2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA) | 978-1-6654-1867-6/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICAICA52286.2021.9498091

Concept of an Intuitive Human-Robot-Collaboration via Motion Tracking and Augmented Reality

Dario Luipers
Cologne Cobots Lab
TH Köln - University of Applied Sciences
Cologne, Germany
dario.luipers@th-koeln.de

Anja Richert
Cologne Cobots Lab
TH Köln - University of Applied Sciences
Cologne, Germany
Anja.richert@th-koeln.de

Abstract—**Human-cobot interaction is one of the main aspects of the 4th industrial revolution. One goal of current robotic research is to enhance safety and efficiency by designing the collaboration in a more intuitive way. The following work introduces two concepts to improve human-cobot collaboration on the basis of deep learning and augmented reality to achieve a more efficient and pleasant working environment. The first concept uses meta learning and Gaussian process to predict the movement of the human worker. The second concept enables the worker to see the next assembly step of the robotic arm via augmented reality.**

Keywords—deep learning, meta learning, motion prediction, Gaussian process, augmented reality, collaborative robotics

I. INTRODUCTION

The industry 4.0 era is rapidly changing the way we work. One main aspect of the changing working environment is the collaborative assembly of parts between human and robots. In this use case cobots are extremely relevant enabling an industry 4.0 production process [1]. To successfully integrate cobots into existing assembly processes the acceptance of the human workers towards the cobot is important. The concept of human factors sums up aspects to enhance the human-robot collaboration (HRC). [2] shows the varying and different effects of a cobot integration in the industry sector. This work concentrates on the cobot movement. Speed, predictability and trajectory of the cobot are important factors to increase the acceptance and reduce the stress for the human workers [3-4]. To overcome these challenges and improve the collaboration between cobots and humans this work presents two concepts. In the first concept the worker's upper body is tracked via RGB-D camera and its trajectory is predicted via machine learning to enable the cobot of moving intuitive and simultaneous with the human worker. In the second concept the next assembly step of the cobot is visualized for the worker via augmented reality (AR) to make the cobot's arm movement more predictable.

The prediction of the human movement affects the motion planning of the cobot to ensure a safe robotic arm trajectory. To improve the interaction of humans and cobots it is also important to design the cobot in a way, that the human worker perceives it as a coworking counterpart. This work proposes a concept to predict the final position of a human worker's hand via Gaussian process (GP), deep learning (DL) and meta learning (ML). The cobot is able to move its joints before a worker has finished the current task. This realizes a human-driven assembly in which the human can act freely and determine the exact position where the robot has to hand the human assembly parts and vice versa. A system with this ability aims to give the cobot a more intuitive behavior.

To make the movement of the cobot more predictable while using the shortest, most efficient path the second concept presents an approach to empower the human to see the next working step and the trajectory of the cobot arms. The representation of the next assembly steps of the cobot and its trajectory is done by using AR and HoloLens glasses (Microsoft, USA). The movement of the cobot becomes more predictable even if it takes the shortest path between two assembly points. This aims to make the assembly process more efficient while reducing the stress on the human worker. The two concepts presented in this paper are promising approaches to significantly improve the human-cobot-interaction.

II. RELATED WORK

Cobots are used to work closely with humans and can take over a variety of tasks [5-6]. The trajectory and the predictability of the cobot movement is a major aspect to realize a non-stressing working environment for humans [7-8]. The focus of recent work shows an increasing interest in the predictability of the human movement as well as in the visualization of the planned arm movements of the cobots.

