Erschienen in: Mensch und Computer 2017 : Tagungsband / Burghardt, Manuel; Wimmer, Raphael; Wolff, Christian; Womser-Hacker, Christa (Hrsg.). - Regensburg : Gesellschaft für Informatik e.V., 2018. - S. 351-356. - ISBN 978-3-7460-8210-3
https://dx.doi.org/10.18420/muc2017-ws08-0311

Human-Machine Interaction in Care-Education

Johannes Reichold[1], Ankita Agrawal[4], Marieke Thurlings[1], Iris Cohen[1], Barbara Weber-Fiori[4], Anita Rölle[4], Muneeb Hassan[4], Maximilian Dürr[3], Ulrike Pfeil[3], Anna-Antonia Pape[1], Gerd Grünert[1], Artur Schmidt[2], Markus Pfeil[1], Victor Fäßler[1], Volker Jauch[2], Harald Reiterer[3], Maik Winter[4], Wolfgang Ertel[4], Jörg Eberhardt[4]

TWT GmbH Science & Innovation[1]
Sarissa GmbH[2]
University of Konstanz[3]
University of Applied Sciences Ravensburg-Weingarten [4]

Summary

The research project ERTRAG, funded by the German Federal Ministry of Education and Research, aims to develop a virtual ergonomics trainer for nursing as well as elderly care education. The trainer supports apprentices conducting the care related motions in an ergonomically correct fashion, thereby helping to prevent work-related health problems such as dorsal pain. The user interface of the ERTRAG system for instruction and feedback adjusts to the individual preferences and needs, as well as to the current cognitive state of the user. This adaptive, highly customizable human-machine interaction fosters effective and long-lasting learning.

1 Introduction

Nursing and elderly care are subject to an increasing shortage of trained personnel, enhanced by the demographic change that societies are facing today. Moreover, employees in care professions have a higher risk of developing musculoskeletal disorders. This not only reduces the inherently small number of care professionals by way of sickness absence and early retirement, but also incurs significant socio-economic penalties. Therefore, measures need to be taken to support current and future care personnel by significantly improving their working conditions. In this regard, the educational system plays a central role, since behavioural prevention can help to reduce work-related health problems. Currently, back-friendly techniques in nursing / elderly care are often taught in block seminars only once or twice during

Veröffentlicht durch die Gesellschaft für Informatik e. V. 2017 in
M. Burghardt, R. Wimmer, C. Wolff, C. Womser-Hacker (Hrsg.):
Mensch und Computer 2017 – Workshopband, 10.–13. September 2017, Regensburg.

the three-year educational programme. The exercise sessions ensure that care-related techniques are being implemented correctly. Moreover, ergonomically unfavourable movements and body positions of the apprentices are identified and corrected. Feedback is generally given by the instructor during the exercise. The research project ERTRAG strives to extend the educational possibilities of nursing and elderly care apprentices, by offering a virtual ergonomics-trainer that provides qualified instruction and feedback outside the standard block seminars. The project started in June 2016 and is being implemented by a consortium of research institutes and industrial firms located in Baden-Württemberg. The project partners are TWT GmbH Science & Innovation (consortium lead), Sarissa GmbH, University of Konstanz and the University of Applied Sciences Ravensburg-Weingarten. Funding is provided by the German Federal Ministry of Education and Research.

2 Virtual Ergonomics-Trainer

In order to optimally support care apprentices in learning ergonomic motion sequences, the ERTRAG system has to address their specific requirements and needs. The first task of the project is thus to perform a thorough requirement analysis. The technical implementation of the system includes the tracking and analysis of movements, as well as a suitable interface for instruction and feedback.

2.1 Requirement Analysis

To identify the requirements concerning the virtual ergonomics-trainer, observerships and interviews were carried out at medical and health care schools and one university. In total, five heads of school, five lecturers and 27 apprentices were interviewed regarding the teaching of ergonomic motion sequences in nursing education. Interviews were audio-recorded. Moreover, four observerships were conducted, including active participation in the respective training-exercises. The sessions were video-recorded and personal experiences and observations were additionally documented in writing. A contextual analysis of the acquired data was performed by way of Affinity Diagramming (Hartson & Pyla, 2012). As a result of the analysis, 111 requirements were defined for the envisioned ergonomics-system. For example, it was observed that apprentices frequently switch roles between caretaker and elderly / patient. Five apprentices indicated that they regard the option to take the role of elderly / patient as an important tool for self-awareness. Consequently, a virtual ergonomics-trainer should support switching between different roles in order to learn from various perspectives. The interviewed heads of school noted that a potential advantage of the virtual ergonomics-trainer is its motivational aspect, since it matches the interests of the younger, technically-oriented generation. Moreover, it would offer the opportunity to demonstrate correct motion sequences whenever the apprentices so require (rather than being restricted to the few demonstration sessions in conventional teaching). Regarding feedback, haptic modalities were described as proven effective, but also optical and acoustic signals were being viewed as useful. The requirements identified in this initial data acquisition initiative form the basis of the technical developments of the ERTRAG system.

