When to assist? – Modelling human behaviour for hybrid assembly systems

Markus Huber, Center for Sensorimotor Research, Clinical Neurosciences, LMU, Germany
Alois Knoll, Robotics and Embedded Systems Lab, Dept. of Computer Science, TUM, Germany
Thomas Brandt, Chair for Clinical Neurosciences, LMU, Germany
Stefan Glasauer, Center for Sensorimotor Research, Clinical Neurosciences, LMU, Germany

Abstract

Over centuries the successful cooperation of foreman and assistant has been an important factor of efficiency, as the mutual understanding of each others actions and intentions facilitates a smooth work flow. Hence, a promising option in the field of automation is to develop models of intuitive and adaptive assistance. An artificial system equipped with such models could improve not only efficiency, but also acceptance and safety in human-robot interaction (HRI). As a first step towards this goal, we investigated the timing of the most common physical interaction occurring in HRI - the handing over of items. In an assembly experiment we measured the duration of the different steps. We found that a linear dependency sufficiently describes the relation between the complexity of each working step and its duration. The parameters of the linear model however largely differ between the individual subjects. This linear dependency was used to develop Kalman filters to predict durations and complexities of different assembly steps. We showed that after a short adapting phase we received an accuracy of 18.03%, comparing the predicted assembling duration to the measured duration. The model was integrated into an assistive robot system to demonstrate its robustness and low requirements of the sensory systems.

1 Introduction

For centuries, artisan work has been accomplished by a well-rehearsed and smoothly cooperating team of foreman and assistant. Such teams always have been very successful in producing complex and individual small batches. Even though artificial systems continuously replace more and more human workers, is not yet possible to replace such an efficient cooperative team by robot systems, especially concerning highly complex and skilful tasks. A reasonable first step for automation in this field is to develop artificial systems replacing the assistant [1][2][3].

To this end, a mutual understanding of each other's actions and intentions is needed. More precisely, this means endowing technical systems with natural, anticipatory, and adaptive behaviour. This would lead to higher acceptance, passive safety and efficiency in human-robot teams [4]. To enable technical systems with the ability of adaption and anticipation, a human model is required in order to predict the duration of actions and to identify cues of future progress. In cognitive science, and recently in robotic research, there is a big focus on understanding the mechanisms of goal inference, action understanding and imitation learning, and their description in a way that they can be transferred to technical systems [5][6][7]. The integration of cognitive concepts in robotics promises essential abilities, such as anticipating, what a human partner needs, and learn from human behaviour. However the obvious importance of exact timing for an effective assistance is rarely investigated in the field of human robot interaction. Apparently, an assistant, who is too early, thus constantly holding a tool in front of a foreman will disturb him. In contrast, passing it to late will decrease the efficiency.

In mass production industries, however the issue of optimal timing is even more present and, there also are workflows, which cannot be fully automated and where human fine motor skills are needed. In contrast to the idea of a team of foreman and assistant in industrial settings, humans mostly work alone and are for security reasons, strictly separated from industrial robots. Certainly, in high-wage countries, an efficient planning of manual object assembly is essential to compete. Therefore in this field, several strategies have been developed to efficiently describe and plan human assembly tasks. A widespread concept in industry is the so-called "Method of Time Measurement" (MTM)[8]. It is an integrant to compute execution times, based on predetermined times of basic motion sequences, body movements and a variety of physical task parameters. Furthermore the concepts "Assembly Evaluation Method" (AEM)[9], or a similar method, the "Design for Assembly" (DFA)[10] are used to evaluate the complexity of an assembly. They also critically examine the necessity of the subtasks and the enhancement of the product's construction. Even if the above mentioned methods work well for planning workflows, there is still undergoing research to improve them. Recent works [11] developed a model, based on Goals, Operators, Methods, and Selection Rules (GOMS) language, which predict error-free assembly

