https://doi.org/10.2352/ISSN.2470-1173.2021.6.IRIACV-310

Towards Large-Scale Evaluation of Mental Stress and Biomechanical Strain in Manufacturing Environments Using 3D-Referenced Gaze and Wearable-based Analytics

Lucas Paletta, Harald Ganster, Michael Schneeberger, Martin Pszeida, Gerald Lodron; JOANNEUM RESEARCH, Graz, Austria
Katrin Pechstädt, Michaela Spitzer, Christiane Reischl; FH JOANNEUM, Bad Gleichenberg / Graz, Austria

Abstract

In future manufacturing human-machine interaction will evolve towards flexible and smart collaboration. It will meet requirements from the optimization of assembly processes as well as from motivated and skilled human behavior. Recently, human factors engineering has substantially progressed by means of detailed task analysis. However, there is still a lack in precise measuring cognitive and sensorimotor patterns for the analysis of long-term mental and physical strain. This work presents a novel methodology that enables real-time measurement of cognitive load based on executive function analyses as well as biomechanical strain from non-obtrusive wearable sensors. The methodology works on 3D information recovery of the working cell using a precise stereo measurement device. The worker is equipped with eye tracking glasses and a set of wearable accelerometers. Wireless connectivity transmits the sensor-based data to a nearby PC for monitoring. Data analytics then recovers the 3D geometry of gaze and viewing frustum within the working cell and furthermore extracts the worker's task switching rate as well as a skeleton-based approximation of worker's posture associated with an estimation of biomechanical strain of muscles and joints. First results enhanced by AI-based estimators demonstrate a good match with the results of an activity analysis performed by occupational therapists.

Introduction

The production industry is currently in a process of continuous transition to the Fourth Industrial Revolution fostering 'smart factories' with modular structured entities and cyber-physical systems monitoring physical processes, creating a virtual copy of the physical world and make decentralized decisions. Recently, human factors engineering has substantially progressed by means of detailed task analysis. However, the extraction of task descriptions still relies on manual analysis and elaborated, time-consuming description of video-based monitoring of the human-machine interaction. In this manner, long-term observation and analysis of work is too costly which implies that there is still a lack in precise measurement of cognitive and sensorimotor patterns for the analysis of long-term mental and physical strain.

From the viewpoint of quality-of-service, caring for attention can play a major role for gain or loss of productivity. Attention metrics transfer into measurements of concentration, cognitive load, situation awareness but also to early indication of fatigue. Quality of attention is directly related to quality of decision making and therefore the related quality-of-service - in terms of increased fatigue for example - indicates less safer decisions due to reduced attention spans, decreased reaction time as well as accuracy. It is well known that undetected and therefore frequently occurring fatigue means more team misunderstanding, decrease of mood and motivation and- due to related physical ailments - consequently short-term and long-term absence from work and increased medical investments. Consequently, quality of attention relates to avoiding loss of productivity, due to distractions, errors, inability to concentrate and lack of motivation.

Our work aims at the challenging long-term objective to enable non-obtrusive long-term analytics of mental, physical, emotional and motivational stress of workers at the manufacturing site. We argue that pure observation based analysis cannot fully interpret the psychophysical processes that reflect psychosomatic stress over long time periods. The technical challenge of this work is to measure both the local infrastructure that determines geometric and functional conditions on the worker's interaction as well as the psychophysiological and biomechanical parameters that describe the human state, intention and activity.

This work presents a novel methodology that enables real-time measurement of cognitive load as well as biomechanical strain from non-obtrusive wearable sensors. In particular, cognitive load is calculated from the analysis of eye movements in reference with the semantics of the recovered 3D structure, such as, task-specific areas-of-interest that are typically viewed at during task execution. From these observations we generate some indicative 'task switching'-related measures which are characteristic for attention-based executive function processes. Finally, these measures enable to provide a better approximation of human states, performance and strain parameters than just external observation would enable.

