International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:8, No:7, 2014

Video-Based System for Support of Robot-Enhanced Gait Rehabilitation of Stroke Patients

Matjaž Divjak, Simon Zelič, Aleš Holobar

***Abstract*—**We present a dedicated video-based monitoring system for quantification of patient's attention to visual feedback during robot assisted gait rehabilitation. Two different approaches for eye gaze and head pose tracking are tested and compared. Several metrics for assessment of patient's attention are also presented. Experimental results with healthy volunteers demonstrate that unobtrusive video-based gaze tracking during the robot-assisted gait rehabilitation is possible and is sufficiently robust for quantification of patient's attention and assessment of compliance with the rehabilitation therapy.

***Keywords*—**Video-based attention monitoring, gaze estimation, stroke rehabilitation, user compliance.

I. Introduction

STROKE is the third most common cause of death in Western society. Prevalence figures are in the vicinity of 5-5.5 % in USA with around 700,000 new cases each year. For every decade after age 55, the relative incidence of stroke doubles. About 4.7 million stroke survivors (2.3 million men, 2.4 million women) are alive today. The effects of stroke depend on several factors including the location of the obstruction and how much brain tissue is affected. One of the hallmark residual deficits of stroke is post-stroke walking disability. Walking incorrectly not only creates a stigma for the patients, but also makes them more susceptible to injury and directly affects their quality of life.

Early rehabilitation therapy is crucial for significant improvements in the treatment outcome [1]. Robotics-based systems are being widely tested and employed to retrain stroke patients. By imposing gait-like movements at a comfortable speed and without restricted duration, such robotic devices are thought to provide many of the afferent cues regarded as critical to retraining locomotion [2].

However, a significant problem with existing stroke therapy is patient non-compliance [3]. Patients often find certain aspects of therapy frustrating, exhausting, boring, annoying, or prone to error. It is recognized that most stroke patients who begin therapy to improve motor control, abandon the therapy because the process is too long, repetitive, and does not provide immediate results [4].

European project BETTER [5] addresses a new approach to rehabilitation therapies of gait disorders in stroke patients by employing non-invasive Brain/Neural Computer Interaction (BNCI) based assistive technologies. One of the main goals of the proposed multimodal BNCI system is to provide immediate visual feedback of the rehabilitation progress to the patient and to characterize the level of patient's involvement in the rehabilitation therapy based on EEG, EMG and IMU sensors.

We propose to extend these modalities with video-based attention monitoring system that allows automatic quantification and long-term monitoring of user attention to the provided visual stimuli. Video-based tracking of patient's attention does not require any sensors to be attached to the patient, making this method fast and easy to apply in stroke rehabilitation. The objectives of this video-based upgrade are threefold:

a) to quantify patient's attention to visual feedback, i.e. the amount of time the patient's gaze is actively following the visual feedback;
b) to systematically test different modalities of visual feedback, their possible impact on motor planning, and their short- and mid-term benefits; and
c) to increase the robustness of the BNCI-based assessment of patient's compliance by merging video-based information with the BNCI system.

II. State of the Art

Patient's responses to visual stimuli can be detected by EEG, but this requires highly controlled experimental conditions and high level of user cooperation. On the other hand, several obvious visual signs of user's attention to visual stimuli, such as the direction of user's eye gaze, his/her responses to dynamically generated visual stimuli, tired eyes and change in blinking patterns can be readily detected by video-based monitoring of a subject's face.

Methods for real-time detection of eye movements and eye gaze direction from video have attracted a lot of research in the past decade [6]. Numerous algorithms for facial feature extraction have been proposed, mostly for face detection and recognition [7]. Currently, the most promising approaches rely on fusing multiple visual cues, for example by combining local feature matching with intensity-based methods [8]. The aforementioned quantifications of user's attention to visual feedback have mostly been developed for Human-Computer Interaction (HCI) applications and in order to help severely disabled people [9]. Since changes in eye blinking rate are known to be related with fatigue and drowsiness, this fact is often exploited for detecting alertness of drivers and signs of fatigue [10], with PERCLOS (Percentage of Eye Closure) and AECS (Average Eye Closure Speed) as the most popular video-based fatigue measures [11]. When combined with

Matjaž Divjak, Simon Zelič, and Aleš Holobar are with the Faculty of Electrical Engineering and Computer Science, University of Maribor, Slovenia (e-mail: matjaz.divjak@ um.si, simon.zelic@ um.si, ales.holobar@ um.si).

facial expressions, stereotypical head movements etc. a probabilistic model of human fatigue can be constructed [12]. We believe a similar approach can be used to detect patient's responses to the rehabilitation and quantify his level of attention.

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 modalities 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 attention to visual feedback have been reported in the literature [13], while video-based assessment of attention has not yet been proposed in the field of rehabilitation.

