Guest Editorial: Introduction to the Special Issue on Long-Term Human Motion Prediction

***Index Terms*—Human motion prediction, motion planning.**

I. Introduction

LONG-TERM human motion prediction is a key ability for advanced autonomous systems, especially if they operate in densely crowded and highly dynamic environments. In those settings understanding and anticipating human movements is fundamental for robust long-term operation of robotic systems and safe human-robot collaboration.

Foreseeing how a scene with multiple agents evolves over time and incorporating predictions in a proactive manner allows for novel ways of planning and control, active perception, or human-robot interaction. Recent planning and control approaches use predictive techniques to better cope with the dynamics of the environment, thus allowing the generation of smoother and more legible robot motion. Predictions can be provided as input to the planning or optimization algorithm (e.g. as a cost term or heuristic function), or as additional dimension to consider in the problem formulation (leading to an increased computational complexity). Recent perception techniques deeply interconnect prediction modules with detection, segmentation and tracking, to generally increase the accuracy of different inference tasks, i.e. filtering, predicting. As also indicated by some of the scientific works accepted in this special issue, novel deep learning architectures allow better interleaving of the aforementioned units. The recent trend shows that the community aims to develop end-to-end approaches that make use of prediction as the gluing unit between perception and planning/control units.

Examples of applications benefiting from predictive human motion models include autonomous driving, intelligent video surveillance, collaborative robots in industry, service and intralogistic robots. Intelligent autonomous vehicles critically depend on predicting the motion of the surrounding vehicles and pedestrians when dealing with busy urban streets, intersections, round-about or crowded highways. State-of-the-art methods are capable of modelling different drivers' behaviors and proactively planning ego-vehicle motion considering those. Modern navigation systems for robots operating in human-shared spaces (e.g. intralogistic- or social spaces) require high predictive capabilities to manage the complexity of those highly dense and dynamic environments.

The growth of the community interest to this area in recent years is evident, in particular from the publication trends in the regular paper sessions of the recent robotics, computer vision and machine learning conferences (ICRA, IROS, RSS, ECCV and others). Furthermore, the area is recognized as a major component/research direction by several big players in the automotive industry.

II. Guide to the Special Issue

Following the growing interest to the the topic of human motion prediction, we present this special issue that collects relevant work and further extends the existing state of the art. The special issue covers different relevant sub-areas, e.g. predictive planning and control, learning-based prediction approaches, human-robot interaction.

Given these topics in motion prediction, human-aware planning and human-robot interaction, the papers, accepted in this special issue, represent a remarkably balanced variety in tasks and application domains.

A. Accepted Papers

The following 14 papers have been accepted to the special issue [item 1), 2), 3), 4), 5), 6), 7), 8), 9), 10), 11), 12), 13), and 14)] in the Appendix. In the following we shortly describe them and categorize in four main sections: methods for trajectory prediction in Section II-A1, methods for planning and control in Section II-A2, methods for full body motion modeling and prediction in Section II-A3 and new context-rich predictions approaches in Section II-A4.

1) Methods for Trajectory Prediction: Several articles focus on the core task of trajectory prediction [item 1), 12), 10), 6), 13), and 4) in the Appendix]. Two papers propose novel representations for encoding interactive dynamics in crowded scenes with multiple people [item 1), and 12) in the Appendix]. Davchev *et al.* [item 1) in the Appendix] design a modular system for learning structured representations of spacial and interactive dynamics for trajectory prediction in crowded scenes. The system includes environment-specific spacial encoding, a global dynamics component to model the evolution of the scene in latent image space and a local behavioral model to predict motion conditioned on it. Zhao and Oh [item 12) in the Appendix] propose an approach to learn, detect and extract joint motion patterns from sequential trajectory data. Huang *et al.* [item 6) in the Appendix] highlight intention reasoning for trajectory prediction using a Mutable Intention Filter (MIF) and a Warp LSTM. MIF performs particle filtering to create a distribution

Digital Object Identifier 10.1109/LRA.2021.3077964

over all potential goal regions, while the Warp LSTM generates offsets on a full trajectory predicted by the nominal intention-aware linear model. Habibi *et al.* [item 1) in the Appendix] tackle learning and updating the knowledge base from multiple sources of trajectories with a similarity-based model fusion algorithm. Finally, two papers focus on pedestrian-vehicle interaction scenarios. Ivanovich *et al.* [item 10) in the Appendix] provide a self-contained tutorial on a conditional-variational autoencoder (CVAE) approach. The latter can produce a multi-modal probability distribution over future human trajectories, conditioned on past interactions and candidate robot future actions. Anderson *et al.* [item 4) in the Appendix] propose a probabilistic method which uses a risk-based attention mechanism to learn when pedestrian should yield to the vehicle, and a model of vehicle influence to learn how yielding affects motion.