***Figure 1**. Wearable-based data acquisition towards large-scale evaluation of mental load and biomechanical strain at a factory workplace. A pick-and-place workplace demonstrates the challenging requirements on cognitive and biomechanical strain.*

Related Work

In the field of Human-machine interaction, human-centered variables have played a major role since quite a while, parameters - such as, situation awareness, workload and the mental model - were described by Steinfeld et al. [1] for the example of human-robot interaction. These human-related variables are essential for the evaluation of human-interaction metrics. Human factors are crucial as industrial robots are enabling human and robot workers to work side by side as collaborators and to assess the user's experience with a robot, while understanding how humans feel during their interaction with it (Figure 1). Attention and gaze provide a means of action prediction and intention as outlined by Huang and Mutlu [2], i.e., to work seamlessly and efficiently with their human counterparts. Machines must similarly rely on predictions of the human worker's behavior, emotions, task specific actions and intent to plan their actions, such as, in anticipatory control with human-in-the-loop architecture [2] to enable robots to proactively perform task actions based on observed gaze patterns to anticipate actions of their human partners according to its predictions.

Measuring and modeling of human attention, concentration and situation awareness based on eye movements and accordingly triggered information recovery is mandatory for the understanding of human action planning, mental workload and, consequently, the achieved quality of service. Gaze as input device has proved to be beneficial, such as, in shop floor management [3]. Finally, the Carnegie Mellon University with Admoni et al. [4] has introduced intuitive gaze-based cyber-physical control as additional asset. There seems to be a wide spectrum of opportunities for gaze to provide information for monitoring, assessment, as interaction device and for action prediction.

The most related work to the presented method is Santner et al. [5] on the 3D gaze recovery by use eye tracking glasses and SLAM methods for 3D information recovery. The presented method builds on the results of that work but extends substantially, applying a more developed handheld device for better infrastructure recovery, as well as adding a complete methodology to extract information about executive functions, and relating to biomechanical strain estimation for a comprehensive evaluation of strain in industrial working cells with human-machine interaction.

Human Factors Integrated Measuring System

Rationale and Data Processing

Overall the methodology consists of two components, (i) the estimation of mental load, and (ii) the estimation of biomechanical strain. The estimation of mental load firstly applies the extraction of working cell infrastructure in terms of its 3D geometry including a computer vision processing stage of 3D information recovery of the working cell using a precise stereo measurement device. A next step requires attention sampling from the focused interaction of the worker with the infrastructure. For this purpose, the worker is equipped with eye tracking glasses as well as with a set of wearable accelerometers with its sensor data transmitted via wireless connectivity to a nearby PC for monitoring.

Data analytics registers the egocentric video frames from the eye tracking glasses with the visual information in the scene (e.g., artificial landmarks) and thereby recovers the 3D geometry of gaze and viewing frustum towards the working cell. A final step relates the semantics of attended areas-of-interest with task references and from this is able to conclude about an estimation on the worker's task switching rate and, consequently, mental work load.

These figures will in the future be part of a more complete spectrum of cognitive, in principle, mental parameters (including affective states) that together will enable - on the basis of more populated and elaborated studies – to develop towards the estimation of long-term generated general definition of human mental state, including stress.

In total, both methods, mental and biomechanical strain, represent together the main source of psychophysiological stress in working cells and will underlie future studies in real working environments.

Development of Digital Twin of Work Cell

An important prerequisite for the computation of mental workload is the definition of work zones that represent regions-of-interest (ROI) in the work cell environment the worker is interacting with during particular tasks. For example, during assembly of a specific part of the product interaction is applied towards a specific part of the environment, such as, a table. The rationale of computation of concentration is that workers are focused while heading towards these environments, being occupied with objects of interest that in this work are not yet considered, this will be done in later work. However, the 3D model serves to define the task interaction zones, but also the overall visualization of the work environment together with the superimposed interaction and related attention behavior.

Figure 2. *Development of a digital twin of the work cell environment and deriving eye movements with respect to the environment by application of eye tracking glasses, image recognition and matching methodology.*

The concrete first task is to extract a 3D model from the real work environment (Figure 2). For this purpose, we applied a mobile, practical stereo system that combines highly accurate and robust projected texture stereo and efficient volumetric integration and allows to easily capture accurate 3D models of indoor scenes [6]. Its methodology optimizes a stereo method for random dot projection patterns and delivers complete and robust results. The hardware is enclosed in a box and contains three Basler dart cameras (2 monochrome cameras for stereo, one for RGB) and active Kinect projector to apply reconstruction via occupancy grid and iterative closest points (ICP) based registration [7]. A typical result is displayed in Figure 3.

