Video-based quantification of patient's compliance during post-stroke virtual reality rehabilitation

M Divjak, S Zelič, A Holobar

Faculty of Electrical Engineering and Computer Science, University of Maribor,
Smetanova ulica 17, 2000 Maribor, SLOVENIA

matjaz.divjak@um.si, *simon.zelic@um.si*, *ales.holobar@um.si*

www.feri.um.si

ABSTRACT

We present a video-based monitoring system for quantification of patient's attention to visual feedback during robot assisted gait rehabilitation. Patient's face and facial features are detected online and used to estimate the approximate gaze direction. This gaze information is then used to calculate various metrics of patient's attention. Results demonstrate that such unobtrusive video-based gaze tracking is feasible and that it can be used to support assessment of patient's compliance with the rehabilitation therapy.

1. INTRODUCTION

Stroke is the third most common cause of death in Western society, with ~4.7 million stroke survivors alive today. One of the hallmark residual impairments of stroke is post-stroke walking disability, which creates a stigma for patients, makes them more susceptible to injury and directly affects their quality of life. Early rehabilitation therapy is crucial for significant improvements (Kollen et al, 2006). In recent years, robotic systems are widely tested and employed to retrain stroke patients. By imposing gait-like movements at a comfortable speed, such robotic devices are thought to provide many of the afferent cues regarded as critical to retraining locomotion (Mehrholz et al, 2007).

However, a major problem with existing stroke therapies is patient non-compliance (Matarić et al, 2007). Many stroke patients abandon the therapy because the process is too long, repetitive, and/or does not provide immediate results (Teasell and Kalra, 2005). European project BETTER (Project, 2013) recently addressed a new approach to gait rehabilitation by employing non-invasive brain-neural computer interaction (BNCI) based assistive technologies based on EEG, EMG and IMU sensors. In this paper we extend the BNCI-based modalities with video-based attention monitoring system that allows automatic quantification and long-term monitoring of user attention. Such attention tracking does not require any sensors to be attached to the patient, making this method easy and fast to apply in stroke rehabilitation. Our main objective is to quantify patient's attention to visual feedback, i.e. the amount of time the patient's gaze is actively following the displayed visual feedback and to inspect the possible impact of visual feedback on motor planning, as well as its short- and mid-term benefits.

Several studies already examined the impact of visual feedback on stroke rehabilitation, but they mostly reported inconclusive results. They focused on quantification of results of rehabilitation enhanced with different kinds of visual feedback, but their experimental designs did not allow for a reliable quantification of the attention a person is paying to stimuli. To the best of our knowledge, only psycho-physiological measures of patient's attention to visual feedback have been reported in the literature (Bakker et al, 2007), while video-based assessment of attention has not been proposed in the field of rehabilitation.

Methods for real-time detection of face, facial features, eye movements and gaze direction from video have attracted a lot of research in the past decade (Hansen and Ji, 2010). Numerous algorithms for facial feature extraction have been proposed, mostly in the context of face detection and recognition (Bagherian and Rahmat, 2008). Currently, the most promising approaches rely on fusing multiple visual cues, such as combining local feature matching with intensity-based methods (Liao et al, 2010). Existing methods for quantification of user's attention to visual feedback have mostly been developed for Human-Computer Interaction applications and in order to help severely disabled people (Poole and Ball, 2005).

2. METHODS

An overview of our approach is depicted in Fig. 1. The patient is fixed in a robotic gait trainer, which moves patient's legs according to predefined rehabilitation scenarios, adapted to patient's current walking abilities. In front of the patient is a large TV screen showing various visual feedbacks. A high speed video camera is mounted above the TV. The camera captures HD video of patient's face, while simultaneously EEG is recorded from 51 scalp sites and EMG is recorded from both legs (tibialis anterior).

The video streams are processed by our algorithms. First, frontal faces are detected and main facial features are extracted (corners of the eyes, pupil centers, mouth, tip of the nose). To suppress jitter from detection errors and body swings during walking, a Kalman filter is applied to the locations of extracted facial features. Next, an active appearance model (AAM – Matthews et al, 2007) of the face with 59 facial landmarks is used to more accurately represent the current facial pose/expression. The distances between extracted landmarks are used to calculate the head pose and gaze direction. Finally, different metrics of patient's attention /compliance are computed: percentage of time the patient is observing the visual feedback, spatial gaze distribution maps, responses to actions in the VR environment, etc.

