Enhancing the Wellbeing at the Workplace

Rainer Planinc and Michael Hödlmoser
Computer Vision Lab
Vienna University of Technology
Email: rainer.planinc@tuwien.ac.at

Martin Kampel
CogVis GmbH
Email: kampel@cogvis.at

***Abstract*—Physical and psychological issues are a common problem for elderly employees. Working in an sitting position with not ergonomically designed working places, too little exercises in combination with unhealthy nutrition and stress are the main causes of the problems. In order to prevent such problems, we propose to monitor the workplace in an holistic way, thus detecting unergonomically workplaces, too little exercises, stress and an unbalanced diet automatically. The proposed system is able to detect unhealthy behavior by using a 3D sensor in combination with a camera and speech recognition. The proposed approach will be evaluated at least at 50 work places in order to prove its feasibility.**

Keywords–workplace; ergonomics; AAL; elderly; physical activity; stress management; nutritional balance;

I. Introduction

A significant amount of jobs require employees to sit for eight hours (or even longer) while performing their tasks [1]. During younger years, employees are not bothered with physical problems, but while getting older, more and more issues due extensive sitting occur. In combination with unergonomically designed working places, malnutrition (nutrition requirements change during our lifetime) and only little exercising, severe problem can arise, especially for people in the age of 50+ [2][3]. Furthermore, the cognitive capability decreases, whereas (together with many other factors) the stress level increases. This leads to employees feeling not very well, which indicates a reduced quality of life due to physical (e.g., back) and psychological (e.g., stress) problems. The aim of this paper is not to counteract on physical and psychological issues, but to prevent them already at an early stage. The proposed system is suitable for all employees performing their job in an sitting position, e.g., secretaries, office clerks, technicians, lawyers, etc. at the same level. Moreover, even employees not performing their jobs using a desk (e.g., bus driver) benefit from this approach since the system is flexible enough to be adapted to different needs easily. The proposed system consists of four different modules, which in combination lead to an enhanced quality of life of older adults in their working environment, but also to an increase of the overall wellbeing. The combination of these modules provides an training program for physical and psychological exercises and consists of the modules physical training, workplace ergonomics, nutritional balance and stress management optimized for the needs and wishes of older adults. Only by combining physical, ergonomic, nutritional and cognitive aspects, a holistic approach to enhance the wellbeing and quality of life of older adults can be developed. The focus is on providing a well-balanced and holistic system that is able to meet the end-users needs and offers benefits and an enhanced quality of life for older adults at the workplace. The rest of this paper is structured as follows: Section 2 provides an overview of the system, its modules are described in detail in Sections 3-6, a conclusion is drawn in Section 7.

II. Overview

The proposed system helps to increase the health of elderly people at their workspaces, focusing on providing an all-in-one solution for increasing the wellbeing of older adults at their workplace. The system focuses on workplaces where employees are performing their work in a sitting position, i.e., office workers. It can mainly be divided into two parts, which are the online wellbeing service for processing the data and generating feedback in real time, for logging, storing and representing results and a combination of sensors in order to gather input data (Figure 1). The quality of the physical training, as well as the performance of the stress management are handled by a 3D sensor (e.g., Kinect), the ergonomic workplace is obtained by exploiting a camera in combination with the 3D sensor, and the nutrition optimization is conducted by a combination of a camera and speech recognition. The workflow of the proposed approach is illustrated in Figure 2 and illustrates the combination of different sensor types.

However, since employees are monitored at their workplaces, privacy issues need to be addressed. The system is developed in order to ensure that only the employee is able to access his or her health data. This is extremely important since 1) only the person itself is allowed to access the data and 2) the obtained data is health data and thus need to be protected. Hence, appropriate measures to ensure privacy are taken and they include the implementation of proper user role management and encryption mechanisms. Moreover, the system needs to be unobtrusive in order to avoid negative side effects (e.g., employees are concerned of a misuse of the system). In combination with guidance of the company medical officer of each company, the system is not expected to trigger negative effects.