Estimation of Mental Load

In order to estimate the actual cognitive load from the worker's interaction with the environment we decided to use eye tracking glasses that continuously determine the eye movements over time. In a previous work we developed a methodology to estimate concentration and mental workload from eye tracking glasses-based data in a factory-like lab environment [8]. Eye tracking glasses generate a continuous video stream of ca. 30 Hz and eye gaze positions within the video frames.

In order to position the extracted gaze with the 3D environment we had firstly to position the video with the environment and then estimate the gaze ray and its geometric recovery with respect to the 3D environment and determine to which ROI it belongs. In order to match the video frames with the 3D model we placed ArUco markers (OpenCV toolbox [9]) within the work cell where we expected human gaze. An alternative method is to base the matching process not on fiducial markers but directly on video-based measurements from the scene [10]. Eye tracking was applyed to record gaze behavior with the purpose to analyze cognitive human factors, such as,

- Gaze behavior in terms of the orientation of the viewing axes of both eyes over time (with 60 Hz) relative to the work environment,
- Indicative estimation of mental stress [11],
- Indicative estimation of concentration and sustained attention with regard to the task [8] (executive function),
- Situation awareness by estimating the reference of gaze to important events [12],
- Task-switching indicating the load of switching attention between mental models [8] (executive function).

Estimation of Biomechanical Strain

For the estimation of biomechanical strain we first conducted an analysis of human posture and from this intended to conclude about general aspects of physical strain. We applied non-obtrusive wearables to the worker (Figure 5) in order to estimate relative positions and further derive the overall worker pose in space, finally, to receive a skeleton-based representation for further processing.

The choice of wearable were Perception Neurons that represent small, adaptive, versatile and affordable motion capture technology. The modular system is based on the NEURON, an IMU (Inertial Measurement Unit) composed of a 3-axis gyroscope, 3-axis accelerometer and 3-axis magnetometer. The system applies proprietary embedded data fusion, human body dynamics and physical engine algorithms that deliver smooth motion with minimal latency. The PERCEPTION NEURON 9-Axis sensor units output data at 60fps or 120fps. The data stream is channelled to a hub where it can then be transmitted to a computer in three different ways: (1) via WIFI, (2) via USB or (3) recorded on-board using the built-in micro-SD slot. The modular system on a Notebook finally applies a skeleton-based representation.

For the analysis of the skeleton-based representation we applied the biomechanical software toolbox BoB (Biomechanics of Bodies). BoB represents a software package that contains a musculo-skeletal model of a digital human [13].

During the import task BoB reads the motion, force and skeleton files. The skeleton file is applied to construct the mechanism containing the joints that are articulated with the joint angles defined in the motion file. The forces in the force file are applied to the mechanism as external forces. BoB then performs an inverse dynamics analysis of the mechanism to calculate the torques at the joints that are required to compute the joint motions in the presence of the external forces and mass model. On the basis of the skeleton-based representation BoB is capable to provide personalized estimates for characteristic biomechanical strain features, such as, muscle forces, joint torques, joint contact forces.

(a)

(b)