A. Prediction of Human Movement using Machine Learning

In most cases, Machine Learning is used to predict the next action or trajectory of a human. A data-driven approach is the use of Gaussian mixture models and Gaussian mixture regression to predict the working area of the human worker for robot motion planning [9]. Another approach is the Gaussian process dynamical model (GPDM). In [10-11], this data driven technique is used to predict human walking motions. GPDM reduces the dimension of the input features to speed up the learning process of the model. The Gaussian Process (GP) regression in [12] is used to predict the next 100 ms of the human's hand trajectory for a human robot handover task with a root mean square error less than 1 cm. The hidden Markov model (HMM) is also a widely used machine learning model to calculate the probability of the future human movements. In [13], an Input Output HMM is used to generate a hierarchical structure to model the dynamics of human movements. Probabilistic movement primitives are used in [14] to predict the trajectory of a human to compute a robotic handover task. The learning parameters

θ are calculated by the Gaussian covariance matrix. Another machine learning approach for the prediction of human motions in the HRC are artificial neural networks (ANN). To realize a quick adapting forecast of human joint positions Cheng et al. [15] train an ANN offline and adapt the last layer of the network online via the recursive least square parameter adaptation algorithm. Recurrent Neural Networks (RNN) use time series information as input to predict future human movements. In [16], regular RNN and long short-term memory (LSTM) are used to classify the intention of a user and the collaborative and non-collaborative working steps. A more advanced approach is the implementation of ML to predict human motion for an intuitive HRC. Model Agnostic Meta Learning (MAML) [17] is one of the most used ML algorithms in research. Different ML models are used in [18] for few-shot learning of human motions on the H3.6M Dataset [19]. To boost the performance of the presented approaches this paper will present a concept to combine ML and GP regression.

B. Advanced Human Robot Collaboration due Augmented Reality

The combination of HRC and AR is another growing research field [20]. The visualization of the next robotic tasks and its trajectory are implemented via AR in [21]. As AR hardware the HoloLens glasses are used. The HoloLens is used in [22] as well to project the robot's picking location for a handover task. In addition to the visualization of the robot movement, AR can be used to project task specific parameters, like the force applied by a hand guided robot in a polishing task [23]. In the second concept these approaches will be combined and ergonomically advantageous positions for humans are implemented for collaborative tasks.

III. Concept for Motion Tracking and Prediction

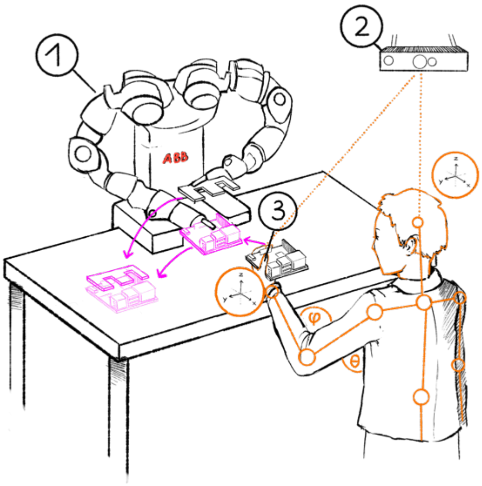

Fig. 1. Sketch of a collaborative assembly with human motion prediction. The predicted assembly part location for the handover, the future movement of the cobot and positions of the assembly parts are drawn in pink. The joints of the human are drawn in orange and are tracked to generate the input data for the machine learning model. 1: YuMi cobot (ABB, Switzerland), 2: Kinect camera (Microsoft, USA), 3: location of the human hand

The scenario presented in Fig. 1 is part of the collaborative assembly cell of the cologne cobots lab [25]. The human worker and the robot work collaboratively on the assembly of sensor cases. To realize an intuitive and natural human-robot interaction (HRI) some of the presented approaches in chapter II will be combined and extended. Like in [12], a GP model is trained to predict the position of the human hand for a given time horizon h. The input to the model is defined as:

$$X_s = (x_s; v_s; \varphi_s; \phi_s; t_s) \in \mathbb{R} \tag{1}$$

x_s, v_s are the position and velocity of the human's hand in the three dimensions of the cartesian coordinate system. φ_s, ϕ_s are the anlges between the upper body and the upper arm, and the upper arm and the underarm. t_s is the corresponding sampling time, where s is the control variable for each data sample. The goal of the model is to determine the final end position of the human hand Y_s to perform the handover task with the robot. This enables the cobot to move his joints simultaneously with the human which reduces the production time and implements an intuitive cobot behavior. The output of the model will be defined as:

$$Y_s = x_{s+h} \tag{2}$$

For simplicity, the training data for the GP model will be defined as $\hat{X}$ and $\hat{Y}$ and the test data will be written as X_* and Y_*. The prediction of Y_* by the assumption of a mean zero GP function is dependent on the covariance matrices K:

$$Y_* = \hat{K}_*\left(\hat{X}, X_*, \theta, \sigma_f\right)^T \left(\hat{K}\left(\hat{X}, \hat{X}, \theta, \sigma_f\right) + \sigma_n^2 I\right)^{-1} \hat{Y} \tag{3}$$

Despite the dependence on $\hat{X}$ and X_s the covariance matrices depend on the length-scale parameter θ, the signal variance σ_f and the noise variance σ_n. Because of the promising results in [12], the covariance matrices are calculated by a squared exponential kernel as follows:

$$\hat{K} = \sigma_f^2 \exp\left(-0.5\, \theta^{-2} \left\|\hat{X} - X_*\right\|^2\right) \tag{4}$$

As $\hat{X}$ and X_* are the input for training and testing of the model, θ, σ_f and σ_n are hyperparameters. The optimization of these hyperparameters is usually done by maximizing the log likelihood. In [25], ANN are used to determine θ and σ_n, which lead to promising results in the field of time series forecasting. The input to the ANN will be the same as for the GP model. The ANN will be a feed forward network with $m \in \mathbb{N}$ layers. The training and test input will be the same as for the GP model. Three separated ANN will be trained, one to predict the length-scale θ, one for the signal variance σ_f and one for the noise variance σ_n. The loss function for the ANN will be the inversed log likelihood of the GP model to use a minimizing algorithm. The most common and efficient optimizing algorithm for the neural networks is Adam [26]. The structure of the combination of ANN and GP is shown in Fig. 2. Human motions can be very varying for each individual. To ensure a quick adaption of the ANN, the network will be trained via ML. This will generate a model that can be adjusted for each user via few shot learning.

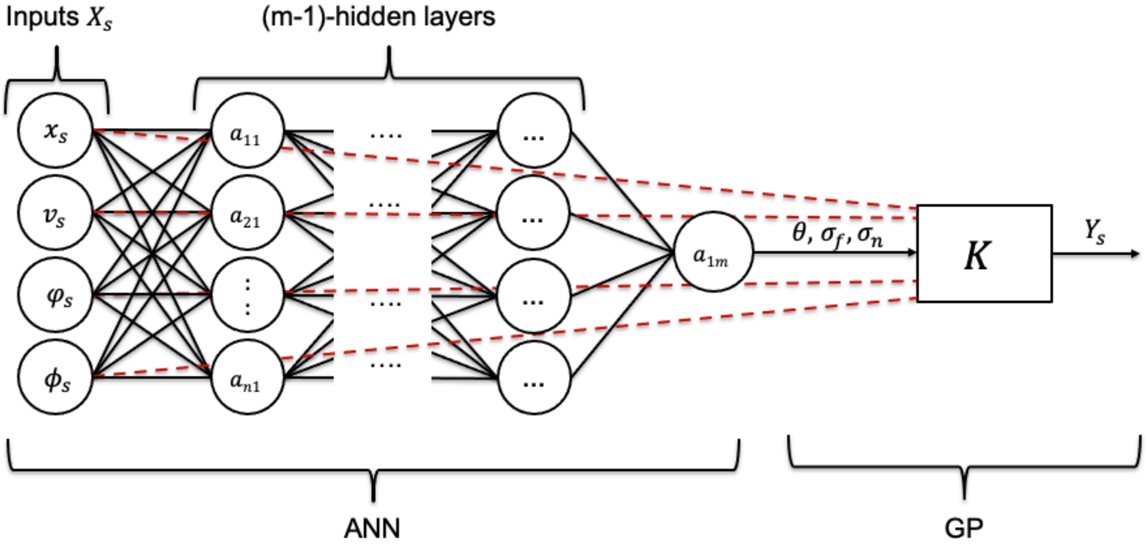

Fig. 2. Combination of ANN and GP. The feed forward network has a variable size of m layers and a variable size of n neurons a per hidden layer m. The input of the ANN is also the input for every covariance matrix K [25].

