Doctoral Dissertation: Summary

Towards Human-in-the-Loop Cyber-Physical Systems

Róbert-Adrian Rill

Babeş-Bolyai University, Cluj–Napoca

Faculty of Mathematics and Computer Science

The Doctoral School in Mathematics and Computer Science

Scientific supervisor:

Dr. Horia F. Pop

2021

Abstract

The present work tackles four common aspects of the human-in-the-loop challenge, one of the major questions in cyber-physical systems (CPS) related research. First, strategic decision measures are identified in a divided attention task not considered by previous studies, and their importance is shown in predicting human performance. This supports the creation of a more complete human behavior model to be integrated into CPS. Second, a generic data-driven approach is proposed for predicting human errors from eye-gaze and hand-motion features with high accuracy. Anticipating human errors facilitates efficient computer intervention and the reliable operation of complex systems. Third, it is shown that demanding gaze-based control of interfaces can be productive in terms of strategies, even though it impairs performance. This promotes intuitive interaction with computers and is especially important in cases where traditional control methods are not feasible. Fourth, an intuitive monocular vision based ego-speed estimation, and a time-to-collision prediction algorithm is investigated using as input two video streams, recording the frontal road view and the driver's perspective. Leveraging smartglasses as sensory devices and combining them with deep learning algorithms improves the decision making of human assistance systems. The results contribute to increasing human-awareness of CPS and to incorporating humans into the loop as an integral part.

Keywords: cyber-physical system, human-in-the-loop, divided attention, human performance, gaze tracking, smartglasses, monocular vision, driver assistance system.

Introduction

Cyber-Physical Systems (CPS) are integrations of computation and physical processes that provide and make use of data-accessing and data-processing services. They represent one of the most promising directions in the development of computer science and information and communication technologies that can change every aspect of life. They have a significant economic and societal potential that could dominate the 20th century information technology revolution. Similarly to how the internet revolutionized the accessibility of information and transformed the way humans interact and communicate with each other, CPS transform the way humans interact with and control the surrounding physical world.

CPS can be regarded as a confluence of wireless sensor networks, Internet of Things, robotics, machine learning for monitoring and controlling the physical world and obtain highly adaptable environments. They form the basis of next-generation infrastructure and emerging intelligent services are improving the quality of life. Humans are always involved as an essential part of any CPS, yet current solutions still leave the human factor behind. Before reaching the full potential of human-in-the-loop CPS, a taxonomic foundation needs to be established that helps in facing the challenges to be overcome: understand human control types, model human behavior comprehensively, incorporate models into the architecture itself and do not treat humans as external elements of the control loop, make interaction with computers intuitive, increase human-awareness of the system so that technology would adapt to humans and not the other way around, recognize the need of keeping humans in the loop in spite of superhuman performance achievements. The exhaustive theoretical foundations are yet to be assembled.

The aim of the present dissertation is to solve the human-in-the-loop challenge, one of the major CPS related questions. Specifically, incorporate the human efficiently into the system, treat human behavior as an integral part, adapt technology to human needs, so that we can interact intuitively with computers in CPS environments in order to achieve common goals. The significance of my research in terms of human-in-the-loop aspects can be summarized in four points: (i) create a more complete human behavior model by identifying and monitoring strategic decision measures, (ii) promote the reliable operation of complex systems by anticipating human errors, (iii) facilitate the intuitive interaction with computers through gaze-based control of interfaces, (iv) improve the decision making of human assistance systems by leveraging smartglasses as sensory devices and deep learning based monocular vision algorithms.

The long term initiative of solving the human-in-the-loop challenge gives rise to several research questions to be investigated. Why do current CPS solutions still leave the human factor behind? What challenges do we still face in order to integrate the human component efficiently and create more human-aware systems? How should the computer adapt to changing human behavior, how can it recognize unusual situations and identify when the human needs help? More generally, what are desirable ways of teaching computers to reach a more beneficial interaction over time? Can we facilitate the decision making of computers, so that they can intervene in time in order to avoid situations dangerous for humans? How should the computer provide assistance to promote human performance in the long run? What are alternate ways of controlling or interacting with computers in situations that entail this? How can we leverage newly emerging technologies to facilitate the solution of CPS problems? How can we take advantage of pervasive intelligent sensing devices? How can we enforce CPS to ensure the privacy of their users?