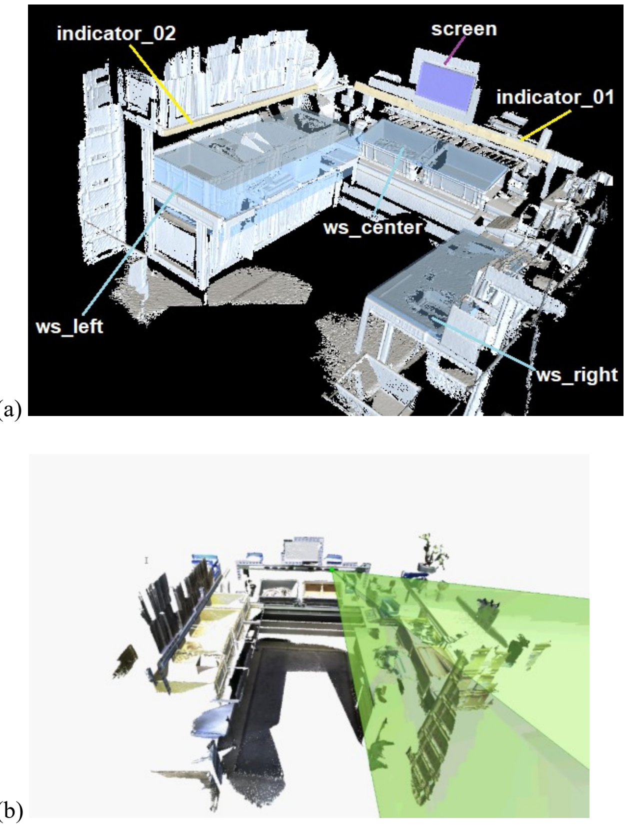

Figure 3. *Result of the 3D information recovery process. (a) Resulting 3D model with regions-of-interest. (b) 3D gaze recovery with the focus of the camera (green node) and the estimated view frustum (green pyramid) together with the 3D infrastructure.*

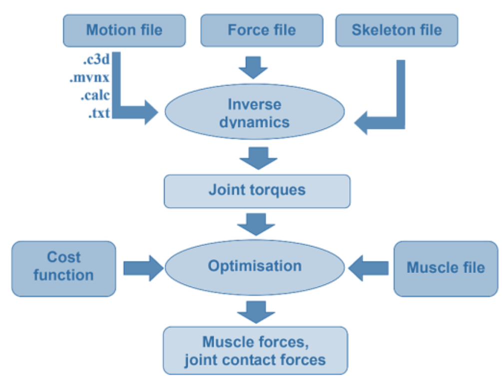

Figure 4. *Data flow from the received motion, force and skeleton-based information, via inverse dynamics and optimization processes that finally lead to the resulting muscle as well as joint contact forces.*

Figure 4 depicts the data flow from the received motion, force and skeleton files, and that an optimization process finally leads to the resulting muscle as well as joint contact forces.

There is no unique solution for the muscle force distribution that is required to generate the joint torques as there are typically approximately 40 joint torques to satisfy with over 600 muscles in the muscle file. Therefore, an optimization approach of a cost function is utilized to identify one of the infinite number of possible solutions. By default BoB minimizes the sum of the square of the muscles' activations where the muscle activations are defined as the instantaneous muscle force divided by the muscle maximum isometric force. This cost function has been shown to correspond to reducing the fatigue of the body. The user can select other cost functions.

Experimental Results

First results were captured from an assembly-oriented working cell where frequent pick-up processes with change of orientation and pose characterize the interaction patterns. The Medical University of Graz granted Ethical approval (No. 31-243 ex 18/19). A typical female worker with 51 years of age was wearing the equipment for a duration of 11 minutes, i.e., 660 seconds (Figure 5). During this time, 15.840 video frames were captured of which ≈27% were referenced in a fully automated way in association with ROIs of the work environment's infrastructure. The remainder of the video frames was manually attributed with the ROIs in sight within the egocentric video frames. A heuristic concentration indicator (Paletta et al. [8]) was then computed on the basis of the gaze-referenced ROI data. This resulted in a mean (standard deviation) of the concentration level M=3.79 (SD=1.39) where the level value range is between 1 and 5 (Figure 6). Within 11 minutes of work, 153 task switches were calculated with a mean (SD) of 4.64 (2.78) switches and some maximum value of ca. 10.0 in a 20 seconds interval (Figure 7). These results demonstrate a very high task switching rate compared to normally reported functioning [14].

In a final step, we associated muscular forces of predetermined cluster of muscles with 'occupational strain categories' given evaluations of the muscle analysis [15] provided by occupational therapists. A support vector machine neural network (Vapnik, 1995) is finally trained to map from muscular forces to occupational strain. A first neural network implementation resulted in 89% accuracy in the prediction for a 1-second time window.