ISBN 978-3-8007-3273-9
© VDE VERLAG GMBH · Berlin · Offenbach

times. However, the model is specified for a circuit assembly and therefore does not provide generic predictions. Other groups [12] investigated physical attributes of objects, which influence the difficulty of their assembly. This is used to predict assembly complexity, thus improving DFE. Also the MTM method is continuously improved, e.g. [13], where methods for multidimensional measurements of degree of detail and complexity of manual assembly tasks are developed.
All these methods used to determine the complexity and assembly time are based on predetermined motor time systems and are optimized for large static productions. Furthermore they do not consider individual preferences of a human worker.
Going back to artificial assistive systems, it is however necessary that the system dynamically adapts to its foreman and his level of experience and strategy. If the human foreman after some assemblies decides to change the assembly strategy, the robot system should dynamically respond to the new timings.
In this context, the present study focus on the investigation of temporal behaviour during an assembly scenario. In order to enable robot systems to assist "just in time", a dynamic model of human assembly behaviour is developed and finally integrated into a robot system [3].

2 Experiment

To investigate the assembly behaviour, especially concerning the timing of assistive actions, we designed an experiment where subjects had to perform a typically assembly task. The task was to build towers with cubes and bolts (Figure 1).
The assistive action, we are interested in, is the delivery of a component (in our case a cube) at the right time stamp. We assume that the right time to deliver a cube is the time, when the subject would be reaching for the cube in the slide.
In the experiment however, the subjects had to perform the task by themselves, without an assistant. The cubes were delivered by a slide at a position equivalent to a natural handover position of an oppositely sitting assistant [14]. Because of the slide, the current cube was available for the subject, at any time.

2.1 Experimental setup and method

25 subjects were asked to sit in front of a table and to build towers with the cubes and bolts (A picture of the setup is shown in Figure 2). While the bolts were accessible at the left side of the workspace, the cubes were available in the slide. The cubes had a different number of holes on two opponent sides. To stick the cubes to a tower, one had to insert the bolts into the holes and stick the cube on the top. The duration for each assembly step varied depending on the number of bolts needed to connect the cubes. The subjects had to build 6 towers with a height of 6 cubes. Therefore, there were 6 grasps, but only 5 assembly steps for each tower; leading to a total number of 30 assembly steps per subject.
We recorded the gaze in space of the subjects using an eye-tracking device[1], combined with infrared LEDs attached at the table. The motions of the pointing fingers, thumbs, hands, head and torso were recorded using a magnetic field based Polhemus Liberty tracking device with a sampling frequency of 240 Hz. The subjects were recorded by two video cameras. The multiplicity of the recorded modalities will allow further investigations and models beyond this report.

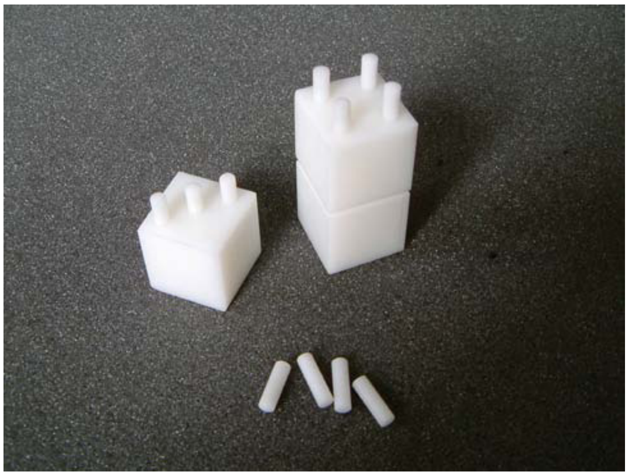

Figure 1 picture of the objects (cubes with bolts) to assemble.

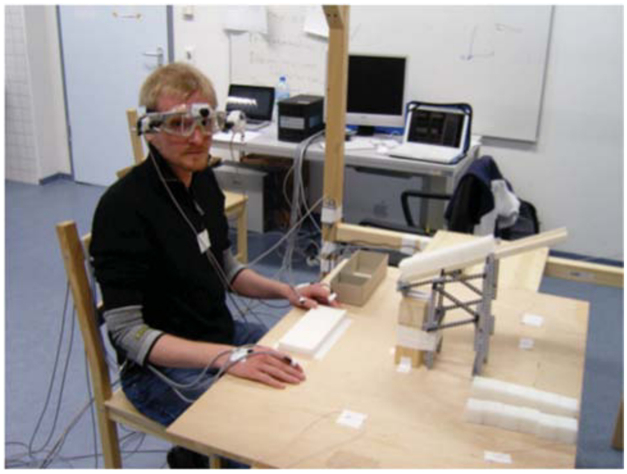

Figure 2 picture of the experimental setup.

