Incorporating Motion Data and Cognitive Models in IPS²

Michael Beckmann and Jeronimo Dzaack

Chair of Human-Machine Systems, Department of Psychology and Ergonomics, Technische Universität Berlin, Franklinstr. 28-29, 10587 Berlin, Germany
{michael.beckmann,jeronimo.dzaack}@mms.tu-berlin.de

Abstract. In the SFB/TR29 a focus lies on human factors and their integration into Industrial Product-Service Systems (IPS²). These innovative systems are complex and dynamic. Human operators need to be able to perform a multitude of complex tasks in such socio-technical systems, providing a challenge to the operators because of the high complexity. Therefore automatic assistance systems are necessary for the overall reliability and effectiveness of such a system. This article describes a theoretical approach for simulating human behavior with cognitive models. The performed actions are recognized with motion capturing in combination with machine learning. By evaluating the perceived action and reality a description for the situation can be automatically generated in real time. This can be used for e.g. providing the human operator with real time contextual feedback.

Keywords: Cognitive User Models, Error Prevention, Human-Machine Interaction, Machine Learning, Gesture Recognition.

1 Introduction

Industrial Product-Service Systems (IPS²) are innovative production systems incorporating human operators and technical systems [1]. They are characterized by an integrated and reciprocal determined development, provision and utilization of products. In modern production systems the strict separation of both services and products is no longer possible. IPS² provides a possibility to substitute partial aspects of services and products during the design and provision phase.

IPS², like other socio-technical systems, consist of human actors and technical systems. IPS² are highly dynamic systems, because they need to be flexible concerning their operation and adapt to different settings. These dynamics lead to a high complexity of the technical system, requiring a vast amount of information about the systems for a multitude of tasks from the human operators. The information needed by a human operator to perform tasks on the technical systems consists of structural and status information, which is provided by e.g. internal heat sensors. This information needs to be evaluated by the human operator for the optimal course of action. In these systems the human operators cannot specialize in only a small subset of procedures, but need to handle diverse tasks and working conditions. The

V.G. Duffy (Ed.): Digital Human Modeling, HCII 2011, LNCS 6777, pp. 379–386, 2011.

optimization of such technical systems in the design phase only considers technical aspects without regard to the special needs of human operators. These are a crucial part for the optimal and error free operation of IPS².

The research in the field of human factors considers the human operator with limited resources [2] and optimizes the human well-being and system performance by applying theories, principles, data and other methods to the design of human-machine systems [3]. Factors which can influence the human in the system are e.g. age, state of mind, emotions and propensity for common mistakes, errors and cognitive biases. Allowing the human operator to control complexity and dynamics of the technical system increases human reliability and performance. Learning from errors of human operators can be used to reduce future errors. Designing IPS² with consideration to human factors can increase the resilience of the overall system and the performance of the human operators, increasing the overall productivity.

The problem of insufficient regard of human needs, in case of knowledge and courses of action, can be addressed by providing human operators with additional contextual information and assistance during the execution of different tasks. In previous work [4] this was addressed by utilizing an expert to support a novice. The novice was the operator on site and the expert could see the gaze and field of view of the novice at a remote location. In this paper an automatic approach is described for combining motion capturing and machine learning to recognize human behavior or actions. Valid behavior is simulated with cognitive user models. By assessing the system state and comparing the observed human behavior with the predicted courses of action dangerous situations, for the human operator as well as the technical systems, can be recognized and system messages with an evaluation of the situation issued. These messages can be used to generate multimodal feedback for the human operator to prevent malpractice (e.g. errors of emission or confusion). Research about the optimal way of providing multimodal feedback to avert the situation is not part of this subproject, but researched in the subproject B4 “Multimodal user interfaces – Interaction specific warnings and user generated instructions for IPS²” [5].