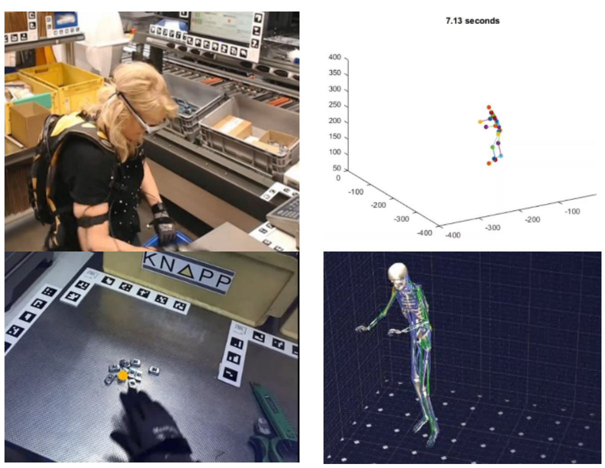

Figure 5. Synchronized videos: (left top) raw video surveillance of the assembly work with wearables, (top right) skeleton estimation, (bottom left) gaze video, (bottom right) visualization of the muscular force (green, yellow, red for low, mid and high muscular strain) and correct posture.

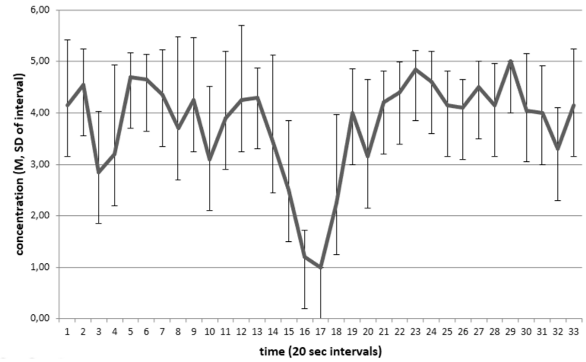

Figure 6. Heuristic concentration level (from 1 (low) to 5 (high) according to Paletta et al. [8], calculated for the pick-up task and an intermediate view on task-irrelevant issues (center).

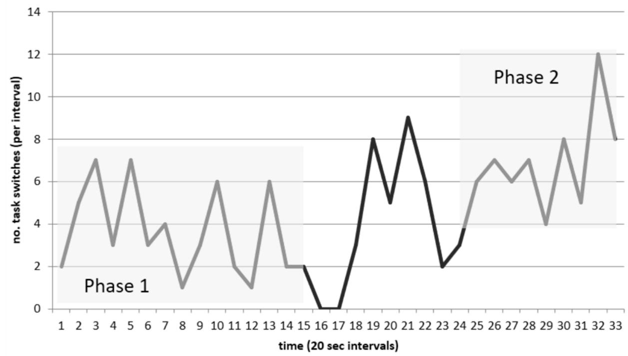

Figure 7. Task switches (calculated according to Paletta et al. [8]), reflecting the high cognitive load in phase 2.

Discussion, Conclusions and Future Work

This experiment demonstrates that a light-weight non-obtrusive wearable equipment consisting of eye tracking glasses and motion capture sensors can be used to derive mental load and biomechanical strain. High concentration and high task switching rates of the worker were reflected in the heuristic scores extracted from the gaze data. From the extracted quantities one can easily derive cognitive workload overcharges, for example, workers are supposed not to work more than 20-30 minutes with high concentration in a row, otherwise fatigue will increasingly take place which in turn highly increases risk factors at work.

Furthermore, the detailed biomechanical strain information was mapped to occupational experts' scores and reflected as well to a high degree the evaluation of the experts.

Future work will focus on the application of more complex eye movement features, studies with larger populations and we are already introducing bio-sensing for more fundamental treatment of psychophysiological analytics underlying human factors and ergonomics in manufacturing environments.