2.2 Experimental Results

We varied the complexity of each working step, and therefore the duration, by varying the number of bolts needed to plug one cube to another. Akin to MTM, the durations for assembling a cube to the tower are measured. The time slot begins when grasping a cube from the slide till grasping the next one. To determine the time stamps, we used the recorded hand trajectories. The grasping time stamps are defined as minima-distance from the hand to the slide.

[1] http://eyeseecam.com/

© VDE VERLAG GMBH · Berlin · Offenbach

Two subjects were removed from the data, because of huge outliers in the duration, e.g. because a bolt was falling on the floor and the subjects reached for it.
The measured durations and the corresponding bolt numbers are demonstrated in Figure 3, where different colours indicate the subjects. The mean of the duration for each number of bolts is indicated as black cross.
Data analysis shows, that a larger number of bolt lead to a longer assembly time. As a fist approximation one can assume a linear dependency between the assembly time and the bolt numbers (R^2=0.548), which is displayed as dashed red line in Figure 3. However, the distinct parameters largely differ from subject to subject, so one linear dependency over all subjects does not describe the collective of subjects. Using a separate linear dependency for each individual subject leads to a better description of the data (R^2=0.669). The individual linear dependencies are plotted as coloured dashed lines.

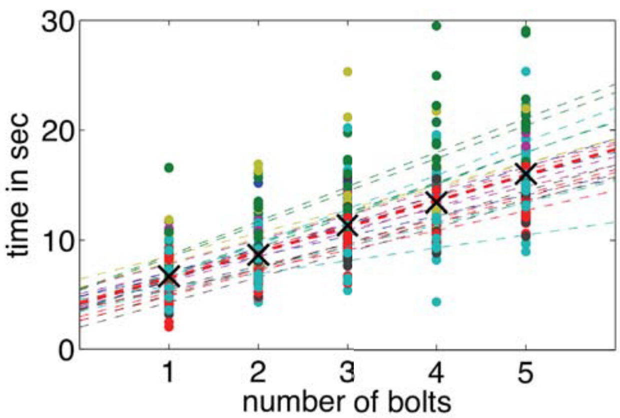

Figure 3 Duration of each working step plotted over complexity (bolts) for all participants (different colours)

Learning effects during the task were observed only at the beginning of the task and only for single subjects. Figure 4 shows the normalized duration, which is calculated by dividing the duration of the trial by the number of bolts. Sudden changes of assembly strategies during the task could not be observed.

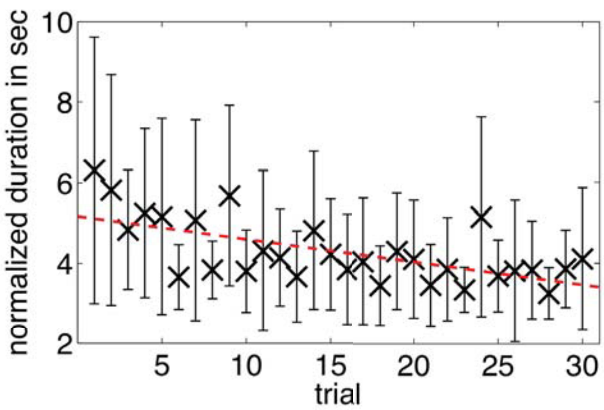

Figure 4 Plot of the normalized duration over the assembly steps.

3 Human assembling model

To develop a robust and precise technique for predicting the duration of each assembly step and thus the upcoming handover assistance of a new item, two separate modules need to be combined: One module that learns the duration of each assembly step depending on its complexity, and an additional module that uses online multimodal sensor input for the prediction of handover probability. The first module will give reliable evidence, on condition that there are no sudden changes and no errors in the production. The second module is intended for unpredictable situations.
To realize this, the present work will suggest a method for the first module, the prediction of durations of consecutive working steps.