III. Algorithms for Video-Based Quantification of Patient's Attention

Two complementary approaches have been developed. The first, feature-based approach is based on a collection of different feature-tracking algorithms and extracts eleven main facial features, i.e. inner and outer corners of eyes, pupil centers, mouth corners, left and right nose corner, and tip of the nose. The second approach builds on active appearance models (AAM) to represent the patient's face with a 2D grid of 59 facial landmarks, and enables more accurate and robust tracking of facial mimics.

The rationale behind building these two approaches is to find the optimal compromise between robustness and efficiency of video-based tracking of facial components during robot assisted walking. Both aforementioned approaches have been systematically tested on video recordings of healthy volunteers as well as stroke patients during their rehabilitation. Both approaches do not exclude, but rather complement each other, as the feature-based facial tracking is also used for initialization and constraint optimization of the AAM model.

A. Feature Based Tracking

First, the OpenCV's Haar detector [14] is used to locate frontal faces in every video frame. In each detected face, the discriminative model of feature appearance in the form of boosted classifiers using Haar-like features [15] is employed to localize the nine facial landmarks denoting the corners of the eyes, mouth, and the tip of the nose (Fig. 1). Centers of pupils are detected by radial symmetry approach [16], which offers the best cost/performance ratio among the tested pupil detectors. In about 3% of the video frames tested, the radial symmetry approach yields an erroneous position of the pupil's center. These sudden deviations of pupil's center can easily be detected online and are corrected by more robust, but also computationally more intensive adaptive approach described in [17].

Afterwards, in order to suppress jitter due to detection errors and body swings during walking, a prediction-correction algorithm based on Kalman filter is applied to the locations of extracted facial features.

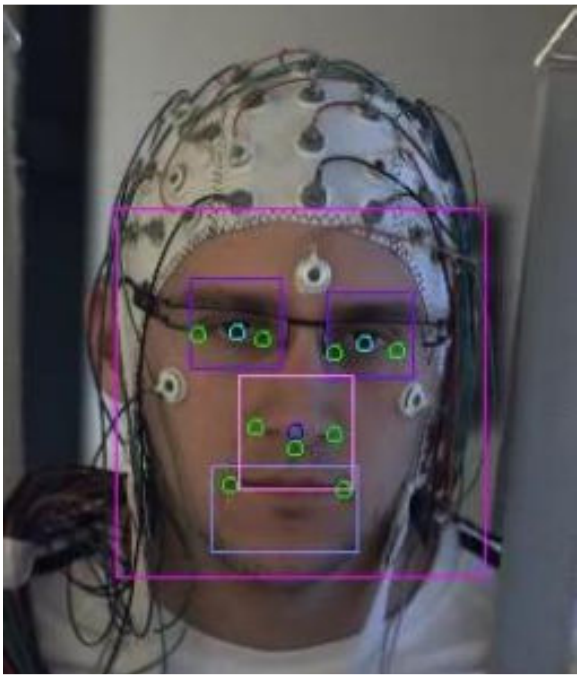


Fig. 1 Representative results of facial tracking in the recorded video of a healthy volunteer. The extracted face and facial regions are denoted by rectangles. Facial landmarks, as identified by the feature extractor are denoted by circles

Finally, the distances between extracted landmarks are used to calculate the head pose and gaze direction. By following the approach described in [18], head pose is calculated as a normal to the planar face region, defined by the outer corners of the eyes and mouth, and the tip of the nose (Fig. 2a). Vertical gaze direction is determined from the projection of the pupil center to the line connecting both corners of the eye (Fig. 2b). Horizontal gaze direction is calculated by mutual comparison of distances d_R in right and d_L in left eye (Fig. 2c).

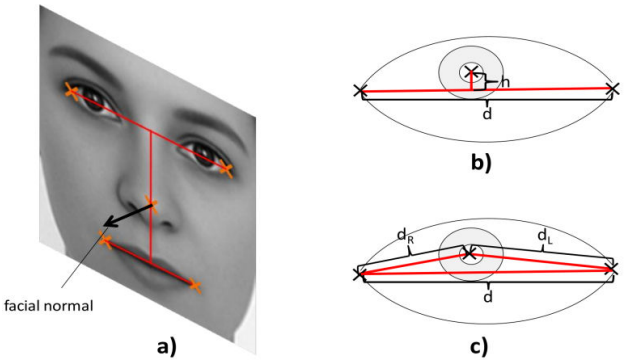


Fig. 2 Schematic representation of low level measures for calculating head pose and gaze direction

B. Active Appearance Based Tracking

The second implemented approach utilizes AAM, a parametric technique for tracking face and facial expressions, which supports more robust personalization and dynamic adaptation to different facial expressions, in comparison to the

feature-based approach. Following the preliminary assessment of different AAM implementations, the algorithm based on the inverse compositional approach (IC AAM) [19] was selected as the most suitable for our purposes. We built a generalized model of human face from a large set of annotated images of different people. This facial model uses 59 facial landmark points to model main facial components from the chin to the forehead and from left to right ear. A few representative examples of the constructed IC AAM fitted to the videos from healthy volunteers are shown in Fig. 3.