An example task is the partial disassembly of a system where a part is mounted by several screws. After the removal of the last screw the part falls down if not held properly, damaging the system. In this setting the cognitive user models predict the holding of the part by one hand. The evaluation of the human movement provides the information on the use of a screwdriver, but does not recognize a holding action in the appropriate place for one of the hands. The prediction of the model and the recognized action from the recorded movement data is compared and evaluated. The invalid course of action is detected in real time and an error message generated. An appropriate reaction to the message might be haptic feedback to the hand holding the screwdriver, getting the attention of the human operator, thus stopping the screwing action and preventing the damage to the part. In case of an automatic screwdriver the unscrew action can be directly intercepted by the technical system.

The previous elaboration of IPS² and the mentioned exemplifications make it evident, that contextual real time assistance for human operators in socio-technical systems will increase the overall robustness and effectiveness of these systems. In this article a theoretical approach for recognizing mismatches between human behavior and valid behavior is presented. This approach combines the fields of cognitive modeling and machine learning in addition to motion capturing. By evaluating the

simulated actions and real operator behavior the situation and the state of the human-machine system can be assessed. This evaluation provides an objective and model-based base for contextual feedback for human operators.

2 Approach

The following chapters describe a theoretical approach to automatically recognize mistakes of a human operator in IPS² in real time. Information about the mistakes can be used to provide the human operator with contextual information and prevent harm to the human as well as the technical system.

Cognitive user models (i.e. optimal user models) can be used to model characteristics of the human operator and to predict expected actions. The combination of motion capturing with machine learning provides the possibility to recognize different actions of the human operator. A mistake is detected, if there is a mismatch between the simulated action and the performed action of the human operator. Such a mismatch might be (1) a deviation from the optimal course of action, (2) a dangerous situation for the machine or the human operator, (3) an omission of previous steps or (4) an invalid course of action for the situation. Figure 1 visualizes the general idea.

The following paragraphs describe the different related fields.

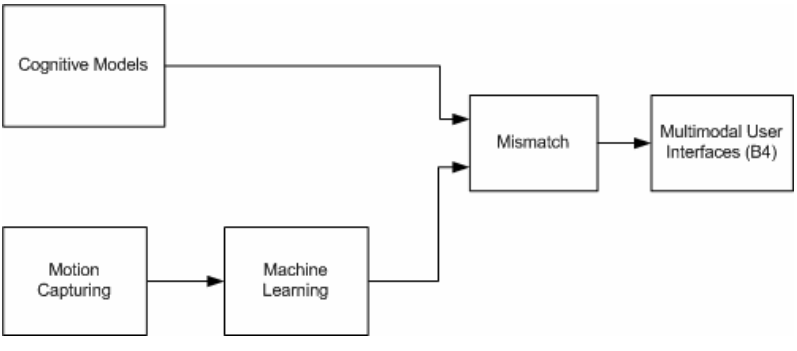

Fig. 1. General visualization of the workflow

2.1 Cognitive User Models

Research in the field of cognitive psychology provides several theories to understand human behavior [6]. Cognitive user models incorporate these theories and make the modeling of selected cognitive processes possible. There are different cognitive architectures for developing models to simulate human behavior. These are basically programming frameworks and can be assigned to the two categories of high-level and low-level cognitive architectures.

High-level architectures use elementary behavior as building blocks and describe cognitive processes as a fixed sequence of actions. They are most suitable for

investigating errors and difficulties in using interfaces in the field of usability and human performance. An example for a high-level architecture is Goals, Operators, Methods and Selection Rules (GOMS) and its derivatives [7], which can be used to predict the time to finish routine tasks in a specific setting. High-level architectures are limited in the description of signal-detection or decision-making processes.

Low-level architectures, like Atomic Components of Thought – Rational (ACT-R) [8] describe human behavior on an atomic level. Most of these architectures use production systems to simulate human processing and cognition. This equals a bottom-up process when processing data and makes the reaction to external stimuli possible, as well as the interruption and resumption of cognitive processes. Complex paradigms, like dual-tasking and decision making, and their underlying processes can be simulated with cognitive models designed with low-level architectures [9]. Because of the bottom-up processing of the information the predictions of cognitive models based on low-level architectures are more natural, taking into account human complexity and dynamics.