The process of ML is shown in Fig. 3. In future work, the number of sampling points per user for an adoption has to be determined. The aim is to realize a one-shot ML network with a prediction accuracy for Y_s of 95%. The used approach will be the MAML algorithm [17]. The main challenge for this ML approach will be the needed training data to train the ML network. To overcome this obstacle, open source data of human motions can be used [19] and synthetic data can be generated. Machine Learning also changed the simulation of human movements in video games. The work in [27] will be used to generate training data for the ML network.

The presented approach is designed to ensure a user adaptive model which combines GP, ANN, and ML. The aim is to have high accuracy for the prediction of the human's hand position at the time t_h. The prediction horizon h will be 0.5 seconds. This enables the system to destinate the user's handover location and the cobot can move its arm simultaneously with the human worker (Fig. 1). This makes the HRI more intuitive and faster.

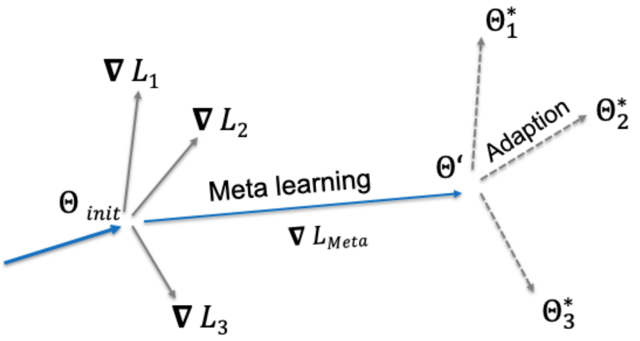

Fig. 3. Meta learning process. Θ_{init} is the initial solution space of the ML network. ∇L is the calculated loss of the ANN for a given task, in this case the prediction of Y_h. Θ' is the solution space of the ML network after the Meta learning phase. $\Theta^*$is the soluition of the specific task after the adaption of the ML network. After calculating the loss for the three specific tasks, in this case the location of the users hand after the time horizon h for three different users, the Meta loss is calculated. This leads to Θ', which is then adapted for the prediction of Y_h of a specific user (1, 2, or 3) [17].

The presented model has a number of advantages over classic DL models like ANN. Due to the GP model the probability of the prediction can be calculated and the interpretation of the hyperparameters σ_f and σ_n increase the transparency of this method. DL models are often black or grey box models which makes it difficult to reason about the model's data processing. Another advantage of the GP approach is the possibility to update the posterior distribution during the assembly by updating $\hat{X}$ and $\hat{Y}$ with live collected data of the human motion. After initializing the system, experiments will be run to validate if the intuitive and anticipating movement of the cobot has a positive effect on the human worker.