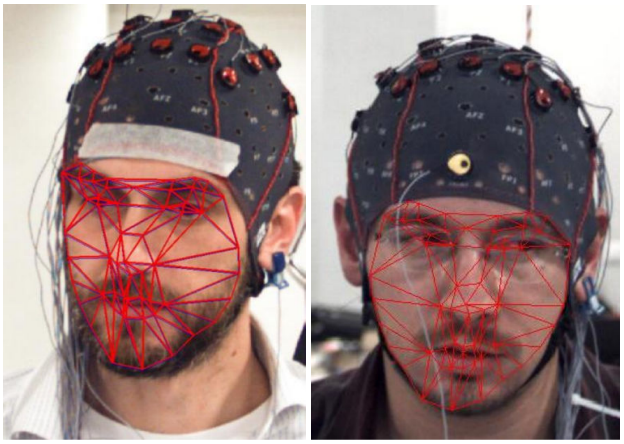


Fig. 3 Facial active appearance model overlaid over images from two healthy volunteers during their robot assisted gait training. The red AAM mesh consists of 59 landmark points, denoting different facial features

Developed AAM models rely significantly on accurate and robust initialization and initial fitting to the facial features of each patient. Although this procedure is theoretically required only once per patient (i.e. during his/her first rehabilitation session), it might prove beneficial also in later rehabilitation sessions, especially when the appearance of the patient's face is significantly altered (e.g. patient shaves, puts glasses on/off, etc.). This makes the initialization of AAM a crucial step of AAM-based tracking.

In the literature, AAM models are typically initialized by manually annotating the crucial anatomical landmarks on a large set of images. However, such approach is slow and requires considerable skill, making it unsuitable for clinical practice. To circumvent these shortcomings, the following automatic constrained fitting of non-person-specific AAM model to images of each individual patient is performed:

1. Image Acquisition during the Calibration Phase:

This step is performed at the beginning of rehabilitation (e.g., during the first rehabilitation session). The patient is asked to look at graphical markers (top-left panel in Fig. 4) displayed in the corners of the screen. The markers appear sequentially and in random order. Each marker appears and then disappears after 2 seconds. During this calibration phase, the video is recorded for offline processing.

2. Automatic Selection of Initialization Images:

The recorded video is analyzed offline by feature-based detector (Section IIIA) to detect 50 frontal faces with different head poses and eye/mouth appearances.

3. Automatic Annotation of Facial landmarks:

The facial landmarks extracted by feature based detectors are used to initialize the generic AAM facial model to each individual frame. Constrained fitting of obtained AAM model is then applied to annotate all 59 facial landmarks in each of the 50 initialization frames.

4. Learning of the Patient-Specific AAM Model:

The newly fitted generic AAM models are used for re-learning of the patient-specific AAM model.

5. Storage of the Patient-Specific AAM Model:

The patient-specific AAM model is stored for use in later rehabilitation sessions.

After fitting the prepared AAM to every frame of the input video, the head pose and gaze vectors are calculated by the approach described in Section III *A*. The nine facial landmarks described in Section III *A* form a subset of AAM landmarks. Thus, only the superior quality of AAM-based feature points is currently exploited by our head pose and gaze estimators, while the information on all other facial components, such as mouth movements and facial mimics is currently ignored.

C. Estimation of Gaze Direction

Two approaches to automatic gaze direction identification have been tested. In the first approach the extracted head pose and gaze directions are fed into a Support Vector Machine (SVM) classifier with a Gaussian radial basis function kernel (sigma=1). First half of video recording with calibration targets (around 120 s) is typically used for classifier training, whereas the second half of the same video is used for assessment of tracking accuracy.

The second approach builds on direct calculation of the gaze vector. First, head pose is calculated as in [18]. Then, d_R, d_L and h metrics (as illustrated in Fig. 2c) are used to rotate the estimated head-pose vector in the direction of the gaze. The standard eye model with diameter of 24 mm is used to calculate the gaze rotation angles. Although slightly slower than SVM, this gaze vector model is more general and does not rely on a completeness of a training set. It still requires a short calibration phase with gaze targets displayed in the corners of the screen whenever the geometric relation between the patient and the visual feedback screen is changed, but this is merely to (re-)calculate the size and the position of the feedback screen in the gaze space.