2) Methods for Robot Control and Motion Planning: Three articles address the problem of motion planning and control considering human motion prediction. Two of them take the approach of simultaneous prediction and planning: Bajcsy *et al.* [item 5) in the Appendix] for a mobile robot and Ghadirzadeh *et al.* [item 7) in the Appendix] for a robot manipulator. Bajcsy *et al.* [item 5) in the Appendix] combine ideas from robust control and intent-driven human modeling to achieve robustness against model errors while reducing the conservatism of the traditional reachability-based predictors. Belief-pruned reachability set is used in the spline-based planner to plan six-second trajectories in a receding-horizon fashion. Ghadirzadeh *et al.* [item 7) in the Appendix] present a reinforcement learning based framework for human-centered collaborative systems. Predictions and proactive robot policies are trained jointly in an end-to-end fashion from full-body motion capture data using graph convolutional networks and Q-learning. Differently, Konar *et al.* [item 7) in the Appendix] learn socially-compliant navigation policies from demonstrations, by mimicking humans. Authors propose an efficient sampling-based approximation to enable model-free inverse reinforcement learning using deep networks to represent both the policy and the reward function. They also propose a goal-conditioned risk-based feature representation for the social navigation problem that captures local information surrounding the agent.

3) Methods for Full-Body Motion Modeling and Prediction: Several works address the full-body motion modeling for collaborative robotics and physical human-robot interaction [item 8), 11), 13) in the Appendix]. Kratzer *et al.* [item 8) in the Appendix] present a dataset of full-body motion capture of people, performing everyday manipulation tasks in a home-like environment. The dataset includes 3D workspace geometry model and eye-gaze data. Abdelkawy *et al.* [item 3) in the Appendix] propose a Spatio-Temporal Joint based Convolutional Neural Network to recognize daily activities, combine them with the ambient events and infer the semantic context of the detected activity. Figueredo *et al.* [item 11) in the Appendix] propose a novel metric for human manipulation comfort, which combines ergonomics and muscular biomechanics. The proposed solution allow to build a quality distribution in the humans workspace which can be quickly tailored to specific tasks and filtered for design purposes.

4) Methods for Context-Rich Predictions: Two articles [item 2), and 14) in the Appendix] explore the new ways of integrating more contextual information in prediction systems. Minoura *et al.* [item 2) in the Appendix] propose a method for crowd density forecasting as an alternative to predicting the individual trajectories of people. To address this task, the authors developed a patch-based density forecasting network, which directly forecasts crowd density maps of future frames based on spatially or spatiotemporally overlapping patches. Rudenko *et al.* [item 14) in the Appendix] investigate the motion priors of walking people in semantically-rich urban environments using only semantic maps as input. An approach based on Convolutional Neural Networks is proposed to this end, which is capable of capturing the local semantic context to infer the probability of observing a person in any state of the environment.

B. Discussion and Conclusions

The papers in this special issue offer a diverse overview on many aspects of human motion prediction and human-awareness for robots and autonomous systems. Despite their diverse nature in specific tasks and application domains, these papers exhibit several dominant trends. For instance, generalization and active adaptation of models to new environments, often overlooked by authors in the prior art, is highlighted and addressed by several works in this special issue [item 1), and 14) in the Appendix]. Several works [item 5), 6), and 8) in the Appendix] studied the prediction of intention of people, which is not observed directly and must be inferred from inputs, an important cue to constraint the uncertainty for trajectory prediction. Several works focused on the key goal of human behavior understanding in robotics, namely the physical and psychological safety and comfort during of the people interacting and co-existing with the robot [item 9), and 11) in the Appendix]. Two works have ventured into co-habitation of pedestrians and vehicles, who need to share the same space and safely avoid collisions [item 4), 14) in the Appendix].