Using the results from the human experiments it is known, that a model for efficiently describing the human behaviour for assembly durations must adapt to the individual parameters of a human. The results are implemented within a probabilistic Bayesian framework, realized as Kalman filter [15].
Our experiment showed, that the time needed to request the next object is well described by an average duration and a linear dependence on the number of bolts, which now defines the complexity.
Thus, the duration z_k for the kth object can be described by

$$z_k = x_{1k} + x_{2k} \cdot u_k + v = \begin{bmatrix} 1 & u_k \end{bmatrix} \cdot \underline{x}_k + v$$

with $\underline{x}_k$ being the state vector of the system, u_k is the complexity (number of bolts), and v is the measurement noise. This equation can be used as measurement model with the measurement matrix

$$H_k = \begin{bmatrix} 1 & u_k \end{bmatrix}$$

The system is very simple, and only assumes that both the offset and the slope (the state $\underline{x}$) are constant for each subject:

$$\underline{x}_k = \underline{x}_{k-1} + \underline{w}$$

The Kalman filter is thus a two-dimensional system with a unity system matrix.
The prediction step is

$$\hat{\underline{x}}_{k|k-1} = \hat{\underline{x}}_{k-1|k-1}$$

$$P_{k|k-1} = P_{k-1|k-1} + Q$$

with the system covariance Q. The update step is

$$y_k = z_k - H_k \cdot \hat{\underline{x}}_{k|k-1}$$

$$s_k = H_k P_{k|k-1} H_k^T + r$$

$$K_k = P_{k|k-1} H_k^T s_k^{-1}$$

$$\hat{\underline{x}}_{k|k} = \hat{\underline{x}}_{k|k-1} + K_k \cdot y_k$$

$$P_{k|k} = (I - K_k H_k) P_{k|k-1}$$

with r being the measurement variance.
To predict the duration until the next transfer, we have to use the prediction step and combine it with the measurement matrix

$$\hat{z}_{k+1} = H_{k+1} \hat{\underline{x}}_{k+1|k}$$

The corresponding variance is

ISBN 978-3-8007-3273-9
© VDE VERLAG GMBH · Berlin · Offenbach

$$\mathrm{var}(\hat{z}_{k+1}) = H_{k+1} P_{k+1|k} H_{k+1}^T$$

The algorithm provides a normal probability distribution of the next duration with the expectation $\hat{z}_{k+1}$ and the standard deviation

$$\sigma_{k+1} = \sqrt{\mathrm{var}(\hat{z}_{k+1})}$$

The system and measurement variance, as well as the initial internal state can be calculated from existing data as follows:

The internal state vector $\underline{x}$ is equal to the coefficient estimates from the first order linear regressions analyse of the whole data set.

The variance of the internal state Q is the median over the variance of the individual slopes x_1 and offsets x_2 for each subject. In our case each subject s had to repeat the task six times ($t=1...6$). The variance of the regression estimates x_i for one subject thus is:

$$varq_{i,s} = \mathrm{var}(x_{t,i,s}),$$

where t is the repetition (trial) and i indicates the offset or slope.

$$Q_{ii} = median(varq_{s,i}) = median(\mathrm{var}(x_{i,t,s}))$$

$$Q_{ij} = 0, \quad i \neq j$$

The variance of the measurement r is calculated as the median over the variance of the residuals for each subject. The variance of the residuals for a subject is:

$$varr_s = \mathrm{var}(y_{k,s})$$

where y is the residual, k is the object or assembly step and s the subject.

Therefore, the variance of the measurement r can be written as:

$$r = median(varr_s) = median(\mathrm{var}(y_{k,s}))$$

4 Evaluation of the model

The described Kalman filter is able to predict the duration of consecutive working steps. It delivers a normal probability distribution of the duration, with expectation value $\hat{z}_{k+1}$ and standard deviation σ_{k+1}. The filter continuously estimates individual parameters describing the behaviour of the human user by taking into account the previously found linear dependency between the duration and complexity of the assembly step.

To evaluate the quality of the model, tests are performed, using the data received by our experiment. Figure 5 shows the root mean squares of the difference (residual) between the measured and the predicted durations over all the 30 assembly steps (dark grey bars). It can be observed that in the first two trials the prediction is not very accurate. In these two trials the parameters of the linear model are adapted to fit the individual subjects. As the subjects themselves do not vary much during the assembly, the following trials are predicted more precisely.