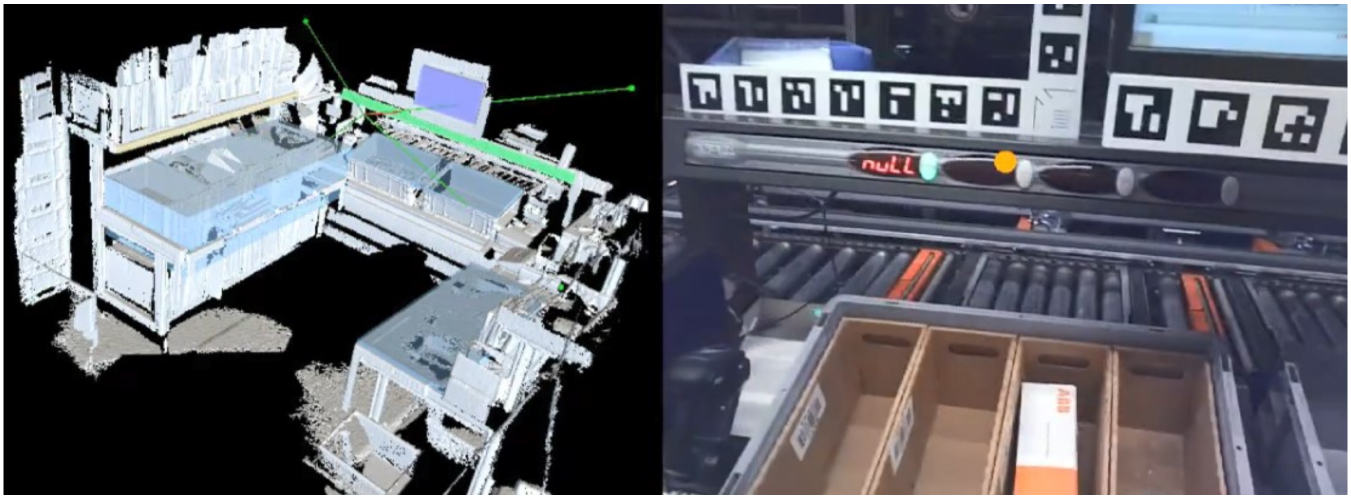

Figure 8. *Recovery of 3D gaze geometry (left) within the 3D model of the work environment and egocentric video frame from the eye tracking glasses with gaze point (orange) superimposed.*