D. Visual Feedback Modalities

Visual feedback consists of two parts. The first part is the calibration screen used during system setup (top-left panel in Fig. 4). In video calibration session, the patient is asked to focus his gaze for about 2 seconds on five graphical markers that are sequentially displayed in the corners and the center of the screen. This helps to adapt the facial and head pose

tracking to a specific patient and compensates for differences in the geometric setup of rehabilitation sessions (e.g. relative position of patient and the screen with the visual feedback).

The second part of the visual feedback is the virtual environment (VE). It utilizes the OpenGL 3.3 library and consists of ground, sky dome and 3D avatar (Fig. 4). The ground limits the VE and provides a feeling of a solid walking surface that can be custom textured or colored. The sky dome limits the space above and gives an open space feeling. The avatar is a virtual representation of the patient and its movement within the VE is controlled by the kinematic data from the robot trainer.

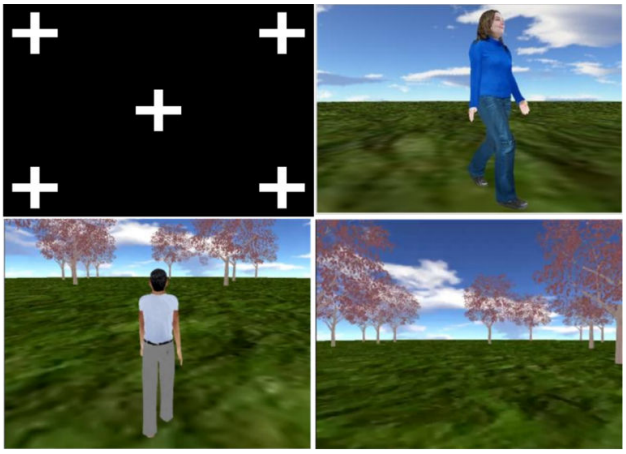


Fig. 4 Visual feedback modalities. Video tracking calibration screen (top-right panel), 3D female avatar as seen by free-form camera (top-right panel), 3D male avatar from the 3rd person perspective (bottom-left panel), and VE from the 1st person perspective (bottom-right panel)

IV. Experimental Results

The experiment involved eleven healthy volunteers and consisted of ten runs of robot assisted walking, with five different feedback modalities (Fig. 5); each modality was presented twice to each participant. In all runs, the participants were instructed to walk actively, maintaining a constant speed, and applying minimum force on the robot. A 42 inch screen was placed in front of the patient, 1.4 m away from his face. Each run lasted four minutes. In all feedback modalities, the walking speed remained constant.

A high-speed video-capture system (Basler Ace acA2000 CameraLink camera with Matrox Radient frame-grabber) was mounted on top of the screen with visual feedback and used to capture high resolution video (2040×1086 pixels) at 100 frames per second. Simultaneously, EEG was recorded from 61 scalp sites, with electrodes placed according to the 5% 10-20 system [20]. The EMG was recorded from the left and right arms (carpi radialis and deltoid posterior) and both legs (tibialis anterior).

Videos recorded during the calibration sessions were visually inspected by an expert. Time intervals with gaze fixed to the visual cues displayed in the center and in the corners of the screen were carefully annotated to serve as a reference. The time intervals corresponding to eye movements or eye blinks were ignored.

Fig. 5 Healthy volunteer performing the robotic gait training in front of the video screen

A. Assessment of Facial Feature Tracking

Results of feature-based tracking of facial components over three different sessions (four minutes long video recordings) are reported in Table I. 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. Detection of facial features was skipped in video frames with no face detected. The average jitter of facial components is reported in Table II, whereas the processing time, as measured on a standard personal computer (Intel Core I7-930 CPU, 6 GB of memory) is reported in Table III. Video processing algorithms were implemented in Matlab with several C++ based MEX files.

The results of AAM-based facial components tracking are summarized in Table IV and Table V. On average, the face was detected in 93 % of video frames recorded during the experimental sessions. The AAM was successfully fitted to all detected faces. In video frames with no face detected, the AAM fitting has been skipped. The average processing time is reported in Table VI.

TABLE I
NUMBER OF FRAMES WITH FACIAL COMPONENTS SUCCESSFULLY DETECTED BY THE HAAR CLASSIFIERS AND THE DISCRIMINATIVE MODEL OF FEATURE APPEARANCE

	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.