Most cognitive user models are applied to predict user behavior offline and the performance of simulations is compared with experiments to validate the model. A valid model explains the observed behavior, but the actual decision process can be completely different from the one simulated.

2.2 Motion Capturing

Motion capturing consists of several techniques to record the movements of real things, like a human being. As an example walking, grasping motion or movements for a serve in tennis can be recorded for later evaluation of the ergonomics of the motion, effectiveness of the motion or realistic animation in a virtual environment. The desired object can be captured using different technologies. An RGB and a depth camera can be used to recognize shapes and match these shapes to some model of the tracked object, recording the information by mapping the shape to the model and mapping the movement of the shape to the movement of the model. Another technique uses markers placed on e.g. the human body in defined positions and tracking of the markers by e.g. infrared cameras and moving the markers in the simulated model under the constraints of a human body. Usually the user needs to assume a specific posture for the initialization of the model. The movement of the model is used for later analysis or online processing.

2.3 Machine Learning

The field of machine learning provides a rich repertoire of methods [10] to learn from observed data. These methods are used to adjust the parameters of a model to a training dataset. Using this trained model new data points from the same distribution as the training data points are transformed, resulting in optimal output.

There are several models with different objectives. The objective can be to extract relevant features of a data set for later processing, which is the field of feature extraction. Another application is the training of a model which learns classes or categories from data points and can assign data points to their respective class. If the labels of the data points in the training set are known, the process is called supervised

learning, otherwise unsupervised learning. In the latter case the model tries to infer classes or categories from the data. Assigning an unknown data point to a class is called classification. An application of supervised learning is handwriting recognition. Images of letters are presented with their label to the model which later discriminates new handwritten digits and assigns the assumed label.

In the case of this approach human movement has to be recognized and classified into actions. In case of human movement the start and end of an action is not known in advance, to solve this problem gesture spotting [11] can be used to get sequences for further classification. This results in the same action not having the same duration or data points. This problem was researched in the field of speech recognition, where timing and pronunciations vary. One approach is to use dynamic time warping [12] which basically stretches or compresses the time axis to generate some matching to a template under several conditions. Hidden Markov models (HMM) represent another prominent approach in the field of gesture recognition [13]. HMMs represent a probabilistic model of latent and observed variables with transition probabilities for the states in the model. For movement the trajectory of the tracking device can be interpreted as the observable variables and the latent variables represent the different actions under consideration.

3 Concept

The aim of this theoretical approach is to provide the means to process human behavior and the technical system state in real time to automatically recognize a mismatch or error in the action of the human operator, so automatic assistance can be provided depending on the type of mismatch and the knowledge of the involved human.

The data flow of this theory can be summarized into four consecutive steps: (1) data gathering, (2) classification and simulation of the human action (3) comparison of the classification with the simulated action and (4) evaluation of the actual system state in respect to the comparison. This process is interrupted when the last step generates a system message indicating an error which needs to be handled by providing the human operator with feedback. The approach is illustrated in Figure 2.

In the first step the data for further processing is gathered. The movement data of the human operator can be collected using motion or gesture tracking systems. The technical systems in industrial systems usually integrate several sensors to observe the system state. Further sensors can be added for additional data. Initial data about the state space of a task can be collected using a setup step.

The second step requires the evaluation of the captured movement using machine learning. Actions like holding some part, moving a part from one location to another or unscrewing a screw are recognized from the recorded movement in real time. Concurrent actions need to be recognized, because the use of different tools at the same time is common in industrial settings. A scenario for concurrent actions is the removal of a part which involves an unscrew action and a holding action at the same time, because the part would fall down otherwise and get damaged. To recognize these concurrent actions, the body movement is segmented into different areas which are involved in different actions. Action recognition is done on each part

separately and the performed actions are deduced from the results of these segments, making recognition of concurrent actions possible. The optimal behavior of the operator is simulated with cognitive user models. These models provide quantitative and theoretical data about ideal operators for specific tasks, like execution times. Action sequences, times and further characteristics are extracted from the classification and simulation for further comparison to determine the validity of the performed action. In the case of the described scenario the model would issue the action of holding the part and unscrewing the screw. These commands also contain timing information for the execution of the task.