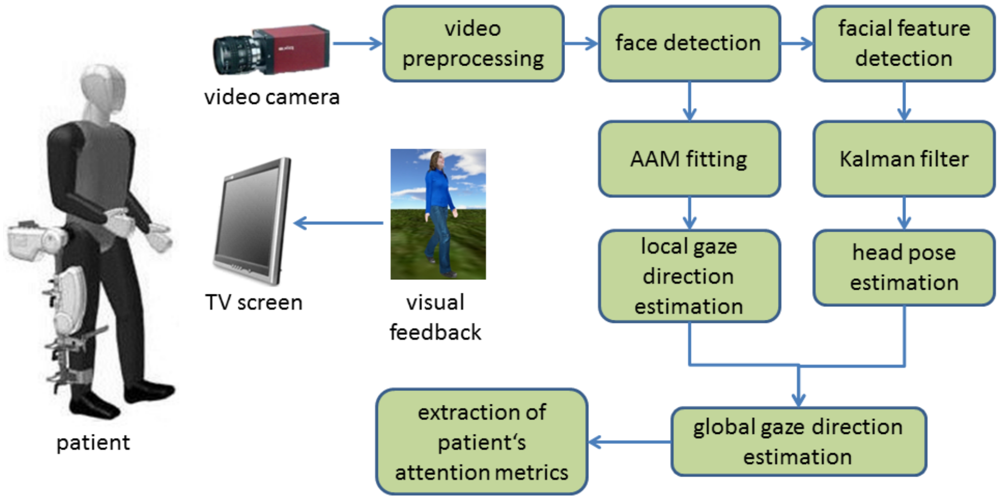

Figure 1. *Overview of the proposed approach for video-based estimation of patient's attention/compliance during robot-enhanced rehabilitation.*

We tested 5 different visual feedbacks (Fig. 2): calibration screen for initialization of our algorithms, 2D plots of current leg positions, 2D plots of muscle activations, and two realistic VR environments (walk in a park from 1st and 3rd person perspective). The VR environments were created in OpenGL 3.3 and consist of ground, sky dome and male/female avatars. The avatars were created in Mixamo Pro Character Creator Tool and animated with AutoDesk 3DS Max 2012. Both avatars have an underlying bone structure with all major joints modeled, allowing animation of almost all human movements. The actual avatar movement within the VR environment is controlled by kinematic data from the robotic trainer, so leg movements in VR world correspond with leg movements in real life.

The whole video processing system was designed for real-time operation, but current version of the software is a mix of C++ and Matlab routines and runs at ~1 frame per second. Therefore, all video processing is currently performed offline, but (near) real-time operation is possible with further optimization of the code.

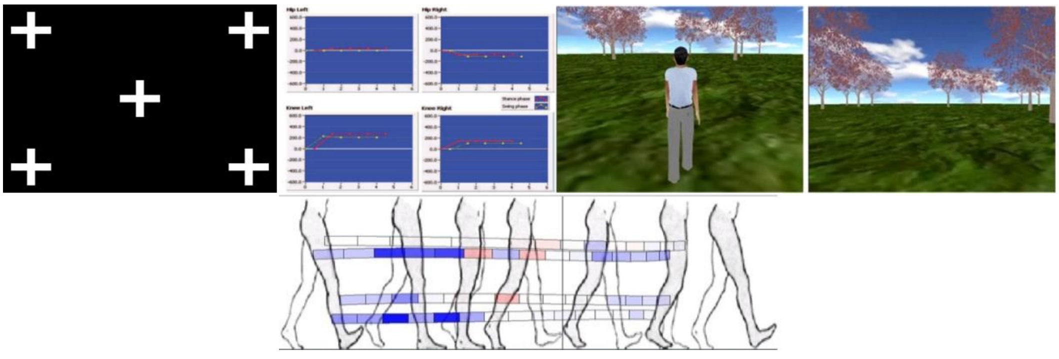

Figure 2. *Five different visual feedbacks that are shown to the patient during therapy.*

Proc. 10th Intl Conf. Disability, Virtual Reality & Associated Technologies
Gothenburg, Sweden, 2–4 Sept. 2014
©2014 ICDVRAT; ISBN 978-0-7049-1546-6

3. EXPERIMENTAL RESULTS

The proposed system was evaluated in an experiment that involved 4 stroke patients and consisted of 5 runs of robot assisted walking with different visual feedbacks. In all runs, the patients were instructed to walk actively, maintaining a constant speed, and applying minimum force on the robot. A 42" screen was placed in front of the patient, 1.4 m away from his face. Each run lasted 4 minutes. For all visual feedbacks the walking speed remained constant. During the experiments videos of the patient's face were recorded to disk and later processed offline.