The over all accuracy in our experiment over the 30 assembly steps is 2.48 s (RMS). Disregarding the initial adaption phase (the first two trials), the model steadily predicts the durations of the following assembly steps with an accuracy of 2.06 s (RMS). In relation to the average duration of an assembly step in our task (11.445 s), the RMS error corresponds to an accuracy of 18.03%. The coefficient of determination is $R^2=0.691$.

If the averaged linear dependency over all subjects is used to predict the durations (see Figure 3, dashed red line), we get an accuracy of 2.50 s (RMS) after two trials. The RMS error using the averaged linear dependency is plotted as light grey bars in Figure 5.

Obviously in the initial phase, the averaged linear model predicts the same duration than the Kalman filter, because in the first trial there is no information about the residual. However in the following trials the Kalman filter predicts a more precise duration of the subsequent assembly steps, than the simple linear model.

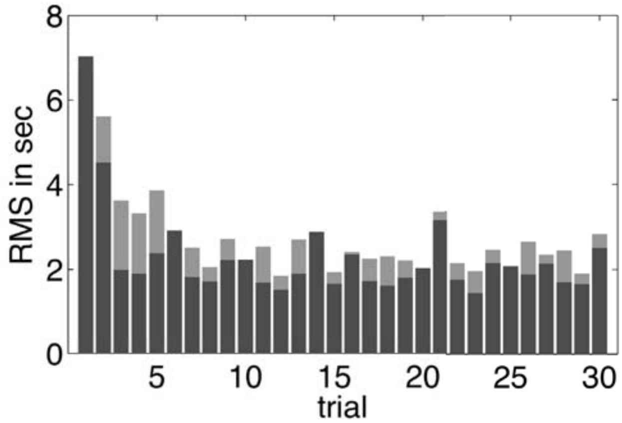

Figure 5 RMS plot of the difference between measured and predicted assembly duration over the trials.

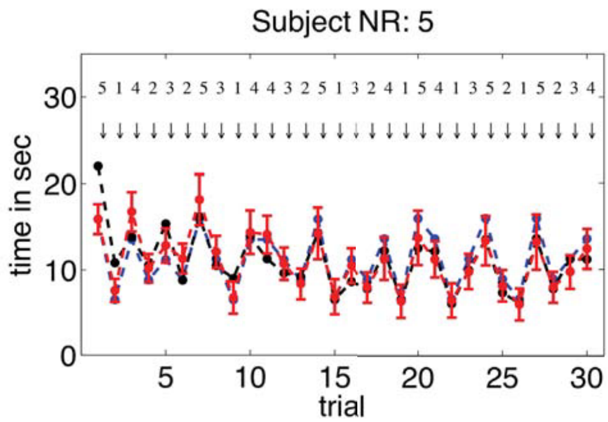

Figure 6 Empirical (black) and predicted duration (red) over all working steps for a single subject.

In Figure 6 the predicted durations (red) are plotted with the measured durations (black) over the trials. The predictions using the averaged linear model, is also plotted (blue). The red error bar indicates the standard deviation of the predicted duration. The numbers above the durations indicate the number of bolts the subject had to use in this assembly step. Here also the initial phase of two trials can be observed, where the model has to adjust the initial parameter to the preferences of the subject. After this adaption phase the model reliably predicts the durations. Fur-

ISBN 978-3-8007-3273-9
© VDE VERLAG GMBH · Berlin · Offenbach

thermore, the measured duration nearly always lies within the standard deviation of the predicted duration (indicated as red error bar).
However, if there is an unexpected event, like an error or a bolt falling down the table, and the subject looks for it, the model obviously is not able include these kinds of events in its prediction. Figure 7 shows a case, where a subject had a problem assembling four bolts in trial (step) 21. It took the subject much longer to assemble a cube with 4 bolts (number over the data point) than in the previous assembly step with 5 bolts. The algorithm interprets the big difference from the predicted duration as general change of the subject's behaviour, thus adjusting the parameter of the linear model. A just one-time disturbance, as shown in this case, leads to defective predictions in the subsequent assembly steps. In order to provide again reliable predictions, the algorithm needs two more trials, as when adapting to a new subject. Therefore after trial 24 the parameters fits the subjects' behaviour again. Such occurring disturbances in the workflow or even worse errors have to be detected by modules using online sensor observations.