TABLE II
JITTER (MEAN ± SD, IN PIXELS) OF FACE AND FACIAL FEATURES LOCATIONS WITH AND WITHOUT KALMAN TRACKING IN TOTAL, 2000 VIDEO FRAMES WITH GAZE FIXED TO TEN DIFFERENT SCREEN POSITIONS WERE ANALYZED IN EACH VIDEO

	Video 1 (gaze targets)		Video 2 (gaze targets)	
	Without tracking	With tracking	Without tracking	With tracking
Face detection: X	2.0 ± 3.3	1.1 ± 0.5	1.4 ± 0.4	0.6 ± 0.2
Face detection: Y	2.5 ± 6.3	1.0 ± 0.3	1.7 ± 0.5	1.1 ± 0.4
Eye detections: X	4.9 ± 2.4	1.3 ± 0.7	3.4 ± 1.1	0.5 ± 0.2
Eye detections: Y	5.0 ± 2.3	1.6 ± 0.9	4.1 ± 1.2	0.8 ± 0.3
Pupil detections: X	2.7 ± 1.8	1.3 ± 0.5	2.0 ± 1.1	0.7 ± 0.3
Pupil detections: Y	3.0 ± 1.9	1.3 ± 0.6	2.8 ± 1.2	0.5 ± 0.2

TABLE III
PROCESSING TIME (IN MS) REQUIRED FOR DETECTION OF FACE AND FACIAL COMPONENTS

	Video 1 (gaze targets)	Video 2 (gaze targets)	Video 3 (VR 3rdperson)
Face detection	43.8 ± 3.5 ms	43.4 ± 8.1 ms	44.1 ± 3.0 ms
Haar based detection of eyes, mouth, and nose	41.8 ± 4.5 ms	32.8 ± 3.5 ms	41.1 ± 3.9 ms
Detection of pupils	5.5 ± 0.7 ms	4.6 ± 2.9 ms	5.4 ± 0.6 ms
SURF/SIFT based detection of eyes, mouth, and nose	38.5 ± 4.5 ms	33.8 ± 2.9 ms	45.2 ± 4.5 ms
Detection of eyes, mouth, and nose by discriminative feature appearance model	87.3 ± 4.2 ms	84.4 ± 3.8 ms	88.5 ± 4.7 ms

As demonstrated by the results in Table V, the AAM fitting is much more robust against gait related head movements than the feature-based tracking presented in Section IIIA. Therefore, it offers superior gaze tracking accuracy at the price of slightly higher processing costs. It is also important to note that AMM tracker relies on feature-based tracker in the initialization step.

TABLE IV
NUMBER OF FRAMES WITH FACIAL AAM SUCCESSFULLY FITTED IN VIDEO FRAMES WITH NO FACE DETECTED, THE AAM FITTING WAS SKIPPED

	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 %
Successful AAM fitting to detected face	9833	100 %*	13193	100 %*	10427	100 %*

* Values normalized by the number of face detections.

TABLE V
JITTER (MEAN ± SD, IN PIXELS) OF FACIAL FEATURES LOCATIONS AS IDENTIFIED BY FACIAL AAM (VIDEO RECORDINGS OF TWO REPRESENTATIVE HEALTHY VOLUNTEERS). IN TOTAL, 2000 VIDEO FRAMES WITH GAZE FIXED TO TEN DIFFERENT SCREEN POSITIONS WERE ANALYZED IN EACH VIDEO

	Video 1 (gaze targets)	Video 2 (gaze targets)
Right eye detection: X	0.52 ± 0.35	0.60 ± 0.38
Right eye detection: Y	0.61 ± 0.43	0.45 ± 0.31
Left eye detection: X	0.44 ± 0.38	0.57 ± 0.32
Left eye detection: Y	0.68 ± 0.45	0.55 ± 0.42
Right pupil detection: X	0.48 ± 0.36	0.52 ± 0.28
Right pupil detection: Y	0.58 ± 0.37	0.95 ± 0.61
Left pupil detection: X	0.51 ± 0.36	0.49 ± 0.25
Left pupil detection: Y	0.58 ± 0.36	0.32 ± 0.21

TABLE VI
PROCESSING TIME (IN MS) REQUIRED FOR AMM DETECTION OF FACE AND FACIAL COMPONENTS

	Video 1 (gaze targets)	Video 2 (gaze targets)	Video 3 (VR 3rdperson)
Facial AAM fitting (per frame)	162 ± 26 ms	161 ± 23 ms	165 ± 24 ms
AAM iterations needed (per frame)	3.01 ± 0.64	2.73 ± 0.48	3.16 ± 0.54

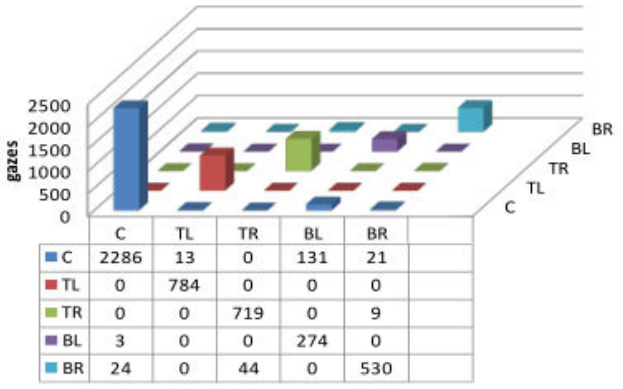
B. Assessment of Video-Based Head-Pose and Gaze Tracking