In this third step real-time comparison of the characteristics is performed. The differences between the simulation and the recognized action can be computed and further processed by the evaluation step. For the scenario described above, if the human operator wants to unscrew the screw, but does not hold the part, there is a mismatch between the required holding action of the cognitive model and the absence of this action in reality. A message about this mismatch will be sent to the following evaluation step.

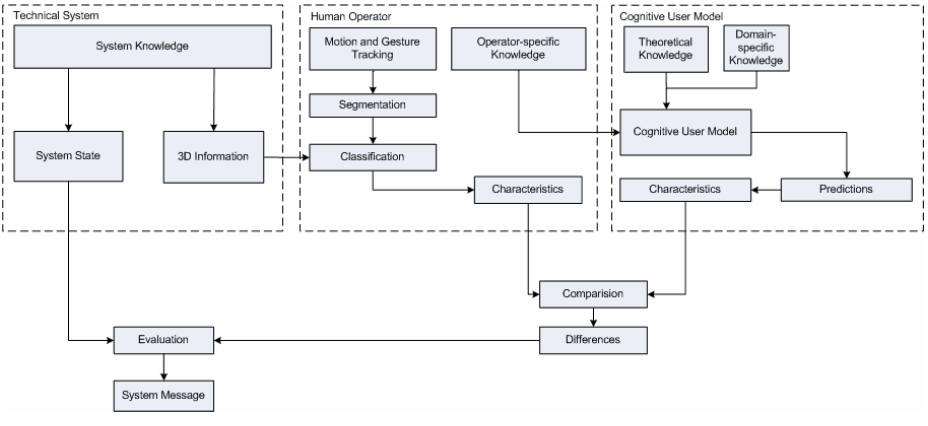

Fig. 2. Conceptual framework to automatically support human-computer interaction in IPS²

The last evaluation step takes into account the differences computed in the previous step and the sensors of the technical system. Dangerous actions can be recognized in this step, e.g. moving a hand near a dangerous area in which the heat did not dissipate yet. On recognizing a valid action the state of the model can be adjusted to the taken action and the process continued. Otherwise appropriate system messages can be generated, which serve as a basis for providing the human operator with contextual feedback, which is part of the subproject B4. The message consists of the system state, the action performed by the human operator, a set of valid actions and the evaluation of the severity of the invalid action. In the described case the mismatch of not holding the part is critical, because unscrewing the screw leads to damage of the technical system, the missing required action is holding the part, which is the required action as well.

4 Discussion

The human operator needs assistance for increasing the robustness and effectiveness of IPS². This article described a theoretical approach to support human operators in such socio-technical systems. The automatic assessment of a working task can be used to generate contextual feedback for the human operator to prevent the execution of improper courses of action. Several challenges need to be tackled to realize the described approach.

The first challenge is the real time processing of the data. This can be solved by predicting courses of action in advance by cognitive user models and adjusting the models to a recognized action, in case it was valid. The models must contain mechanism for interruption, modification and resumption. Differences in action must be recognized as early as possible, because of the real time processing. Thus low-level architectures should be used, because human behavior is simulated on an atomic level. This makes an earlier comparison of actions possible. Tools exist to map high-level models to low-level models [14], making the design of such models possible at a more abstract level and transforming them to an easier level for evaluation.

Another challenge is the action recognition using machine learning. The model can be trained offline, considering the past data. The transformation of such a learned model is fast for new data and can be done in real time. An open problem is the type as well as the amount of movement information, e.g. trajectories, which is necessary to reach a reliable recognition of different actions. The previously mentioned work in gesture recognition and other fields suggest that this problem can be solved by studying different practical settings.

A first step in the realization of this approach is the simulation and recognition of a basic normalized industrial process involving human operators and technical systems. For this data from the technical systems, movement data and additional sensory data is used for the evaluation of the situation. In a second step the data provided by additional sensors is reduced. In a further step individual cognitive user models for the human operators will be employed, providing the human operators with individualized contextual support.