The problem of human motion prediction is quite challenging and still requires a lot of effort and novel ideas to handle the complexity of the dynamic world in which our autonomous systems work. The high-quality papers presented in the special issue represent important milestones towards solving the aforementioned problem, we hope that they can be of inspiration for novel works that will further advance this growing field.

Acknowledgment

This work has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant agreement No. 732737 (ILIAD).

Appendix

Related Work

1) T. Davchev, M. Burke, and S. Ramamoorthy, "Learning structured representations of spatial and interactive dynamics for trajectory prediction in crowded scenes,"

IEEE Robotics and Automation Letters, vol. 6, no. 2, pp. 707–714, 2021.

2) H. Minoura, R. Yonetani, M. Nishimura, and Y. Ushiku, "Crowd density forecasting by modeling patch-based dynamics," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 287–294, 2021.
3) H. Abdelkawy, N. Ayari, A. Chibani, Y. Amirat, and F. Attal, "Spatio-temporal convolutional networks and n-ary ontologies for human activity-aware robotic system," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 620–627, 2021.
4) C. Anderson, R. Vasudevan, and M. Johnson-Roberson, "Off the beaten sidewalk: Pedestrian prediction in shared spaces for autonomous vehicles," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 6892–6899, 2020.
5) A. Bajcsy, S. Bansal, E. Ratner, C. J. Tomlin, and A. D. Dragan, "A robust control framework for human motion prediction," *IEEE Robotics and Automation Letters*, vol. 6, no. 1, pp. 24–31, 2021.
6) Z. Huang, A. Hasan, K. Shin, R. Li, and K. Driggs-Campbell, "Long-term pedestrian trajectory prediction using mutable intention filter and warp lstm," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 542–549, 2021.
7) A. Ghadirzadeh, X. Chen, W. Yin, Z. Yi, M. Bjrkman, and D. Kragic, "Human-centered collaborative robots with deep reinforcement learning," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 566–571, 2021.
8) P. Kratzer, S. Bihlmaier, N. B. Midlagajni, R. Prakash, M. Toussaint, and J. Mainprice, "Mogaze: A dataset of full-body motions that includes workspace geometry and eye-gaze," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 367–373, 2021.
9) A. Konar, B. H. Baghi, and G. Dudek, "Learning goal conditioned socially compliant navigation from demonstration using risk-based features," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 651–658, 2021.
10) B. Ivanovic, K. Leung, E. Schmerling, and M. Pavone, "Multimodal deep generative models for trajectory prediction: A conditional variational autoencoder approach," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 295–302, 2021.
11) L. F. C. Figueredo, R. C. Aguiar, L. Chen, S. Chakrabarty, M. R. Dogar, and A. G. Cohn, "Human comfortability: Integrating ergonomics and muscular-informed metrics for manipulability analysis during human-robot collaboration," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 351–358, 2021.
12) D. Zhao and J. Oh, "Noticing motion patterns: A temporal cnn with a novel convolution operator for human trajectory prediction," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 628–634, 2021.
13) G. Habibi and J. P. How, "Human trajectory prediction using similarity-based multi-model fusion," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 715–722, 2021.
14) A. Rudenko, L. Palmieri, J. Doellinger, A. J. Lilienthal, and K. O. Arras, "Learning occupancy priors of human motion from semantic maps of urban environments," *IEEE Robotics and Automation Letters*, pp. 1–1, 2021.

Luigi Palmieri is a Senior Expert at Robert Bosch GmbH - Corporate Research. His research currently focuses on kinodynamic motion planning in dynamic and crowded environments, control of non linear dynamic systems, hybrid systems of learning-planning-control, MPC and numerical-optimization techniques, planning considering human motion predictions and social constraints. He earned his Ph.D. degree in robot motion planning from the University of Freiburg, Germany. During his Ph.D., he was responsible for the motion planning task of the EU FP7 project Spencer. Since then, he has the same responsibility in the EU H2020 project ILIAD. He has co-authored multiple papers at RA-L, ICRA, IROS, FSR on the combinations of motion planning with control, search, machine learning, and human motion prediction.