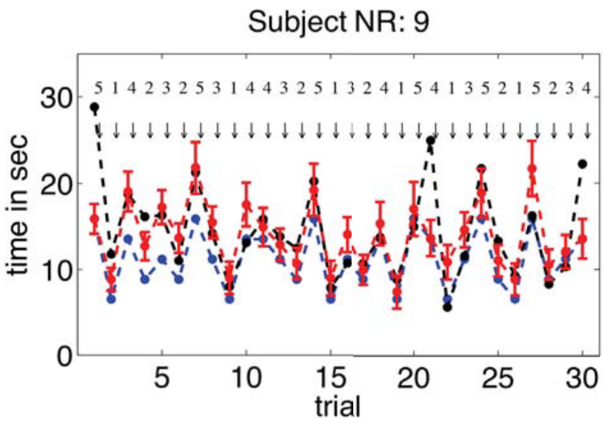

Figure 7 Empirical (black) and predicted duration (red). This subject had an assembly problem at trial 21.

The robust prediction of qualitatively good assembly durations, in workflows without bigger disturbances is remarkable. This is achieved due to its simple underlying mechanism, which is a widely know basic Kalman filter, using a linear model.

5 Implementation in robot systems

For a proof of principle, the developed algorithm has been implemented in an industrial robot JAHIR [3].
Doing so, our concept calculates the duration of a subsequent assembly step. The robot system uses the predicted duration to pass the next object and it reaches the handing over position after the calculated expectation value $\hat{z}_{k+1}$.
The sequence of the cubes and the corresponding bolt numbers were deposited as list in the computer system.
Contrary to the first experiment, the cubes are not always present. The cubes are only available after the predicted duration, when delivered by the robot. In contrast to the previous experiment, the assembly duration is not the time from grasping the first object, until picking the second one from the robot's gripper, but the time from picking an object from the gripper till the user finishes the assembly. In case the calculated duration is longer than the actual assembly and the subject has to wait for the delivery, he has to move his hand to the handing over position, signalling he has finished the assembly and is waiting for the next cube. This signal is used as breakpoint for the measured assembly duration. The hand was detected by an infrared tracking system [2].
Furthermore, the tracking of the hand allowed considering the case, where the algorithm calculates too long durations, by adding a simple logic to the system: If the subjects' hand is on the handing over position, the robot starts to hand over the cube immediately, ignoring predicted value. This avoids waiting times longer than 1.5 seconds, which is the time the robot takes for delivering a cube.

Figure 8 Picture of the JAHIR robot system while it is assisting a user by an assembling task.

For simulating a smooth and safe handover feeling, a handover-position, corresponding to [14] was programmed. As trajectory we used a modified version of the minimum jerk trajectory [16].
A first test has shown that the implementation in robot systems has demonstrated a promising interaction between the robot system and human user. Figure 8 shows a picture of the JAHIR robot system, which is assisting a subject to assemble the towers.
The present model can be easily integrated in most robot systems. The demand on the sensory systems is marginal. For labelling the objects a deposited list in the software, RFID tags or optical tags on the objects are sufficient. For detecting the hand at the handing over position, a simple light barrier would be sufficient.

6 Discussion

Our results provide a first step towards endowing robotic systems with the ability for timely assistance in a cooperative human-robot task. While the present algorithm, which

[2] http://www.ar- tracking.com

ISBN 978-3-8007-3273-9
© VDE VERLAG GMBH · Berlin · Offenbach

implements a primitive intention prediction, only uses information about task complexity and previous durations, a combination with online sensor fusion will provide more robust, time-optimized assistance. Furthermore, comparison of the predictions of both modules would allow detecting irregularities in the workflow and provide the possibility to react adequately. Accurate timing of robotic action in assistance scenarios will thus be an important step toward adapting technical systems to human users in a safe and convenient way.

In human experiment we could show that a linear dependency sufficiently describes the relation between the defined complexity of each working step and its duration. However, the distinct parameters largely differ from subject to subject. These results were implemented within a probabilistic Bayesian framework realized as Kalman filter that is able to predict the duration of consecutive working steps (for an example of its performance, see Figure 6). The filter continuously estimates individual parameters describing the behaviour of the human user by taking into account the previously found linear dependency between the duration and complexity of the assembly step. The simple underlying model enables the method to adapt to subjects within the first two assembly steps. After the short adaption phase, our model predicted the durations of subsequent assembly steps with an accuracy of 2.06 sec (RMS). Referred to an average duration of an assembly step in our task, this is an accuracy of 18.03%. Even though the model is making good predictions in a constant workflow, it is unable to predict unforeseeable event, such as human disturbances and errors.