IV. Concept for Motion Vizualization and Assembly Status via Augmented Reality

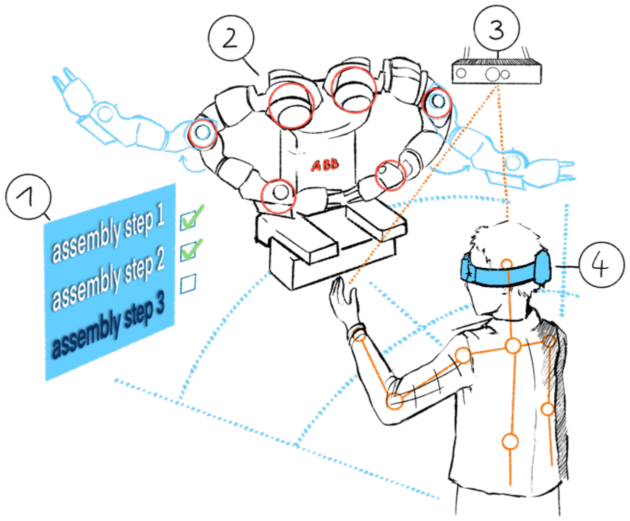

Fig. 4. Sketch of a collaborative assembly supported by AR. The AR features are drawn in blue. The next working steps of the cobot and the assembly status are visualized in AR via Unity (Unity Technologies, USA). The joints of the human are drawn in orange and are tracked to realize an ergonomic handover task. 1: finished and unfinished assembly steps, 2: YuMi cobot (ABB, Switzerland), 3: Kinect camera (Microsoft, USA), 4: HoloLens glasses (Microsoft, USA).

The second concept realizes a collaborative assembly that enables the human to see the next working steps of the cobot via AR. This addresses the problem described in section II of unpredictable cobot movement and trajectories. As shown in Fig. 4, the end position and the trajectory of the robotic arms will be visualized via AR for the human worker. This increases the transparency of the collaborative assembly. The following technical steps have to be implemented in the future system:

- Implementing the YuMi URDF files in Unity to visualize the arm movement in AR
- Detecting the joints of the robotic arm to project the arm movements
- Implementing a data pipeline from the cobot to the computing unit to calculate the future arm trajectories and assembly steps via Robot Operating System (ROS)
- Building a data pipeline from ROS to Unity to generate the virtual arm movements

Another feature is the display of the ordered finished and unfinished assembly steps of the current working process. This helps the human worker to track the progress of the production process. Further visualizations are also possible like the working speed of the cobot.

Human-cobot handover tasks profit in a variety of aspects from an AR support HRC. The human worker has a positive experience in the aspects of fluid interaction, trust in the robot, sense of safety, mental workload, and collaboration [22]. To improve the advantages of this system further, the handover position will be ergonomically individualized for the user. Because the Kinect RGB-D camera tracks the human position and its joints the system is able to calculate the ergonomically most advantageous position for the human-cobot handover. The technical concept of an ergonomic HRI has been developed by [28]. The visualized robotic hand moves to the user-specific ergonomically optimal position to perform a handover. This handover position reduces the pressure on the human joints and has a positive effect on the HRI. This concept also reduces the processing time, as the human can move to the visualized position without waiting for the cobot to reach its end position.

The assumed positive effects of this system have to be evaluated via experiments. The users will interact with the cobot while receiving different levels of AR support. The support varies from no AR features to all of the presented features. The User Experience is evaluated by means of a questionnaire.

V. Conclusion and Future Work

This article presents two concepts to develop an advanced HRI and increase the human acceptance of cobots. The first concept combines different Machine Learning approaches to predict the next state of the human motion. Due to the combination of GP model, ANN and ML, this system can be more accurate for a variety of different users. The ML network has the capability to quickly adapt its weights after a small sample size of task data. This allows the system to determine a user-specific prediction of the handover position after a small number of assembly runs. Due to the diversity of humans (sex, height, weight, etc.) the ML network approach seems very promising. The GP model has shown promising results in related works. One major advantage of this model is the transparency and interpretability in comparison to DL approaches. The uncertainty of the prediction can be calculated with the GP model as well as the variance of different input data. The GP model can also quickly adapt because of the updated posterior distribution with online data, which is collected during the assembly process. Recent research in the field of human motion simulation can be used to sample synthetic data. The synthetic datasets can be used to train the ML network. Further work regarding this concept aims to build a proof of concept and evaluate the User Experience of the system. After realizing the technical system, the following research questions will be answered:

- Is the combination of GP and ML more accurate for predicting human motion than the state of the art work presented?
- Which synthetic data points can be gathered from computer animations and what is the impact on the accuracy of the model?
- What is the minimum of real human motion data points needed to reach a prediction accuracy of 95%?
- Does the simultaneous movement of cobot and human to perform a handover task have a positive impact on the User Experience?
- Does the predictive motion of the cobot create a sense of intuitive behavior in the user?