All modules, except the one for nutritional balance, are based on a correct pose estimation of the user. Pose estimation can be done by first extracting features in the image which describe local changes between different poses. After collecting the images, the system needs to be trained and novel input poses need to be classified based on this training data. Determining the pose of a human is done in 3D space. To overcome the ill-posed problem of extracting 3D poses out of 2D images, the RGB images were replaced by depth images obtained from a 3D sensor. By using depth images,

 ISBN: 978-1-61208-384-1

Figure 1. Wellbeing platform and modules

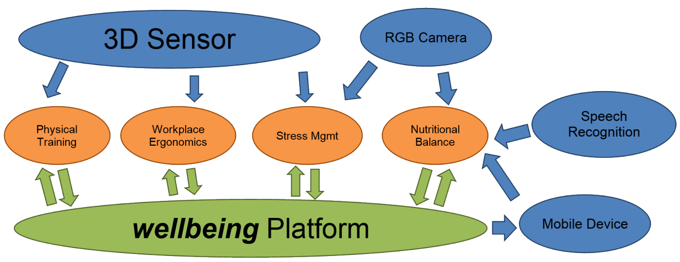

Figure 2. Workflow of the proposed approach

Fanelli et al. in [4] use regression in combination with a random forest is used for determining a human's head pose. Similar features are also used by Shotton et al. [5] where the first system for obtaining a human's pose is described. Each body part casts votes for a single class. The final pose is estimated by generating confidence-scored 3D proposals of how the body parts are connected. As a 3D sensor is able to overcome the ill-posed problem of extracting 3D information out of 2D images and is proven to be very robust against clutter and environmental changes, the input for the proposed system will also be based on this sensor for human pose estimation. Nevertheless, RGB images provide valueable information necessary to perform other task than 3D pose estimation (e.g., detection of a chair or a desk). As this information is also needed for the proposed framework, RGB images serve as an additional input source. The aim of speech recognition is to enable a very natural interaction with the computer by speaking instead of using traditional input devices and not only have the machine understand the verbal content, but also more subtle cues such as affect that any human listener would easily react to. However, so far real-time emotion recognition has scarcely been attempted and if so, only in prototypical applications. Technical challenges that arise when equipping human-computer interfaces with the ability to recognize the users vocal emotions are almost endless. Lately, several tentative studies were published recently on the interpretation of speech signals in terms of certain application-dependent affective states [6]. As can be seen, there is a research shift towards the analysis of spontaneous human behavior [7], which means that the analysis of acoustic information will not only suffice for identifying subtle changes in vocal affect expression. There are two driving trends focusing on (i) audiovisual analysis which combines linguistic and nonlinguistic analysis and (ii) visual analysis coming from multiple cues (facial expressions, head movements, and/or body gestures). This way, great outcomes are expected when merging both frontline cues into one single technology.

III. Physical Training

Health and well-being oriented companies today offer their employees either the use of a gym or even provide the possibilities for other forms of exercising (e.g., aerobics). However, providing such facilities is very cost intensive and hence, only a small number of companies offer these activities to their employees. Nevertheless, companies are getting aware that many health related issues their older and more experienced employees are suffering from, could have been prevented if detected already at an early stage. However, since aerobics lessons and gyms are expensive, especially in bigger companies with a high number of employees, other ways of exercising were found. A common way of motivating employees to exercise is to provide them information about small exercises to perform on their own. There are different ways of presenting these exercises, e.g., in books, videos or on the Internet. The main problem with these techniques in comparison to a personal trainer is the lack of feedback since employees cannot verify if their movement during the exercise is correct. This issue is tackled by the use of a 3D sensor, which is able to detect the movement of the person. Furthermore, together with the proposed framework, wellbeing is able to provide feedback during the exercises automatically. The advantages of using this sensor is twofold: on the one hand, feedback about the accuracy of the movement is gathered and tips to improve the exercises on an individual level can be given without having high costs of a personal trainer. On the other hand, the activity can be logged automatically and thus can be used to motivate the end-user to exercise more often if only sporadic use is detected. The exercises itself can be presented in different ways, e.g., by using the conventional way where the exercise is shown and explained using a video, but also more entertaining ways of explanation can be used, e.g., as mini games, combining physical exercise together with gaming elements. These mini games are called exergames and are not simply video games, but are based on fundamental research by physicians. Hence, this type of games being able to tackle health related problems in a very entertaining way will be integrated during the project.