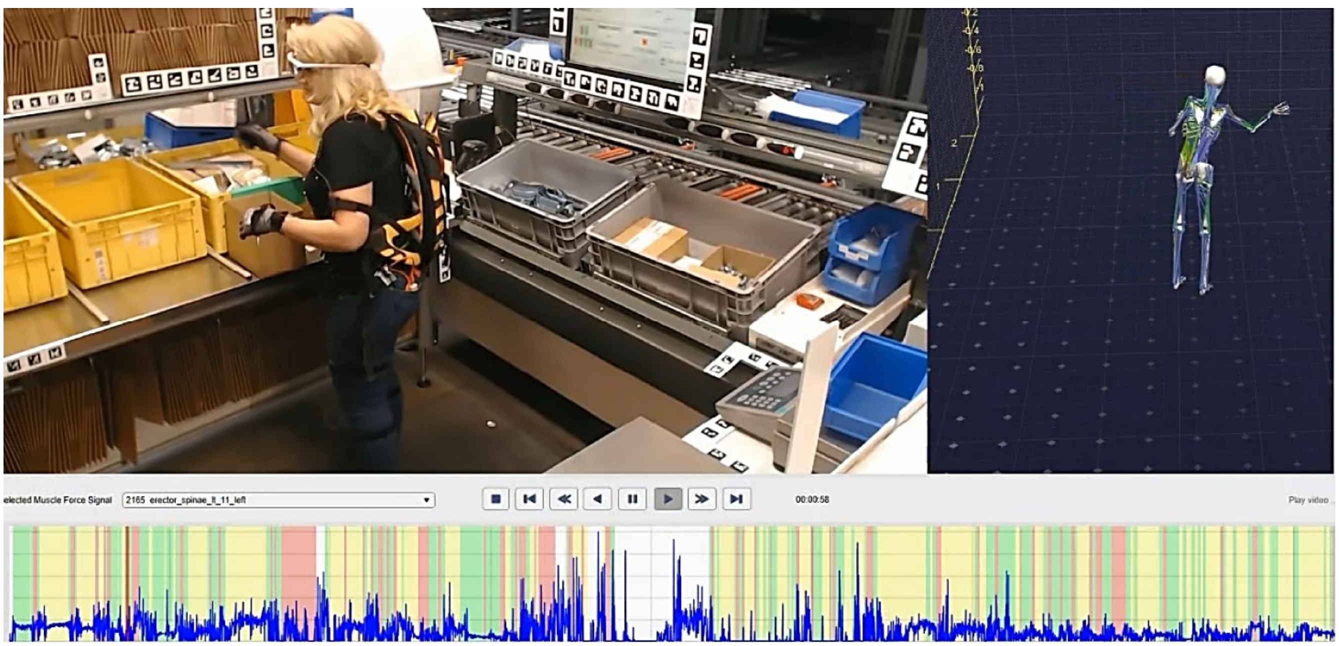

Figure 9. *Work in a modern factory representing pick-up processes in an assembly working cell (top left). The worker is equipped with eye tracking glasses, wearables in terms of gyroscopes/accelerometers/magnetometers in gloves and at extremities as well as with a backpack containing a notebook. From the wearables we compute a skeleton-based representation of human pose (right). The force on each muscle is quantified from an estimation model derived from the current pose and plotted over time (blue, below). Furthermore, we have a ground truth-like annotation of occupational therapists describing low (green), medium (yellow) and high (red) degree of challenging the muscle functions by rating video-recorded sequences using an occupational-based activity analysis [15].*

Acknowledgments

This work has been supported by the Styrian Future Fund ("Zukunftsfonds Steiermark“, project INCLUDE, no. 1036) and by the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology (BMK) within projects MMASSIST (no. 858623) and FLEXIFF (no. 861264). Special thanks go to Mr. Wolfgang Silly (KNAPP AG) for the great support in the project INCLUDE.

References

[1] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz and M. Goodrich, "Common metrics for human-robot interaction," in *Proc. Human-robot interaction, ACM SIGCHI/SIGART*, 2006.

[2] C.-M. Huang and B. Mutlu, "Anticipatory robot control for efficient human-robot collaboration," in *Proc. ACM/IEEE HRI 2016*, 2016.

[3] K. Fischer, L. Jensen, F. Kirstein, S. Stabinger, Ö. Erkent, D. Shukla and J. Piater, "The Effects of Social Gaze in Human-Robot

Collaborative Assembly," in *Tapus A., André E., Martin JC., Ferland F., Ammi M. (eds), Social Robotics. ICSR 2015*, Lecture Notes in Computer Science, vol 9388. Springer, Cham. https://doi.org/10.1007/978-3-319-25554-5_21, 2015.

[4] H. Admoni, A. Dragan, S. Srinivasa and B. Scassellati, "Deliberate delays during robot-to-human handovers improve compliance with gaze communication.," in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 49*, 2014.

[5] K. Santner, G. Fritz, L. Paletta and H. Mayer, "Visual recovery of saliency maps from human attention in 3D environments," in *Proc. Intrnational Conference on Robotics and Automation (ICRA)*, pp. 4297-4303, 2013.

[6] M. Klopschitz, R. Perko, G. Lodron, G. Paar and H. Mayer, "Projected Texture Fusion," in *10th International Symposium on Image and Signal Processing and Analysis (ISPA 2017)*, Ljubljana, Slovenia, September 18-20, 2017,.

[7] R. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. Fitzgibbon, A. Davison, P. Kohli, J. Shotton and S. Hodges, "KinectFusion: Real-Time Dense Surface Mapping and Tracking," in *10th IEEE International Symposiumon on Mixed and Augmented Reality (ISMAR 2011)*, Oct. 2011..

[8] L. Paletta, M. Pszeida, H. Ganster, F. Fuhrmann, W. Weiss, S. Ladstätter, A. Dini, S. Murg, H. Mayer, I. Brijacak and B. Reiterer, "Gaze based Human Factors Measurements for the Evaluation of Intuitive Human-Robot Collaboration in Real-time," in *Proc. 24th IEEE Conference on Emerging Technologies and Factory Automation, ETFA 2019*, Zaragoza, Spain, 2019.