2.2 Motion Tracking

Three dimensional motion capture is an essential part of the virtual ergonomics-trainer. From the requirement analysis follows that the ERTRAG system should be unobtrusive and easy to use. In compliance with these requirements, a holistic system for motion capture is being implemented.

The motion capture makes use of 3D cameras that function on the principle of (light) time of flight. In contrast to conventional cameras, the 3D cameras not only generate standard image data, but also provide depth information. This results in three dimensional image data with color and distance values for each pixel. In recent years, gaming console manufacturers such as Microsoft and Sony have successfully contributed to market readyness of these cameras. The IR-lightsource integrated in the 3D camera emits a periodically modulated signal that is being reflected from the objects in the scene and received by the camera sensor. Emitted and reflected signal are compared by evaluation electronics and from the phase difference between the two signals, the distance to the reflecting object can be computed on a per-pixel basis. The modulation of the emitted signal allows to reliably filter-out ambient light – therefore 3D cameras are especially suitable for use in environments with a high degree of interfering light, specifically including outdoor use.

The acquired 3D data are mapped onto a virtual human model, consisting of abstract skeleton segments and joints. This mapping facilitates automatic processing of the data, e.g. differentiation between the individual persons visible in the scene and analysis of their respective posture. The transfer of camera data to virtual human model is implemented without additional markes attached to the person performing the motion. This is a highly nontrivial task, especially when two people are in close interaction, as is the case for caretaker and elderly person. Many movements in such scenarios will result in image data that contain occlusions or are ambiguous, leading to a false assessment of the motion performance. Garcia and Zakhor have shown in their work that a multi-view setup can mitigate this problem substantially (Garcia & Zakhor, 2013). Therefore, the ERTRAG system makes use of two or more 3D cameras that capture the care scenario from different viewing angles and whose data is merged to a joint data model. The mapping of camera data onto the virtual human model is based on the previous work of (Shotton et al., 2013) and is in principle independent of the type of camera used. The ERTRAG 3D motion capture employs mitiaturised 3D cameras of the Siegen-based company pmdtechnologies (pmdtechnologies GmbH, 2015). The consortium member Sarissa GmbH manufactures ultrasound-based positional markers that provide 3D location information with high frequency (kHz) and spatial resolution (+/- 5mm). These markers are being used to validate the position of segments and joints of the virtual human model, computed from the optical data.

In addition to posture and movement measurements, ground reaction forces are measured by way of pressure sensitive shoe inlays. This additional information is often essential in evaluating the ergonomic quality of a motion sequence.

2.3 Motion analysis

Motion analysis deals with the recognition of correct or incorrect human stances or movements while performing an activity. Typically, skeleton or silhouette (depth) data are used for motion tracking and pose estimation (Ye et al., 2013; Du et al., 2015). In the ERTRAG project, we develop classifier software that analyses the care apprentice's skeleton data in order to estimate the movement quality and provide feedback accordingly. Because of the inherent complexity of this task, model-based classical software development is inappropriate, and we rely on machine learning for automatic training of the classifier. Interviews conducted with kinaesthetic and physiotherapy experts revealed that there is no gold standard movement sequence for any given nursing scenario. However, it is possible to identify universally wrong movements that should be avoided in order to maintain good musculoskeletal health.

In order to apply machine learning algorithms, labelled training data is required. State of the art datasets[1] based on depth and skeleton models are publicly available for activity and gesture recognition. However, since they assume a gold standard for motions, these data cannot be used in ERTRAG. Therefore, we aim at generating a training data set that captures the variability of motions (e.g. due to varying physiognomy of patient and care apprentice) and explicitly also contains ergonomically incorrect motion sequences. The dataset will comprise image and video sequences of the activities that are frequently performed in day to day nursing care, such as transfer of the patient from a lying position in bed to a sitting position at the edge of the bed. The depth data are in the form of images and the skeleton data in the form of position and orientation of the joints for each frame. The first batch of the training data were recorded using single-view Microsoft Kinect v2 with the help of a kinaesthetic-expert and two nursing students. It comprises data of both correct and incorrect movement sequences.

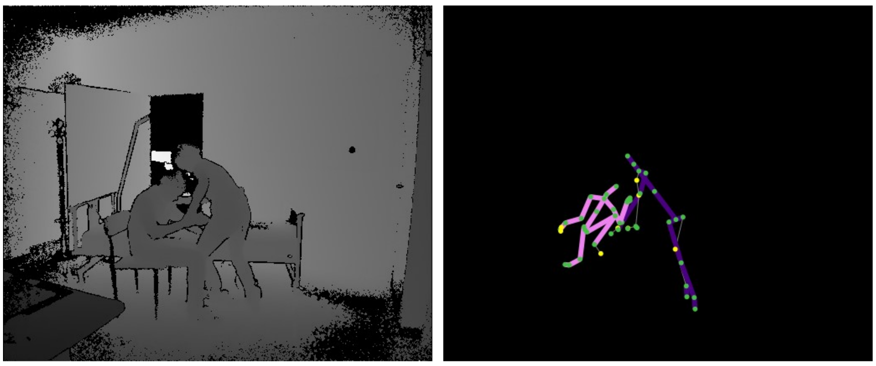

Figure 1: Training data. Depth and Skeleton images for the scenario, transfer of the patient from chair to bed.