Physical training is based on pose estimation, during the performance of exercises in front of the system. These exercises are divided into conventional training and so called exergames, which should help to motivate the user for the workout. There is already a variety of different fitness games on the market, which use 3D information as input for tracking the user (e.g., Kinect Fitness, Nike+ Kinect Training, Kinect Sports, Kinect Zumba, etc.). Studies performed at the University of Chester show that the fitness games boosted the heart rates by up to 194 percent over sedentary games and increase the energy expenditure by 263 percent over resting values [8]. When having the human pose by using the real-time method described in [5][9], it is also possible to track body parts over time. Then, this enables keeping track of certain regions, which

 ISBN: 978-1-61208-384-1

are more important than others for playing exergames and performing conventional training (e.g., arms, legs, head). As the input coming from the sensors is then sent to an online service, it is implicitly possible to also play against other users over the Internet. Having the components of first packing exercises in games and second competing against colleagues or other persons over the Internet increases the motivation for the user to perform the workout needed for a healthier working environment.

IV. Workplace Ergonomics

Work place ergonomics is the key factor of a healthy work place. Paying attention to ergonomic standards cannot only prevent back pain or postural deformity, but also vision problems and tensions. Work place ergonomics consists of guidelines for finding optimal distances to e.g., a monitor, adjusting the height of the chair and table, but also recommends regular breaks for relaxation, e.g., of the eyes. These rules already exist, but it is often hard to apply them in practice since you might sit correctly for a specific time after reading these rules, but getting lazy and forget them after a while. Since especially older adults are prone to postural deformity and vision problems, we propose to detect and track the position of the employees during their work using the 3D sensor introduced above and offer possibilities for improving the actual position. These offers range from reminders for taking breaks at regular intervals, analysing the sitting position while performing tasks and reminding the person to correct the pose (together with tips on how to correct them) and to provide relaxation exercises for the muscles, as well as the eyes. Since the system is based on an end-user centred approach, a non-intrusive system meeting the end-users requirements is designed. In contrast to regular reminders, the system does not remind the user at pre-defined intervals but takes the overall fitness of the user into consideration. The system is highly customized in order to consider the individual needs of different users, where one might benefit from more breaks or others from specific exercises.

Also for this module, it is necessary to obtain the users 3D pose first by using the 3D sensor. Additionally, objects in the environment need to be detected (e.g., desk, chair, etc.). As the 3D sensor is not providing any colour or texture information, an RGB camera is used to extract additional information necessary in order to assure these object detection tasks. Conventional object detection is performed by extracting some features first which represent the local structure of a patch. A classifier (e.g., support vector machine, random forest) is then used to train the system and to obtain a classification scheme. As in [10] or [11], the system can be trained on a variety of different objects. Felzenszwalb et al. [10] present a system for 2D object detection of rigid objects with manually labeled training images. The object is not detected as a whole but by using the combination of the parts, which make up the object. Since the images are labeled manually, the system is also able to estimate the 3D pose of the object in the scene. Liebelt et al. [11] provide a different approach, which is based on existing 3D models. Features are extracted from projections of these rigid models and the system is then able to estimate both the type of object and its pose in 3D space. These methods need to be extended in order to handle the type of objects needed for wellbeing.

V. Nutritional Balance

According to the WHO many diseases of older adults and elderly are diet-affected. These diseases are, amongst others, cardiovascular and cerebrovascular disease, diabetes, osteoporosis, and cancer, which are among the most common diseases affecting older persons. However, the WHO found out that only increasing the consumption of fruit and vegetables by one or two servings per day could reduce the cardiovascular risk by 30%. Since these results are achieved by only very simple measures, the proposed system aims at even increasing this percentage by ensuring the nutritional balance in a holistic way. The integration of this module is strongly motivated by the fact that older adults often do not have many possibilities to pay attention to their nutritional balance when they are in the office due to reduced time and resources. Furthermore, old habits and stress during work even worsens these problems and thus the consortium focuses on providing information for enhancing the nutritional balance. This module consists of a food recognition unit, with which the food can be recognized by providing a picture and/or spoken information. The food recognition unit consists of computer vision algorithms in order to estimate the amount of carbohydrate, protein and vegetables. These results are enhanced by providing spoken information, e.g., the approximate weight of meat, the type of protein/carbohydrate/vegetables. Offering this easy to use multi-modal interface, an individual nutrition schedule is recommended and monitored. In addition, information about the nutrition schedule, as well as unobtrusive reminders (e.g., ensuring the intake of the right amount of water) are provided.