On average, the face was detected in 93 % of video frames recorded (i.e. in practically all the frames with frontal faces whereas the profile faces were not detected). Eyes, mouth, nose, and pupils were accurately detected in more than 99% of detected faces (Table 1). Detection of facial features was skipped in video frames with no face detected. The average jitter of detected facial features was 1.0 ± 0.4 pixels. The AAM was successfully fitted to all frontal faces. The average jitter of AAM landmarks was estimated at 0.6 ± 0.4 pixels.

Table 1. *Facial feature detection performance, estimated on three test videos.*

	Video 1 (gaze targets)		Video 2 (gaze targets)		Video 3 (VR 3rd person)	
	frames	%	frames	%	frames	%
Whole video	9999	100 %	13391	100 %	10551	100 %
Face detection	9833	93.7 %	13193	98.5 %	10427	98.8 %
Detection of left eye	9805	99.7 %*	13174	99.9 %*	10419	99.9 %*
Detection of right eye	9769	99.3 %*	13179	99.9 %*	10408	99.8 %*
Detection of left pupil	9800	99.9 %#	13191	99.9 %#	10419	100 %#
Detection of right pupil	9769	100 %#	13187	100 %#	10408	100 %#
Detection of nose	9827	99.9 %*	13150	99.8 %*	10423	99.9 %*
Detection of mouth	9742	99.1 %*	13174	100 %*	10414	99.8 %*

* Values normalized by the number of face detections.
# Values normalized by the number of eye detections.

Videos recorded during sessions with the screen displaying calibration targets were inspected by an expert and 9 approximate gaze direction (top-left, top-centre, top-right, bottom-left, bottom-centre, bottom right, left-centre, right-centre and centre of the screen) were manually annotated. The time periods corresponding to eye movements or eye blinks were ignored and were not annotated. Gaze direction was then calculated automatically by our algorithm and compared to manually annotated gaze locations. The gaze directions were identified with average accuracy of 94% ± 6%. Most errors originated from distinguishing between top-left vs. bottom-left and top-right vs. bottom-right gaze directions.

Fig. 3 shows an example of patient's detected level of attention to visual feedback, estimated as percentage of time in 1 second intervals with gaze fixed to the TV screen; all gazes outside the screen were classified as non-attention. Fig. 4 presents an example of the estimated spatiotemporal gaze distribution metric. This metric shows relative frequency of identified gaze locations over 4 minutes long gait rehabilitation. Brighter spots in the gaze distribution plots indicate areas with more frequent attention. Such maps support identification of gaze targets (i.e. gaze hot-spots) and assessment of spatiotemporal correlation between patient's attention to visual feedback and other BNCI-based performance indices of gait rehabilitation.

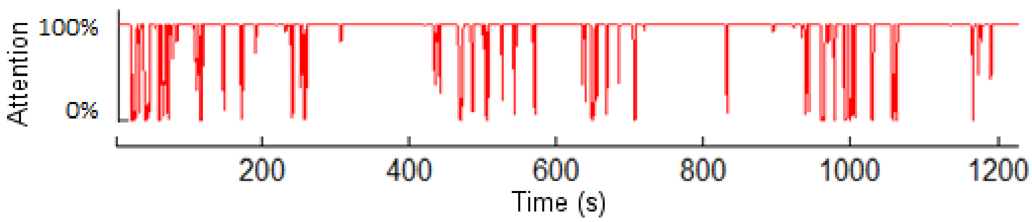

Figure 3. *An example of estimated level of patient's attention to visual feedback.*