The developed algorithm has been implemented in an industrial robot [3] to further investigate its performance. The implementation in robot systems is noncritical; the needs to the senor systems are marginal. The present algorithm works with a simple detection of the current object, e.g. a list, RFID tags or colour codes, and a light barrier to detect the finished assembly step.

Acknowledgement:

This work is supported by the DFG excellence initiative research cluster "CoTeSys" (www.cotesys.org). We thank Helmuth Radrich for his help with the experiments and the robot system.

7 Literature

[1] Foster M.E., Giuliani M., Müller T., Rickert M., Knoll A., Erlhagen W., Bicho E., Hipolito N., Louro L.: Combining Goal Inference and Natural-Language Dialogue for Human-Robot Joint Action, in Proceedings of the CIMA Workshop at ECAI, 2008

[2] Hinds P.J., Roberts T.L., Jones H.: Whose job is it anyway? A study of human-robot interaction in a collaborative task, in Human-Computer Interaction 19, p. 151-181, 2004

[3] Lenz C., Nair S., Rickert M., Knoll A., Rösel W., Gast J., Wallhoff F.: Joint-action for humans and industrial robots for assembly tasks, In Human Interactive Communication, RO-MAN, Munich, Germany, p. 130 -135, 2008

[4] Fong T., Nourbakhsh I., Dautenhahn K.: A survey of socially interactive robots, Robotics and Autonomous Systems, 42, p. 143-166, 2003

[5] Oztop, E. and Wolpert, D. and Kawato, M.: Mental state inference using visual control parameters, Cognitive Brain Research, 22, p.129-151, 2005

[6] Cuijpers, R. H. and van Schie, H. T. and Koppen, M. and Erlhagen, W. and Bekkering, H.: Goals and means in action observation: A computational approach, Neural Networks, 19, p. 311-322, 2006

[7] Erlhagen W., Mukovskiy A., Bicho E., Panin G., Kiss C., Knoll A., van Schie H., Bekkering H.: Action-understanding and Imitation Learning in a Robot-Human Task, Artificial Neural Networks: Biological Inspirations, p. 261-268, Lecture Notes on Computer Science, Springer Verlag, 2005.

[8] Maynard H.B., Stegemerten G.J., Schwab J.L.: Methods-time measurement, McGraw-Hill New York, 1948

[9] Miyakawa S., Ohashi, T.: The Hitachi Assembly Evaluation Method (AEM), Proc. International Conference on Product Design for Assembly, Newport, Rhode Island, 1986

[10] Boothroyd G., Alting L.: Design for assembly and disassembly, CIRP annals, 41, p.624-636, 1992

[11] Fish L.A., Drury C.G., Helander M.G.: Operator-specific model: An assembly time prediction model, Human Factors and Ergonomics in Manufacturing, 7, p.211-235, 1997

[12] Richardson M., Jones G., Torrance M., Baguley T.: Identifying the task variables that predict object assembly difficulty, Human factors, 48, p.511-525, 2006

[13] Zaeh M.F., Wiesbeck M., Stork S., Schubö A.: A multi-dimensional measure for determining the complexity of manual assembly operations, Production Engineering, p.1-8, 2009

[14] Huber M., Knoll A., Brand T., Glasauer S.: Handing-over a cube: spatial features of physical joint action. Annals of the New York Academy of Sciences 1431, p. 380-382, 2008

[15] Kalman R.E.: A new approach to linear filtering and prediction problems, Journal of basic Engineering, 82, p.35-45,1960

[16] Huber M., Radrich H., Wendt C., Glasauer S., Knoll A., Brandt T.: Evaluation of a novel biologically inspired Trajectory Generator in Human-Robot Interaction, in Human Interactive Communication, RO-MAN Toyama, Japan, p. 639 - 644, 2009

ISBN 978-3-8007-3273-9
© VDE VERLAG GMBH · Berlin · Offenbach