[9] S. Garrido-Jurado, F. Madrid-Cuevas, R. Muñoz-Salinas and M. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition,* pp. 47, 6 (June 2014), 2280-2292. doi=10.1016/j.patcog.2014.01.0, June 2014.

[10] L. Paletta, K. Santner, G. Fritz, A. Hofmann, G. Lodron, G. Thallinger and H. Mayer, "FACTS - A Computer Vision System for 3D Recovery and Semantic Mapping of Human Factors," in *Proc. 9th International Conference on Computer Vision Systems, ICVS 2013*, LNCS 7963, pp. 62-72, Springer-Verlag Berlin Heidelberg, Sankt Petersburg, Russia, July 16-18, 2013, 2013.

[11] L. Paletta, M. Pszeida, B. Nauschnegg, T. Haspl and R. Marton, "Stress Measurement in Multi-tasking Decision Processes Using Executive Functions Analysis," in *Ayaz H. (eds) AHFE 2019. Advances in Neuroergonomics and Cognitive Engineering. Advances in Intelligent Systems and Computing*, Springer, 2019, pp. vol 953, pp. 344-356.

[12] A. Dini, C. Murko, L. Paletta, S. Yahyanejad, U. Augsdörfer and M. Hofbaur, "Measurement and Prediction of Situation Awareness in Human-Robot Interaction based on a Framework of Probabilistic Attention," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vancouver, Canada, 2017.

[13] J. Shippen and B. May, "BoB – biomechanics in MATLAB," in *Proceedings of 11th International Conference BIOMDLORE*, 2016.

[14] M. Steyvers, G. E. Hawkins, F. Karayanidis and S. D. Brown, "A large-scale analysis of task switching practice effects across the lifespan," *PNAS ,* pp. 17735-17740, September 3, 2019 116 (36) .

[15] H. Thomas, Occupation-based activity analysis, Thorofare, NJ: SLACK Incorporated., 2015.

Author Biography

Lucas Paletta received his PhD in Computer Science from Graz University of Technology (2000). Since then he has worked at the JOANNEUM RESEARCH (Institute DIGITAL), Graz, Austria, where he is a key researcher and a research manager. He has been leading the Human Factors Lab since 2017.

Harald Ganster received his PhD in Mathematics from Graz University of Technology (1999). He has been working as a key researcher in the area of machine vision at the JOANNEUM RESEARCH (Institute DIGITAL), Graz, Austria since 2005.

Michael Schneeberger received his MSc (DI) in Telematics from Graz University of Technology (2018). He has been working as researcher in the field of machine vision and bio-sensing at JOANNEUM RESEARCH (Institute DIGITAL), Graz, Austria, since 2010.

Martin Pszeida received his BSc in Telematics from Graz University of Technology (2014). He has worked at the JOANNEUM RESEARCH (Institute DIGITAL), Graz, Austria since 2012, where he is a researcher in AI and eye tracking analysis.

Gerald Lodron received his DI in Health Care IT from Carinthia University of Applied Sciences (2009). Since 2009 he has worked at the JOANNEUM RESEARCH (Institute DIGITAL), Graz, Austria, where he is a senior researcher in machine vision.

Katrin Pechstädt has been a state-certified occupational therapist since 2006 and received her MSc in Occupational Therapy in 2019. She has been working at FH JOANNEUM – University of Applied Sciences as a research assistant (2017-2018) and as a lecturer (since 2018)

Michaela Spitzer has received her BSc in Occupational Therapy in 2015 and her MSc in Occupational Therapy and Occupational Science in 2020 and has been working at FH JOANNEUM – University of Applied Sciences at the Institute of Occupational Therapy since 2019, where she is a research assistant.

Christiane Reischl has received her MA in Sociology and her MSc in Environmental Systems Sciences at the University of Graz and has been working at FH JOANNEUM – University of Applied Sciences at the Institute of Social Work since 2018, where she is a research assistant and lecturer.