Additional value is added by this approach in addition to the contextual information by evaluation of the rich set of sensory data collected for this approach. Collected data from different operators, tasks and systems can be evaluated in a further step to plan training workshops for skill enhancement for those tasks. Training videos can be created semi-automatically from the recorded movement data and optimal process patterns derived from recorded work of an expert with additional information added by a human about critical situations. The additional information is necessary, because dependencies and reasoning cannot be reconstructed on movement data alone. Detected common problems in operation can be perceived and the process for a task or the technical system adjusted.

The method of providing the human operator with appropriate feedback to avoid malpractice or error is not covered in this approach. This is a well studied field in human-computer interaction. The subproject B4 of the SFB/TR29 is researching the optimal methods for providing this feedback.

5 Summary

IPS², like other socio-technical systems, are complex systems. Human factors are not considered appropriately when these systems are designed. The multitude of tasks a human operator needs to perform and the know-how necessary for these tasks increases rapidly, reducing the practice of the human operator in each task and thus the familiarity. This, in turn, has a negative impact on the performance of the involved humans. The described approach for observing human operators and recognizing their behavior from collected data with the help of machine learning and simulating valid behavior with cognitive models, identifying mismatches, can be used to provide the human operator with contextual feedback. This reduces the complexity of the technical system and helps the human operator to cope with the multitude of facets in these socio-technical systems.

Acknowledgments. We express our sincere thanks to the Deutsche Forschungsgemeinschaft (DFG) for funding this research within the Collaborative Research Project SFB/TR29 on "Industrial Product-Service Systems – dynamic interdependency of products and services in the production area". We thank Bo Höge for his support concerning this work.

References

1. Meier, H., Völker, O.: Industrial Product-Service-Systems – Typology of Service Supply Chain for IPS^2 Providing. In: Mitsuishi, M., Ueda, K., Kimura, F. (eds.) Manufacturing Systems and Technologies for the New Frontier, pp. 485–488. Springer, London (2008)
2. Jones, B.D.: Bounded Rationality. Annu. Rev. Polit. Sci. 2, 297–321 (1999)
3. Human Factors and Ergonomics Society, `http://www.hfes.org/web/abouthfes/about.html`
4. Höge, B., Schlatow, S., Rötting, M.: A Shared-Vision System for User Support in the Field of Micromanufacturing. In: Proceedings of HCI International 2009 – Posters, pp. 855–859. Springer, Heidelberg (2009)
5. Schmuntzsch, U., Rötting, M.: Multimodal User Interfaces in IPS^2. In: Proceedings of HCI International 2011. Springer, Heidelberg (in press, 2011)
6. Solso, R.L.: Cognitive Psychology, 6th edn. Allyn and Bacon, Nedham Heights (2000)
7. John, B.E., Kieras, D.E.: The GOMS family of user interface nalysis Techniques: Comparison and contrast. ACM Transactions on Computer-Human Interaction 3(4), 320–351 (1996)
8. ACT-R Theory and Architecture of Cognition, `http://act-r.psy.cmu.edu/`
9. Dzaack, J., Kiefer, J., Urbas, L.: An approach towards multitasking in ACT-R/PM. In: Proceedings of the 12th Annual ACT-R Workshop, Italy (2005)
10. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, New York (2006)
11. Junker, H., Amft, O., Lukowicz, P., Tröster, G.: Gesture spotting with body-worn inertial sensors to detect user activities. Pattern Recognition 41(6), 2010–2024 (2008)
12. Salvador, S., Chan, P.: Toward accurate dynamic time warping in linear time and space. Intelligent Data Analysis 11, 561–580 (2007)
13. Ying, Y., Randall, D.: Toward natural interaction in the real world: real-time gesture recognition. In: ICMI-MLMI 2010, pp. 15:1–15:8. ACM, New York (2010)
14. Amant, R.S., Freed, A.R., Ritter, F.E.: Specifying ACT-R models of user interaction with a GOMS language. Cognitive Systems Research 6(1), 71–88 (2005)