Examples of real-life professions in CPS scenarios where the intuitive interaction with computers is desired include, but are not limited to air traffic (airspace) controllers, nuclear power plant operators, manufacturing engineers, surgeons, physicians, medical providers (assisting the elderly and people with impairments), civil emergency operators, call center agents, aircraft pilots, automobile drivers.

The outlined long term ambition calls for the measurement of human parameters, the monitoring of human actions and behavior, the identification of factors influencing human performance, the anticipation of human errors – all this in environments requiring the divided attention of participants, which is generally the case in CPS scenarios. Besides this, it is desired to consider the substitution of traditional ways of interacting with computers in specific cases with innovative approaches and to take advantage of ubiquitous smart devices for sensing purposes, not to mention the incorporation of state-of-the-art artificial intelligence advancements into the overall solutions.

The main contributions of the present dissertation to the existing scientific literature can be summarized as follows:

- the identification of strategic decision measures in a divided attention task not considered by previous studies,
- demonstrating the importance of strategies in predicting human performance,
- the proposition of a generic data-driven approach for predicting human errors from

eye-gaze and hand-motion features with high accuracy,

- showing that even though switching from the mouse cursor control to the more demanding gaze-based control impairs performance, it is possible to make progress in terms of strategies – this is especially important for people for whom the traditional control method is not an option,
- the investigation of an intuitive monocular vision based ego-speed estimation, and a time-to-collision prediction algorithm using as input video stream from a spherical camera and smartglasses, with the aim to enhance driver assistance systems.

Thesis statements and contributions

The work presented in the dissertation can be superimposed along the lines of the human-in-the-loop challenge. To address parts of this large-scale challenge, one mock-up interface environment and one real CPS scenario was designed and investigated, and data collection was performed with human participants. Firstly, a dashboard environment was implemented for divided attention tasks in order to mimic CPS situations and to investigate human performance. Secondly, a real setup was built for data collection in a driving scenario, composed of smartglasses and a 360-degree (spherical) camera.

In the first data collection instance a divided attention task was designed and implemented, and a longitudinal study was performed with 10 participants [6]. After an extensive qualitative and quantitative evaluation of the experimental data, I characterized the strategies of subjects, i.e. their method of problem-solving/decision-making. In order to demonstrate the importance of identifying human problem solving strategies in divided attention environments, I propose the following thesis statement.

Thesis 1. **Strategic predictors of performance.** In divided attention environments the performance of humans can be predicted by identifying and measuring their ability to make strategic decisions, without analyzing ability constructs and personality traits. Moreover, my finding is that the strategy of planning ahead and executing an action before a situation would become critical is a more influential predictor than the strategy of selecting the most urgent task or action.

The most important strategy is called planning and has the effect of reducing later cognitive load or timing constraints and the statistical analysis showed that it explains almost as much variance in performance (47%) as the other three, more straightforward predictors together (51%): selecting the more urgent task and user action between multiple simultaneous possibilities, respectively, and choosing a response within the same task when the opportunity is present.

The results of study [6] indicate that considerable differences in the divided attention ability of normal subjects can be identified early, with minimal efforts, using a small sample and applying a relatively short period of practice. The carefully crafted circumstances regarding the design of our special divided attention task and the experimental procedure helped to find and highlight relevant explanatory variables, called strategic decisions. The findings indicated that distinct strategies influence general performance and give rise to different and diverging learning trajectories. My work emphasizes the importance of describing and analyzing strategies, which in turn can substantially influence performance in complex tasks and may serve training needs.

Measuring the ability of making strategic decisions contributes to the development of a more complete model of human behavior and facilitates the intuitive assistance from computers when the human deviates from the right strategy, if those have been identified beforehand. The performance of participants in our divided attention task is determined by the number of errors committed. Therefore, to anticipate human errors from their behavior, I examined various algorithms in an attempt to predict omission errors before they would occur [9, 10]. The long-term goal is to decide when and how the computer should intervene in order to avoid critical situations. The following thesis statement summarizes these efforts.

Thesis 2. **Predicting human errors.** Quantitative features of eye-gaze movement and hand motion (e.g., changes of positions over time) can be used to predict human errors, i.e. to classify successful and failed user actions with high accuracy.

Using a data-driven approach to predict human errors, I evaluated several classical machine learning algorithms and compared them with a more traditional temporal modeling approach and a deep learning based LSTM model. Employing a leave-one-subject-out cross-validation procedure I achieved a best classification accuracy of up to 86%.