Both approaches to gaze classification (Section IIIC) were compared to manually annotated gaze directions. The following metrics have been used to compare the performances of SVM-based and vector-based gaze classifiers:

a) Accuracy of SVM-based gaze detection during robot-assisted gait rehabilitation with calibration target screen;
b) Performance of vector-based gaze estimation during robot-assisted gait rehabilitation with calibration target screen;
c) Agreement between SVM-based and vector-based gaze detection during robot-assisted gait rehabilitation with various feedback modalities.
d) Spatiotemporal gaze distribution during robot-assisted gait rehabilitation.

The first three metrics measure the head-pose and gaze tracking accuracy in general, whereas the fourth metric measures the spatiotemporal dynamics of patient's gaze.

In healthy volunteers the SVM-based gaze directions were identified with average accuracy of 94% ± 6% (Fig. 6). Most of errors originated from distinguishing between top-left vs. bottom-left and top-right vs. bottom-right gaze directions.



	C	TL	TR	BL	BR
C	2286	13	0	131	21
TL	0	784	0	0	0
TR	0	0	719	0	9
BL	3	0	0	274	0
BR	24	0	44	0	530

Fig. 6 Typical example of confusion matrix for SVM-based gaze estimation in a healthy volunteer

Vector-based gaze tracking was assessed by comparison to manually annotated video recordings of gaze target feedback. Videos recorded during sessions with the screen displaying calibration targets were inspected by an expert and nine 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 vector-based gaze tracker and compared to manually annotated gaze directions.

Representative results of vector-based gaze tracking with gaze targets displayed in the center of the screen and in the centers of all four screen edges are depicted in Fig. 7. Each marker denotes the gaze direction as assessed by vector-based gaze tracking model. Different classes of gaze directions as determined by manual (left panel) and SVM-based classification (right panel) are denoted by different colors. Both vector-based and SVM-based models largely agreed with the manual classification, with errors mostly appearing in distinction of vertical directions (e.g. top-left vs. bottom-left and top-right vs. bottom-right corners).

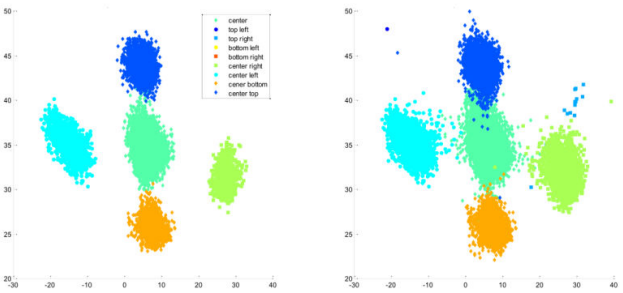


Fig. 7 Representative results of gaze direction assessment by vector-based AAM gaze tracking, compared to manual gaze annotation (left) and SVM-based automatic classification (right)

C. Assessment of Patient's Attention to Visual Feedback

The following metrics were used to assess patient's attention to the visual feedback:

a) Attention to visual feedback: percentage of time in 1 second intervals with gaze fixed to the feedback screen; all gazes outside the feedback screen were classified as non-attention. An example of this metric is displayed in Fig. 8.
b) Spatiotemporal gaze distribution plots: cumulative plots of gaze directions revealing the hot-spots of user's attention (Fig. 9).

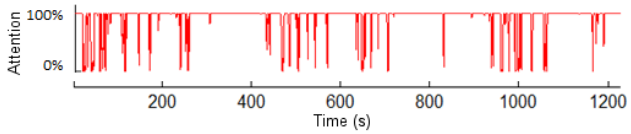


Fig. 8 An example of estimated level of patient's attention to visual feedback during a 20 min session

The spatiotemporal gaze distribution metric is exemplified in Fig. 9. It supports 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. The examples in Fig. 9 show relative frequency of identified gaze classes over 4 minutes long robot-assisted gait rehabilitation of a healthy volunteer, with lighter spots indicating areas with more frequent attention.