The second presented concept showcases the advantages of an AR supported HRC. State of the art approaches are used to make the collaborative assembly more visual and give the human worker the ability to see the next working step of the cobot via AR and HoloLens glasses. The predictability of the cobot movement is increased which should reduce the stress of the human worker in a close and collaborative working environment. Another aspect of the concept is the improvement of the HRI ergonomically. The handover location visualized in AR is positioned at an ergonomically beneficial position to reduce the pressure in the human joints and to increase the User Experience. After realizing the technical system, the following research questions will be evaluated:

- Does the visualization of the cobot movement reduce the stress of the human worker and increase the trust in the system?
- Does the ergonomically optimized handover position visualized in AR have a noticeably positive effect for the user?

A combination of the two concept is also planned. This would enable the cobot in the second scenario to move his joints simultaneously with the human worker. The presented concepts will help to further increase the acceptance of close human-cobot-interactions and collaborations.

Acknowledgment

The authors acknowledge the financial support by the Ministry of Innovation, Science and Research of North Rhine-Westphalia Germany in the framework of Zukunftsfonds.

References

[1] M. Bortolini, E. Ferrari, M. Gamberi, F. Pilati, and M. Faccio, “Assembly system design in the Industry 4.0 era: a general framework,” *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 5700–5705, 2017.

[2] W. P. Neumann, S. Winkelhaus, E. H. Grosse, and C. H. Glock,

"Industry 4.0 and the human factor – A systems framework and analysis methodology for successful development," *International Journal of Production Economics*, vol. 233, p. 107992, 2021.

[3] A. Ajoudani, A. M. Zanchettin, S. Ivaldi, A. Albu-Schäffer, K. Kosuge, and O. Khatib, "Progress and prospects of the human–robot collaboration," *Autonomous Robots*, vol. 42, no. 5, pp. 957–975, 2017.

[4] G. Gulletta, E. C. Silva, W. Erlhagen, R. Meulenbroek, M. F. Costa, and E. Bicho, "A Human-like Upper-limb Motion Planner: Generating naturalistic movements for humanoid robots," *International Journal of Advanced Robotic Systems*, vol. 18, no. 2, p. 172988142199858, 2021.

[5] C. A. Moore, M. A. Peshkin, and J. E. Colgate, "Cobot implementation of virtual paths and 3-D virtual surfaces," *IEEE Transactions on Robotics and Automation*, vol. 19, no. 2, pp. 347–351, 2003.

[6] S. El Zaatari, M. Marei, W. Li, and Z. Usman, "Cobot programming for collaborative industrial tasks: An overview," *Robotics and Autonomous Systems*, vol. 116, pp. 162–180, 2019.

[7] Á. Castro-González, H. Admoni, and B. Scassellati, "Effects of form and motion on judgments of social robots′ animacy, likability, trustworthiness and unpleasantness," *International Journal of Human-Computer Studies*, vol. 90, pp. 27–38, 2016.

[8] M. Koppenborg, P. Nickel, B. Naber, A. Lungfiel, and M. Huelke, "Effects of movement speed and predictability in human-robot collaboration," *Human Factors and Ergonomics in Manufacturing & Service Industries*, vol. 27, no. 4, pp. 197–209, 2017.

[9] J. Mainprice and D. Berenson, "Human-robot collaborative manipulation planning using early prediction of human motion," *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013.

[10] J. M. Wang, D. J. Fleet, and A. Hertzmann, "Gaussian Process Dynamical Models for Human Motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 283–298, 2008.

[11] R. Quintero, J. Almeida, D. F. Llorca, and M. A. Sotelo, "Pedestrian path prediction using body language traits," *2014 IEEE Intelligent Vehicles Symposium Proceedings*, 2014.