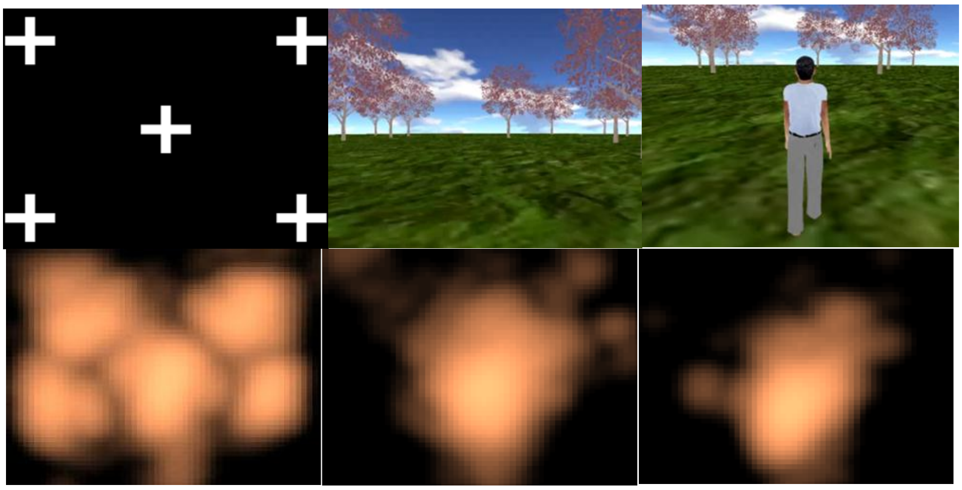

Figure 4. *Example of spatial gaze distribution plots (bottom row) for three different visual feedbacks (top row) shown to a patient during 4 minutes long gait rehabilitation.*

4. CONCLUSIONS

We developed and validated a video-based system for robust tracking of patient's face pose and gaze direction during robot-assisted lower limb rehabilitation therapy. The results (Fig. 3 and Fig. 4) show that such system can be used to unobtrusively analyze patient's attention to displayed visual feedback and to provide means for long-term quantification of visual feedback effects on the rehabilitation progress. It can also serve as an additional feature for other BNCI performance indices, such as similarity of motor modules, kinetic and kinematic profiles and brain patterns. For example, preliminary results show significant correlation between attention to muscle activation plots and improvements in muscle activations during walking.

Acknowledgements: This work was supported by the Commission of the European Union within Framework 7, grant agreement FP7-2009-7.2–247935 - "BETTER – Brain-Neural Computer Interaction for Evaluation and Testing of Physical Therapies in Stroke Rehabilitation of Gait Disorders". The authors would like to thank the personnel of Fondazione Santa Lucia, Rome, Italy, for organizing and performing the experimental recordings.

5. REFERENCES

Bagherian, E, and Rahmat, RWOK, (2008), Facial feature extraction for face recognition: a review, *International Symposium on Information Technology*, pp. 1 – 9.

Bakker, M, de Lange, FP, Stevens, JA, Toni, I, and Bloem BR, (2007), Motor imagery of gait: a quantitative approach, *Exp. Brain Res.*, 179, pp. 497–504.

Project BETTER, (2013), http://www.car.upm-csic.es/bioingenieria/better/.

Hansen, DW, and Ji, Q, (2010), In the eye of the beholder: A survey of models for eyes and gaze, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 32, Iss. 3, pp. 478-500.

Kollen, B, Kwakkel, G, and Lindeman, E, (2006), Functional Recovery After Stroke: A Review of Current Developments in Stroke Rehabilitation Research, *Reviews on Recent Clinical Trials*, 1, pp. 75-80.

Liao, WK, Fidaleo, D, and Medioni, G, (2010), Robust, real-time 3D face tracking from a monocular view, *EURASIP Journal on Image and Video Processing*, Vol. 2010, article ID 183605.

Matthews, I, Xiao, J, and Baker, S, (2007), 2D vs. 3D Deformable Face Models: Representational Power, Construction, and Real-Time Fitting, *Internat. J. of Comput. Vision*, 75(1), pp. 93-113.

Matarić, MJ, Eriksson, J, Feil-Seifer, DJ, and Winstein, CJ, (2007), Socially assistive robotics for post-stroke rehabilitation, *J. Neuroeng. Rehabil.*, 4(5).

Mehrholz, J, Werner, C, Kugler, J, and Pohl, M, (2007) Electromechanical-assisted training for walking after stroke, *Cochrane Database Syst Rev*, 17(4).

Poole, A, and Ball, LJ, (2005), Eye tracking in human-computer interaction and usability research: Current status and future. In *Encyclopedia of Human-Computer Interaction* (C Ghaouli), Idea Group, Pennsylvania.

Teasell, R, and Kalra, L, (2005), What's new in stroke rehabilitation: Back to basics, *Stroke*, 36, pp. 215-217.