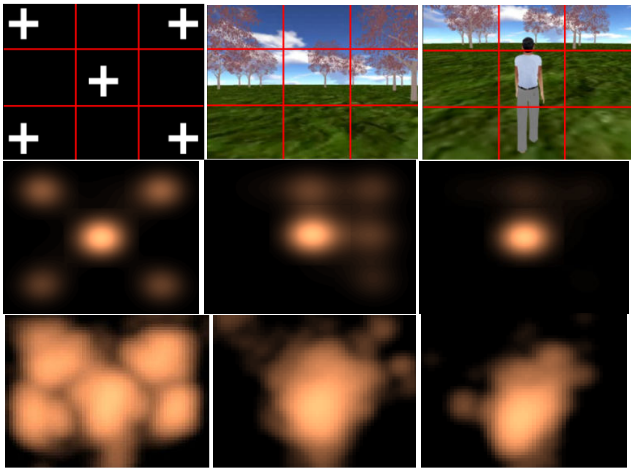


Fig. 9 Spatial gaze distribution plots for different visual feedback modalities (top panels) for a healthy control subject during 4 minutes long gait rehabilitation. Middle panels: SVM-based gaze classification. Bottom panels: vector-based gaze tracking

Spatial gaze distribution can also be combined with other BNCI metrics of patient's performance. For example, it can be accumulated only across time moments of wrong muscle activation patterns, or over gait cycles with large hip/knee forces as assessed by the robot trainer. This adds a temporal dimension to the gaze distribution plots, giving spatio-temporal gaze distributions.

In Fig. 10, gaze distributions have been accumulated over the entire 4 minutes of rehabilitation runs and correlated to the similarity of motor modules (SMM) during each gait cycle.

SMM metric measures the global similarity of muscle activations (for all the measured muscles and motor modules throughout the whole gait cycle) in stroke patients with the muscle activation patterns in healthy control subjects. Its value ranges between 0% and 100 %, with the value of 100% representing the equality to the muscle activation patterns in healthy control subjects.

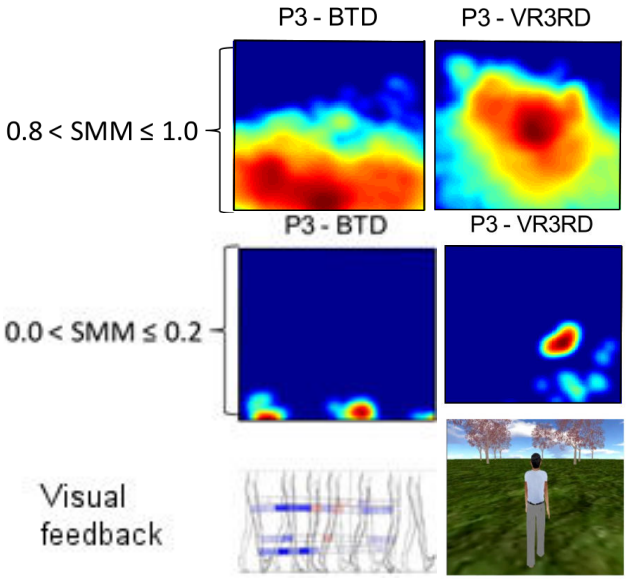


Fig. 10 Spatiotemporal gaze distribution plots for two different visual feedback modalities as determined by vector-based gaze tracking in healthy volunteer during 4 minutes long gait rehabilitation

V. Discussion

In the context of video-based assessment of attention to visual feedback two different facial tracking approaches have been functionally validated: feature-based discriminative model proposed by Sivic et al. [15] and the AAM model proposed by Matthews et al. [19]. Results on eleven healthy volunteers demonstrated that unobtrusive gaze tracking based on video is possible during the robot-assisted gait rehabilitation. The feature-based discriminative face tracker is robust to large and sudden head movements, but offers inferior accuracy in detection of facial components when compared to AAM model. As a result, it is more prone to the perturbations due to body and head movements during robot-assisted rehabilitation. These perturbations can partially be compensated by filtering, e.g. by Kalman filter. Nevertheless, the feature-based discriminative model supports only coarse gaze estimation, discriminating between attention and non-attention to the visual feedback screen.

When compared to the feature-based tracker, the AAM tracker is much more robust to head and body swings, but suffers from occasional problems in refitting to the user's face after intensive movements away from the camera. The errors of AAM fitting appeared mainly in the cases of severe facial occlusion or/and large head movements. Tracking 59 facial landmark points, the AAM model also requires more computational power than the feature-based discriminative model.

Two different gaze classifiers have been evaluated: SVM and gaze vector model. The first one enables fast and fully automatic classification of identified facial features into 3x3 gaze classes, but requires relatively large learning sets, and thus long (typically 4 minutes) calibration sessions to yield the accuracy of ~ 90%. The second, so called gaze vector model calculates the gaze orientation directly from the measured distances between the centers of pupils and eye corners, and requires much shorter calibration sessions (i.e. 2-3 second long gazing into all four corners of the visual feedback screen).

The attainable spatial resolution of gaze tracking (i.e., reliable distinction of different gaze directions during robot-assisted rehabilitation) varies among subjects. We limited our tests to gaze classification into 3x3 classes, though more dense distribution of gaze screen was feasible in many of the participants.