[12] M. Wu, B. Taetz, E. Dickel Saraiva, G. Bleser, and S. Liu, "On-line Motion Prediction and Adaptive Control in Human-Robot Handover Tasks," *2019 IEEE International Conference on Advanced Robotics and its Social Impacts (ARSO)*, 2019.

[13] Z. Wang, P. Jensfelt, and J. Folkesson, "Multi-scale conditional transition map: Modeling spatial-temporal dynamics of human movements with local and long-term correlations," *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015.

[14] G. J. Maeda, G. Neumann, M. Ewerton, R. Lioutikov, O. Kroemer, and J. Peters, "Probabilistic movement primitives for coordination of multiple human–robot collaborative tasks," *Autonomous Robots*, vol. 41, no. 3, pp. 593–612, 2017.

[15] Y. Cheng, W. Zhao, C. Liu, and M. Tomizuka, "Human Motion Prediction using Semi-adaptable Neural Networks," *2019 American Control Conference (ACC)*, 2019.

[16] D. Nicolisl, A. M. Zanchettin, and P. Rocco, "Human Intention Estimation based on Neural Networks for Enhanced Collaboration with Robots," *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018.

[17] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in Proc. 34th Int. Conf. Mach. Learn., vol. 70, Aug. 2017, pp. 1126–1135.

[18] L.-Y. Gui, Y.-X. Wang, D. Ramanan, and J. M. Moura, "Few-Shot Human Motion Prediction via Meta-learning," *Computer Vision – ECCV 2018*, pp. 441–459, 2018.

[19] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu. Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(7):1325–1339, 2014.

[20] Z. Makhataeva and H. Varol, "Augmented Reality for Robotics: A Review," *Robotics*, vol. 9, no. 2, p. 21, 2020.

[21] U. Gruenefeld, L. Prädel, J. Illing, T. Stratmann, S. Drolshagen, and M. Pfingsthorn, "Gruenefeld, U., Prädel, L., Illing, J., Stratmann, T., Drolshagen, S. and Pfingsthorn, M., 2020, September. "Mind the ARm: realtime visualization of robot motion intent in head-mounted augmented reality," *Proceedings of the Conference on Mensch und Computer*, 2020.

[22] R. Newbury, A. Cosgun, T. Crowley-Davis, W. P. Chan, T. Drummond, and E. Croft, "Visualizing Robot Intent for Object Handovers with Augmented Reality," *arXiv preprint arXiv:2103.04055*, 2021.

[23] A. De Franco, E. Lamon, P. Balatti, E. De Momi, and A. Ajoudani, "An Intuitive Augmented Reality Interface for Task Scheduling, Monitoring, and Work Performance Improvement in Human-Robot Collaboration," *2019 IEEE International Work Conference on Bioinspired Intelligence (IWOBI)*, 2019.

[24] C. Neef, D. Luipers, J. Bollenbacher, C. Gebel, and A. Richert, "Towards Intelligent Pick and Place Assembly of Individualized Products Using Reinforcement Learning," *Human Systems Engineering and Design III*, pp. 325–331, 2020.

[25] K. Cremanns and D. Roos. "Deep Gaussian covariance network." *arXiv preprint arXiv:1710.06202* (2017).

[26] D. P. Kingma and J. Ba. "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980*, 2014.

[27] S. Starke, Y. Zhao, T. Komura, and K. Zaman, "Local motion phases for learning multi-contact character movements," *ACM Transactions on Graphics*, vol. 39, no. 4, 2020.

[28] W. Kim, M. Lorenzini, P. Balatti, P. D. H. Nguyen, U. Pattacini, V. Tikhanoff, L. Peternel, C. Fantacci, L. Natale, G. Metta, and A. Ajoudani, "Adaptable Workstations for Human-Robot Collaboration: A Reconfigurable Framework for Improving Worker Ergonomics and Productivity," *IEEE Robotics & Automation Magazine*, vol. 26, no. 3, pp. 14–26, 2019.