Different to all other modules, surveying the nutritional balance is performed without using the 3D sensor but by exploiting a RGB camera and a speech recognition module. As both an RGB image and speech recognition is also available on a mobile device, this task can also be performed on such devices. Pouladzadeh et al. [12] use a support vector machine classifier and feature extraction using colour, texture, size and shape of the food. Bosch et al. [13] also use food identification for diet measurement. Due to the significant variations in appearance, Yang et.al. introduce an algorithm, which exploits the spatial relationship between ingredients [14]. Different to existing approaches, the goal is to gather a rough idea of the ingredients of the meals and determine e.g., how much carbohydrates, fat, or vitamins are present in the meal. Answering these questions is a different task compared to state of the art methods, which aim for determining e.g., which meat is on the plate. As the complexity of finding the ingredients of a meal is much lower compared to the complexity of a full food recognition system, the proposed task can be solved much more accurate and more efficient and is therefore a perfect fit for the needed purpose. To make the recognition even more robust, the visual detection system is combined with a speech recognition module, which provides the possibility to record spoken information about the meal, e.g., I had smashed potatoes and beef, a cup of peas and a cup of sauce. After a few weeks of taking pictures of daily meals, there is already enough data to suggest a perfect food plan. The user then receives individualized suggestions for improvements, or recipes. As can be seen, for a full food recognition system it is necessary to outsource the computation of segmentation, features and classification on an external server for a fast and reliable recognition. This makes it a perfect fit for the proposed wellbeing system. Image processing on RGB images in combination with speech recognition are exploited in order to detect the food on a plate. In order to deal with non-centred food in the image or cluttered background, the system has to segment the important parts of the image (the food) and discard objects in the background. The final decision is passed to the user interface from the computation server. If the result is incorrect, the user can send the feedback from the user interface to the server with the correct food information.

VI. Stress Management

Job stress is a serious problem among older workers that leads to physical and psychological health conditions and, in a mid-term perspective, to early retirement [15][16]. Stress at work occurs when occupational demands are higher than resources (excessive demand), and the mediating variables job autonomy and social support are low [17]. Our technology will prevent job stress by a) increasing personal internal resources (cognitive and physical abilities) and b) increasing social interaction among the employees by playing exergames together and in competition among each other. In order to enhance the communication and exchange between different employees, the games developed in wellbeing also contains a social component and can therefore either be played together in teams or the competition between different players is used as motivating factor for performing exercises from different modules. Furthermore, tools for stress management will be provided. These tools can also be combined with exergames introduced earlier and thereby these games not only offers physical training, but also relaxation and fun.

The proposed solution analyses stress related factors in order to estimate the current stress level. This analysis is mainly based on physiological factors, caused by stress (e.g., more reddish colour of the face). The application then gives each indicator a certain weight and measures the stress level. If the stress level exceeds a certain threshold, short exercises are proposed by the system, e.g., relaxation, acupressure, breathing-technique or any other slight movement. Similar to physical training, exergames can also be proposed here which can be played against other persons. The type and length of an exercise proposed depends on the level of stress. This diversity of different types of exercise and stress-reducing measures increases the motivation of the user. The stress management system is using the 3D sensor as an input device and the workflow of the module is based on the scientific background in the fields of the analysis of stress related facial signs and the execution and the impact of stress-reducing exercises.

VII. Conclusion

Monitoring ergonomical aspects at the workplace together with motivating user to exercise regularly reduces physical problems. In combination with the monitoring of stress related factors, as well as the nutrition ensures that employees stay healthy and fit. The wellbeing platform uses a 3D and color (RGB) sensor in order to motivate elderly employees to exercise more often by providing exercises on the one hand, and social exergames on the other hand. Moreover, ergonomical aspects about the workplace (sitting posture, distance to the monitor, etc.) are analyzed by the platform and feedback to correct the position is provided immediately. By analyzing the employees food and stress level, appropriate interventions can be suggested at the right time, ensuring a healthy lifestyle. The system will be evaluated on more than 50 users during the course of the wellbeing project, testing the system for the duration of 12 months in order to obtain reliable results.