My results and efforts have implications for the design and evaluation of predictive interfaces involving decision making under time pressure. Such intelligent interfaces are being increasingly integrated into diverse technological areas. In complex high-risk environments, where humans represent a crucial part of the system and their attention is often divided between simultaneous activities, imminent human errors may have serious consequences. Computers may need to anticipate user actions and errors in order to provide assistance and avoid dangerous situations. Enhancing interfaces with predictive capabilities can facilitate efficient human-computer interaction, and therefore promote the safe and reliable operation of complex systems.

The studied divided attention task uses mouse cursor based control, which is a traditional approach for computerized applications. However, interfaces with human eye-gaze control represent a promising alternative by widening the possibilities for personalized interaction. Gaze-based control causes a demanding burden in dynamic tasks, but traditional control methods may often not be feasible, such as in the case of systems for people with disabilities, or tasks where the hands of the human are occupied (e.g. a surgeon in the operating room, bomb disposal). Accordingly, I investigated the effects of switching to exclusively gaze-based control [3, 4] and I introduce the thesis statement below.

Thesis 3. **Demanding gaze-based control.** Even though gaze-based control of interfaces is more demanding in divided attention environments than mouse cursor control for instance, learning to use the right strategy allows human participants to perform sufficiently well. Therefore, gaze-based control can make interaction with computers productive, especially for people with restricted capabilities.

After the longitudinal study with the mouse control version of the task, the experiments were repeated with nine out of the ten original participants. Despite carefully controlling experimental and design aspects (the participants were experienced users of the mouse control version of the task, the difficulty was reduced to the more demanding conditions and the parameters of gaze input were selected based on previous research findings), the performance of subjects was considerably impaired. In contrast to initial assumptions, experienced users could not get used to gaze-based control in the amount of experiments performed. On the other hand, I considered the previously identified strategies of users, and found that it is possible to make considerable progress even during a short amount of practice.

The results of this study provide evidence that the adoption of interfaces controlled by human eye-gaze in cognitively demanding environments require careful design, proper testing and sufficient user training. This is especially important in the case of people with physical disabilities (for instance amyotrophic lateral sclerosis), for whom gaze-based interaction might represent the only means to communicate and interact with technology and other people.

In the experiments a commercial gaze tracking device was used. Nonetheless, I also contributed in developing appearance-based gaze tracking algorithms [8, 12]. Gaze direction can be tracked with smartglasses, an element of the list of intelligent devices carried by people that turn humans into "walking sensors" in CPS settings, not to mention that smartglasses can record a video stream from the users' perspective as well. The next and final thesis statement demonstrates the usefulness of this device in a real divided attention environment, namely the driving scenario, along with the possibility of leveraging the quickly developing deep learning based computer vision algorithms with the long-term objective of improving and/or complementing human assistance systems.

Thesis 4. Smartglasses as sensors. Smartglasses are valuable sensor devices, and combining them with deep learning based monocular vision algorithms facilitates the decision making of human assistance systems, e.g., in self-driving cars.

In the accompanying studies I explored two monocular vision based automated driving tasks, namely ego-speed estimation [2] and time to collision prediction [5], with the long-term objective of improving and/or complementing driver assistance systems. Car stop situations were utilized as collision surrogates to obtain training data, in order to overcome the data scarcity problem regarding collisions. I exploited deep learning based object detection to identify the lead vehicle during driving, and investigated the object detection as well as monocular depth based features to estimate time-to-collision.

Conclusion

The dissertation elaborates the significant impact of CPS on economy and society, and how they will change every aspect of life by forming the basis of next-generation infrastructure and emerging intelligent services. Because humans are always involved as an integral part, the human-in-the-loop challenge is one of the major questions research has to address. The works superimposed along the lines of this large-scale challenge tackle common human-in-the-loop aspects and demonstrate the four thesis statements confirming them by scientific publications.

Thesis 1 highlights the importance of identifying strategic decision measures, which may be critical when humans control processes in divided attention environments. Monitoring the ability of making strategic decisions (instead of or together with ability constructs and personality traits) also contributes in creating a more complete human behavior model to be integrated as an essential part of CPS.

When the human is an active participant in decision making, the probability of human error causing a system failure can be high. Robust CPS call for real-time predictive models that are able to recognize dangerous situations, control the outcomes, maintain stability and accuracy and adapt to changing human behavior and to dynamic environments. Thesis 2 and the related studies contribute in this regard, by introducing a generic data-driven approach for predicting omission errors from gaze-movement and hand-motion features in dynamic environments.