As demonstrated in Section IV, feature- or AAM-based facial tracking extends the multivariate patient analysis, by assessing attention to visual feedback as an additional feature to the BNCI performance indices, such as the similarity of motor modules (SMM), kinetic and kinematic profiles and brain patterns. Moreover, the connections between the existing BNCI metrics and the coarse gaze orientation can be assessed, supporting the analysis of impact of different visual feedback elements on gait rehabilitation.

Acknowledgment

The authors are grateful to Teodoro Solis-Escalante, Johanna Wagner and Prof. Gernot R. Müller-Putz from Graz University of Technology and to Rehabilitation Clinic Judendorf-Strassengel (Austria) for organization of video recordings of robot-assisted walking. This work was funded by the Commission of the European Union, within Framework 7, under Call "ICT restoring and augmenting human capabilities compensating reduced motor functions or disabilities", 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".

References

[1] B.Kollen, G.Kwakkel, and E. Lindeman, "Functional Recovery After Stroke: A Review of Current Developments in Stroke Rehabilitation Research," *Reviews on Recent Clinical Trials*, 1, 2006, pp. 75-80.
[2] J.Mehrholz, C. Werner, J.Kugler, and M. Pohl, "Electromechanical-assisted training for walking after stroke,"*Cochrane Database Syst. Rev.*, 17(4), 2007.
[3] M.J.Matarić, J. Eriksson, D.J.Feil-Seifer, and C.J.Winstein, "Socially assistive robotics for post-stroke rehabilitation,"*J.Neuroeng.Rehabil.*, 4(5), 2007.
[4] R.Teasell, and L.Kalra, "What's new in stroke rehabilitation: Back to basics," *Stroke*, 36, 2005, pp. 215-217.
[5] Project BETTER, http://www.car.upm-csic.es/bioingenieria/better/, 2013.
[6] D.W. Hansen, and Q.Ji, "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, 2010, pp. 478-500.

[7] E. Bagherian, and R.W.O.K.Rahmat, "Facial feature extraction for face recognition: a review,"*International Symposium on Information Technology*, Kuala Lumpur, Malaysia, 2008, pp. 1 – 9.
[8] W.K. Liao, D.Fidaleo, and G.Medioni, "Robust, real-time 3D face tracking from a monocular view,"*EURASIP Journal on Image and Video Processing*, Vol. 2010, article ID 183605, 2010.
[9] A. Poole, and L.J Ball, "Eye tracking in human-computer interaction and usability research: Current status and future", *Encyclopedia of Human-Computer Interaction*, C. Ghaouli, Pennsylvania, Idea Group, 2005.
[10] Q.Ji, and X. Yang, "Real-time eye, gaze and face pose tracking for monitoring driver vigilance,"*Real-Time Imaging*, 8, 2002, pp. 357-377.
[11] L. Lang, and H. Qi, "The study of driver fatigue monitor algorithm combined PERCLOS and AECS", *Proc. Int. Conf. on Comp. Science and Software Eng.*, Vol. 1, 2008.
[12] Q.Ji, P.Lan, and C. A. Looney, "Probabilistic framework for modeling and real-time monitoring human fatigue", *IEEE Trans. on Systems, Man and Cyb.*, Vol. 36, Iss. 5, 2006, pp. 862-875.
[13] M. Bakker, F. P. de Lange, J. A. Stevens, I. Toni, and B. R. Bloem, "Motor imagery of gait: a quantitative approach", *Exp Brain Res*, 179, 2007, pp. 497–504.
[14] OpenCV, Open source computer vision library, http://opencv.org/, 2014.
[15] J.Sivic, M.Everingham, and A.Zisserman, "Who are you? Learning person specific classifiers from video," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1145-1152.
[16] G. Loy, and A. Zelinsky, "A Fast Radial Symmetry Transform for Detecting Points of Interest,"*IEEE PAMI*, 25 (8),2003, pp 959-973.
[17] M.Asadifard, and J.Shanbezadeh, "Automatic Adaptive Center of Pupil Detection Using Face Detection and CDF Analysis,"*Proc. of IMECS 2010 conf.*, Vol. I, Hong Kong, 2010.
[18] A.H. Gee, and R. Cipolla, "Determining the gaze of faces in images," *Image and Vision Computing*, 12, 1994, pp. 639-647.
[19] I. Matthews, J. Xiao, and S. Baker, "2D vs. 3D Deformable Face Models: Representational Power, Construction, and Real-Time Fitting,"*Internat. J. of Comput. Vision*, 75(1), 2007, pp. 93-113.
[20] R.Oostenveld, and P.Praamstrac, "The five percent electrode system for high-resolution EEG and ERP measurements,"*Clinical Neurophysiology*, 112, 2001, pp. 713-719.