Acknowledgment

This work is supported by the European Union under grant AAL 2013-6-063. The authors want to thank the whole wellbeing project team.

References

[1] S. McCrady and J. Levine, “Sedentariness at work; how much do we really sit?” Obesity, vol. 17, no. 11, 2010, pp. 2103–2105.

[2] R. Castanharo, M. Duarte, and S. McGill, “Corrective sitting strategies: An examination of muscle activity and spine loading,” Journal of Electromyography and Kinesiology, vol. 24, no. 1, 2014, pp. 114–119.

[3] J. P. Caneiro, P. OSullivan, A. Burnett, A. Barach, D. ONeil, O. Tveit, and K. Olafsdottir, “The influence of different sitting postures on head/neck posture and muscle activity,” Manual Therapy, vol. 15, no. 1, 2010, pp. 54–60.

[4] G. Fanelli, J. Gall, and L. Van Gool, “Real time head pose estimation with random regression forests,” in Conference on Computer Vision and Pattern Recognition, 2011, pp. 617–624.

[5] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, “Real-time human pose recognition in parts from single depth images,” in Conference on Computer Vision and Pattern Recognition, 2011, pp. 1297–1304.

[6] J. B. Hirschberg, S. Benus, J. M. Brenier, F. Enos, S. Friedman, S. Gilman, C. Girand, M. Graciarena, A. Kathol, L. Michaelis, B. Pellom, A. Stolcke, and E. Shriberg, “Distinguishing deceptive from non-deceptive speech,” in Proc. of European Conf. Speech Comm. and Technology, 2005, pp. 1833 – 1836.

[7] J. Liscombe, J. Hirschberg, and J. Venditti, “Detecting certainness in spoken tutorials dialogues,” in Proc. European Conf. Speech Comm. and Technology, 2005, pp. 1837 – 1840.

[8] A. Park, “Xbox or Wii: Which Is Better for Sedentary Kids?” Sep. 2012, accessed: 2014-12-18. [Online]. Available: http://healthland.time.com/2012/09/26/xbox-or-wii-a-scientific-argument-for-why-xbox-is-better-for-sedentary-kids/

[9] G. Rogez, J. Rihan, S. Ramalingam, C. Orrite, and P. H. S. Torr, “Randomized trees for human pose detection,” in Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.

[10] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, “Object Detection with Discriminatively Trained Part-Based Models,” Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 9, Sep. 2010, pp. 1627–1645.

[11] J. Liebelt, C. Schmid, and K. Schertler, “Viewpoint-independent object class detection using 3D Feature Maps,” in Conference on Computer Vision and Pattern Recognition, vol. 64. IEEE, Jun. 2008, pp. 1–8.

[12] P. Pouladzadeh, G. Villalobos, R. Almaghrabi, and S. Shirmohammadi, “A novel svm based food recognition method for calorie measurement applications,” in Proc. of ICME Workshop, 2012, pp. 495 – 498.

[13] M. Bosch, F. Zhu, N. Khanna, C. Boushey, and E. Delp, “Combining global and local features for food identification in dietary assessment,” in Proc. of ICIP, 2011, pp. 1789 – 1792.

[14] S. Yang, M. Chen, D. Pomerleau, and R. Sukthankar, “Food recognition using statistics of pairwise local features,” in Proc. of CVPR, 2010, pp. 2249 – 2256.

[15] J. Kulmala, M. von Bonsdorff, S. Stenholm, T. Trmkangas, M. von Bonsdorff, C. Nygrd, M. Klockars, J. Seitsamo, J. Ilmarinen, and T. Rantanen, “Perceived stress symptoms in midlife predict disability in old age: A 28-year prospective cohort study,” The Journals of Gerontology Series A, vol. 68, no. 8, 2013, pp. 984 – 991.

[16] M. Marmot, “Whitehall I & II,” accessed: 2014-12-18. [Online]. Available: http://www.ucl.ac.uk/whitehallII

[17] J. Siegrist, M. Wahrendorf, O. von dem Knesebeck, H. Jrges, and A. Brsch-Supan, “Quality of work, well-being, and intended early retirement of older employees baseline results from the share study,” European Journal of Public Health, vol. 17, no. 1, 2006, pp. 62 – 68.