We developed a tool built on the Kinect API 2.0 that can capture the RGB, skeleton and depth data simultaneously for a given scenario. Options for recording data, converting image sequence into a video and filtering out non-desirable skeleton data are available. An example

[1] Available at: *http://users.eecs.northwestern.edu/~jwa368/my_data.html*

of the depth and skeleton images, that were captured for classifier-training is shown in Figure 1.

A user-friendly tool was developed, which enables experts to label the data efficiently. The tool samples the video data at one image per second and presents the expert with these images in a randomized order. If an image depicts unergonomic behaviour, the expert can assign one or more ergonometric error classes and annotate the severity of these errors. Additional error classes can be added if the default classes are not sufficient. The tool will be made accessible over the internet, eliminating the need for local installation, thereby reducing the overhead for the experts. Labelling of and classifier-training and evaluation on the preliminary training data will indicate whether static images are sufficient for movement assessment, or if dynamic (video) data are required. Based on these findings, suitable supervised machine learning techniques will be implemented and tested, such as Support Vector Machines (Cortes & Vapnik, 1995), Hidden Markov Models (Baum et al., 1970) and Artificial Neural Networks (McCullough & Pitts, 1943).

2.4 Adaptive User Interface

The simplest form of custom learning supported by the ERTRAG system is the possibility to choose learning content individually. Moreover, the learning history and achievements, as well as possible deficits can be retrieved and analysed in order to visualise the individual learning progress. In addition, the research project ERTRAG investigates how different modalities for instruction and feedback can be employed to dynamically adjust interaction according to personal learning preference and cognitive user state. The cognitive state of the apprentice is of importance, since cognitive over- or understraining due to instruction and feedback may stand in the way of successful learning. Therefore, the virtual ergonomics-trainer strives to achieve a level of interaction complexity that is adjusted to the individual cognitive state, preference, ability and experience in order to foster effective learning.

In order to put these goals technically into practice, the ERTRAG system records physiological parameters such as heartrate, skin conductivity, eye-movements and facial expressions of the apprentice while learning and exercising the care-related motion sequences. The acquired data are analysed with machine learning techniques such that the cognitive load of the user can be assessed with low latency (ideally in real-time). In this fashion, different modalities for instruction and feedback (visual, auditory, haptic) can be evaluated and multimodal interaction concepts are developed. Moreover, the ideal time for instruction and feedback can be determined. The classifier for cognitive load is developed in a series of studies. Currently, laboratory tests are performed with tasks of well-defined difficulty levels, such as the N-Back paradigm, using the full range of sensors (heart rate belt, skin conductivity sensor, high-resolution video data and eye-tracking glasses for facial expressions and eye-movements). At later project stages, studies will be conducted in realistic educational elderly care settings with the aim of reducing the sensors to a minimal (yet reliable) set.

3 Conclusion

The ultimate goal of ERTRAG is to support nursing and elderly care apprentices with a virtual coach that teaches workplace ergonomics in a highly personalised, motivating and enjoyable manner. The ERTRAG system supplements regular educational courses and aims to be unobtrusive and intuitively usable. In contrast to conventional desktop systems, the ERTRAG system supports the concepts of situated learning and embodied cognition. Instruction and feedback are given realistic scenarios, the system is not limited to conveying theoretical knowledge, but emphasizes the actual performing of the care related motion sequences. User requirements and feedback guide the entire development process in order to achieve high acceptance. Notably, this will have an influence on the final set of sensors and feedback modalities, as well as on the cost of the ERTRAG system.

Literature

Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The annals of mathematical statistics*, 41(1), 164-171.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273-297.

Du, Y., Wang, W., & Wang, L. (2015). Hierarchical recurrent neural network for skeleton based action recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1110-1118).

Garcia, R.R. & Zakhor, A. (2013). Multi-view dynamic geometry capture using structured light, *SPIE Newsroom*, DOI: 10.1117/2.1201301.004724

Hartson, R., & Pyla, P. S. (2012). The UX Book: Process and guidelines for ensuring a quality user experience. Elsevier.

McCullough, W. S. & Pitts, W. (1943), A logical calculus of the ideas immanent in nervous activity, *Bulletin of Mathematical Biophysics*, 5 (4): 115–127

pmdtechnologies GmbH (2015). Reference Design Brief CamBoard pico flexx, Datasheet

Shotton, J., Fitzgibbon, A. Cook, M. et al. (2015). Real-Time Human Pose Recognition in Parts from Single Depth Images. *Machine Learning for Computer Vision*, Volume 411, 119-135, Springer Berlin Heidelberg

Ye, M., Zhang, Q., Wang, L., Zhu, J., Yang, R., & Gall, J. (2013). A survey on human motion analysis from depth data. In *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications* (pp. 149-187). Springer Berlin Heidelberg.