Andrey Rudenko is a Research Scientist at Robert Bosch GmbH - Corporate Research. He earned his Ph.D. at the University of Örebro (Sweden) in the field of human motion prediction. He has received the M.Sc. degree in robotics from the University of Freiburg. His research tackles predicting long-term motion trajectories of people in social and industrial settings, contributing to the EU FP7 project Spencer and the EU H2020 project ILIAD. He has co-authored multiple papers at RA-L, ICRA, IROS and RO-MAN, which tackled various aspects of anticipating human motion, human-aware path planning and human-robot interaction.

Jim Mainprice is leading a research group at the University of Stuttgart, Germany. His research interests include motion planning, machine learning and human-robot interaction. He holds a M.Sc. degree from Polytech'Montpellier, France, and a Ph.D. degree from the University of Toulouse, France, which he received in 2009 and 2012 respectively. While completing his Ph.D. at LAAS-CNRS, he took part in the EU FP7 projects Dexmart and Saphari. From January 2013 to December 2014, he was a Postdoctoral Researcher in the Autonomous Robotic Collaboration Lab at the Worcester Polytechnic Institute (WPI) located in Massachusetts, USA, where he participated in the DARPA Robotics Challenge as a member of the DRCHubo team. Since January 2015, he is affiliated with the Autonomous Motion Department (AMD) of the Max Planck Institute for Intelligent Systems in Tbingen, Germany, and he leads the Humans to Robots Motion (HRM) research group at the University of Stuttgart since April 2017.

Marc Hanheide is a Professor of Intelligent Robotics and Interactive Systems in the School of Computer Science at the University of Lincoln, U.K.. He received the Diploma and Ph.D. degrees (Dr.-Ing.) in computer science from Bielefeld University, Germany, in 2001 and 2006, respectively. From 2006 to 2009 he held a position as a Senior Researcher in the Applied Computer Science Group. From 2009 until 2011, he was a Research Fellow at the School of Computer Science at the University of Birmingham, U.K.. Marc Hanheide is a Principle Investigator in many national and international research projects, funded by H2020, EPSRC, InnovateUK, DFG, industry partners, and others. The STRANDS, ILIAD, RASberry, and NCNR projects are among the bigger projects he is involved with. In all his work, he researches on autonomous robots, human-robot interaction, interaction-enabling technologies, and system architectures. Marc Hanheide specifically focuses on aspects of long-term robotic behaviour and human-robot interaction and adaptation.

Alexandre Alahi is currently an Assistant Professor at EPFL. He spent five years at Stanford University as a Post-doc and Research Scientist after obtaining his Ph.D. from EPFL. His research enables machines to perceive the world and make decisions in the context of transportation problems and Smart Environments. He has worked on the theoretical challenges and practical applications of socially-aware Artificial intelligence, i.e., systems equipped with perception and social intelligence. He was awarded the Swiss NSF (Swiss National Science Foundation) early and advanced Researcher grants for his work on predicting human social behavior. He won the CVPR Open Source Award (2012) for his work on Retina-inspired image descriptors, and the ICDSC Challenge Prize (2009) for his sparsity-driven Algorithm that has tracked more than 100 million pedestrians to date.

Achim Lilienthal received the M.S. degree in physics from Konstanz University in 1998 and the Ph.D. degree in computer science Ö from University of Tbingen, Germany, in 2004. He is currently a Professor of computer science with Örebro University, Sweden, where he is also the Head of the Mobile Robotics Olfaction Laboratory, Center of Applied Autonomous Sensor Systems. His research interests include rich 3-D perception, robot vision, and mobile robot olfaction.

Kai O. Arras is the Head of the Robotics Program and Chief Expert in Robotics at Robert Bosch GmbH - Corporate Research. Prior to joining Bosch, he was an Assistant Professor at the University of Freiburg. He served as PC member for conferences on robotics (RSS, ISER, ECMR), human-robot interaction (HRI, ICSR, HAI), artificial intelligence (ICAPS, AAAI SA, ECAI), and computer vision (CVPR, ICCV), was Associate Editor of IJSR, ICRA, IROS, and authored over 100 peer-reviewed publications in those areas. He co-organized several workshops and conferences in different roles, and was a Member of the HRI Steering Committee and Co-ordinator of EU-FP7 project SPENCER.