According to CPS requirements, in order to reach common goals efficient human-computer interaction is needed. Thesis 3 suggests that in order to make interaction with computers intuitive, gaze-based interface control should be combined with traditional methods, especially in cases when using hands is not a viable option. Monitoring the strategic decisions can also facilitate adequate machine intervention when the human deviates from the right strategies, if those have been identified beforehand (cf. Thesis 1).

Thesis 4 proposes to improve the decision making of human assistance systems by leveraging smartglasses and deep learning based monocular vision algorithms. Additionally, smartglasses are members of the various intelligent wearable sensory devices and assist in alleviating the difficulty of monitoring and modeling human behavior.

The extensive plan of my research involves the construction of a general framework for the design and realization of CPS that facilitates the efficient inclusion of the human into the (control) loop. Related to the divided attention task and the driving scenario, I also worked on proposing an architecture for goal-oriented CPS that considers the spatio-temporal context of events, promotes anomaly detection, and facilitates efficient human-computer interaction. The accompanying research has been started [1, 7, 11] – nevertheless this points beyond the scope of the dissertation.

Notably, the human-in-the-loop concept is not specific just for CPS, but is also in the focus of the current Human-Centered AI initiative of the European Union: the goal of the HumanE AI[1] project is to *"design and deploy AI systems that enhance human capabilities and empower both individuals and society as a whole to develop AI that extends*

[1] https://www.humane-ai.eu/

rather than replaces human intelligence". The vision involves new solutions to human-computer interaction problems, with a strong emphasis on ethics and related legal and social considerations.

References

[1] R. A. Rill. (2016). *Measuring Human Divided Attention in Cyber-Physical Systems.* In 11th Joint Conference on Mathematics and Computer Science, Eger, Hungary.

[2] R. A. Rill. (2020). *Intuitive Estimation of Speed Using Motion and Monocular Depth Information.* Studia Universitatis Babeş-Bolyai Informatica, 65(1):33–45.

[3] R. A. Rill and K. B. Faragó. (2018). *Gaze-based Cursor Control Impairs Performance in Divided Attention.* In The 11th Conference of PhD Students in Computer Science, pages 140–143, Szeged, Hungary.

[4] R. A. Rill and K. B. Faragó. (2018). *Gaze-based Cursor Control Impairs Performance in Divided Attention.* Acta Cybernetica, 23(4):1071–1087.

[5] R. A. Rill and K. B. Faragó. (2021). *Collision Avoidance Using Deep Learning-Based Monocular Vision.* SN Computer Science, 2:375.

[6] R. A. Rill, K. B. Faragó, and A. Lőrincz. (2018). *Strategic Predictors of Performance in a Divided Attention Task.* PLOS ONE, 13(4):1–27.

[7] R. A. Rill and A. Lőrincz. (2019). *Cognitive Modeling Approach for Dealing with Challenges in Cyber-Physical Systems.* Studia Universitatis Babeş-Bolyai Informatica, 64(1):51–66.

[8] R. A. Rill, Z. Tősér, and A. Lőrincz. (2015). *Facial Landmark Based Gaze Direction Estimation.* In KEPT: Knowledge engineering Principles and Techniques, Cluj-Napoca, Romania.

[9] R. R. Saboundji and R. A. Rill. (2019). *Predicting User Actions Under Time Constraints in a Divided Attention Task.* In Pannonian Conference on Advances in Information Technology (PCIT 2019), pages 77–83, Veszprém, Hungary.

[10] R. R. Saboundji and R. A. Rill. (2020). *Predicting Human Errors from Gaze and Cursor Movements.* In International Joint Conference on Neural Networks (IJCNN), pages 1–8, Glasgow, United Kingdom.

[11] Z. Tősér, R. Bellon, D. Hornyák, H. Zoltán, T. Kozsik, R. A. Rill, and A. Lőrincz. (2015). *Functional Programming Framework for Cyber-Physical Systems.* In KEPT: Knowledge engineering Principles and Techniques, Cluj-Napoca, Romania.

[12] Z. Tősér, R. A. Rill, K. Faragó, L. A. Jeni, and A. Lőrincz. (2016). *Personalization of Gaze Direction Estimation with Deep Learning.* In G. Friedrich, M. Helmert, and F. Wotawa, editors, KI 2016: Advances in Artificial Intelligence, pages 200–207, Cham. Springer International Publishing.

Contents of the Doctoral Dissertation