Proceedings of the 2011 IEEE
International Conference on Robotics and Biomimetics
December 7-11, 2011, Phuket, Thailand

A multi-modal human machine interface for controlling an intelligent wheelchair using face movements

Lai Wei and Huosheng Hu, *Senior Member, IEEE*

***Abstract*— This paper introduces a novel face movement based human machine interface (HMI) that uses jaw clenching and eye closing movements to control an electric powered wheelchair (EPW). A multi-modality HMI derived from both facial EMG and face image information is developed and testified in comparison with a traditional joystick control in an indoor corridor environment. In the experiment, ten repetitive experiments are carried out in a navigation task by controlling an EPW using either face movement control or joystick control. Wheelchair trajectories and execution time during the task are recorded to evaluate the performance of the new face HMI.**

***Index Terms* – Multi-modality, Human machine interface, HMI, face movement detection, eye close detection, jaw clenching, EMG, Intelligent Wheelchair**

I. Introduction

THE development of hands-free human machine interface for elderly and disabled people to operate electric powered wheelchairs is a challenging issue to be solved in a modern society. Over last two decades, many researchers have worked on different multi-modality HMIs by using human biometric information such as human electromyography (known as EMG) [1][2][3], brain wave signals (known as EEG and BCI) [4][5], and human image information [6][7][8][9][10]. These novel approaches and novel HMI systems can contribute greatly to the benefit and well-being of people with disabilities resulted from coma, spinal-cord injury, amputation or ageing.

Compared with single modality HMIs such as either a single EMG controller or single face movement based control method, multi-modality HMIs have the following benefits:

1). Expanded usability and controllability - Integration of multiple modalities capability into a HMI would potentially enrich the controllability of the interface. The user can benefit from deploying alternative communication methods under different situations and environments. Moreover, the extended communication channels can boost the interaction speed and efficiency between machine and human in two directions: from human to machine, human can choose more efficient communication methods, and from machine to human, machine can supply multi-media feedback to indicate about how much information are understood from the user.

2). Refined adaptability and flexibility – A multi-modality HMI can optimize the combination from a number of potential communication methods, and customize itself into a special communication method for individual users. In other words, it provides more choices for people with severe disabilities who have very limited control motions.

3). High accuracy and robustness – Properly fusing various complementary information into a multi-modality HMI could improve the overall performance of the interface, which is much better than traditional mouse and key board. For example, touch screen interface and touch pen interface [13] etc. are a paradigm of MMHMI applications.

As a natural way of communication between people and machine, multi-modality human machine interaction (MMHMI) [11][12] has unique features that are especially helpful for people who have severe disability and have very limited capability in controlling assisted tools such as intelligent wheelchairs and robotic artificial limbs.

However, in order to design a MMHMI, components of the interface need to be fine-tuned to obtain good performance and it is not always that a multi-modality system will certainly outperform single modality ones. There are reasonable factors need to be considered such as tactics in choosing and matching between modalities including human movement pattern or biometric feature. Also, the trade-offs between efficiency against accuracy, flexibility against robustness are important concepts and need to be taken into account carefully.

Meanwhile, intelligent wheelchairs can be classified as autonomous mobile service robots that are a helpful tool for assisting elderly and disabled people with their daily life. Efficiently interacting with such tools will benefit those who lost their motilities; therefore, developing communication interfaces between human and such assisted rehabilitation robots are necessary.

The rest of this paper is organized as follows. Section II introduces an intelligent wheelchair system developed at the University of Essex and its hands-free MMHMI based on human facial movements. In Section III, the signal processing techniques in MMHMI concerning EMG signal

Manuscript received 26 August 2011.
Lai Wei and Huosheng Hu are with the School of Computer Science and Electronic Engineering, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom (e-mail: lwei@essex.ac.uk, hhu@essex.ac.uk).

and image information are briefly explained. Section IV shows the experiment setups and some obtained results. Finally, a brief conclusion and future prospective are proposed in Section V.

II. EXPERIMENT SETUP

A. Intelligent Wheelchair

Fig. 1 shows the picture of our intelligent wheelchair, namely RoboChair, which was built at University of Essex in 2006.

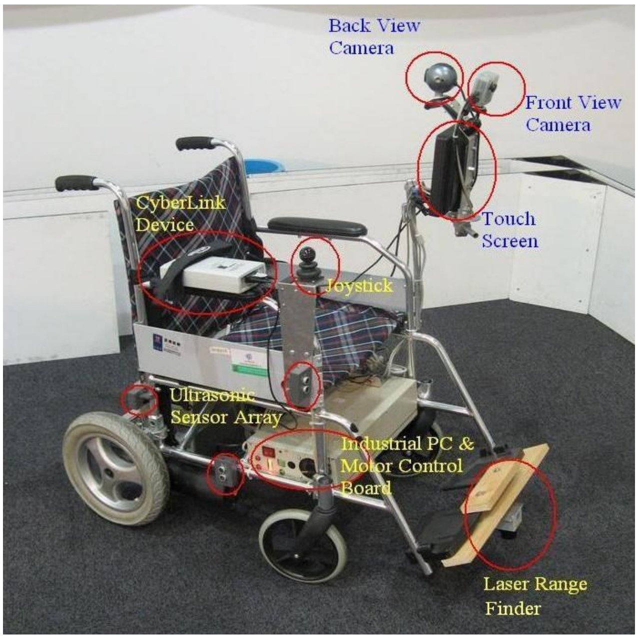


Fig. 1 Intelligent electrical powered wheelchair platform used in the experiment

The wheelchair system has following components:

- ➢ Six ultrasonic sensors at a height of 50 cm (four at the front and two at the back) and one laser range sensor fixed under footplates for obstacle avoidance
- ➢ DSP TMS320LF2407-based PID motor controller for motion control of 2 differentially-driven wheels
- ➢ A local joystick controller to connect to an A/D converter of the DSP-based controller
- ➢ A Logitech 5000 Pro Webcam for recognising the user's facial movements
- ➢ A firmware industrial camera for recording and analysing front view images
- ➢ Intel Centrino 1.8 G dual core CPU with 4 GB physical memory, integrated graphic card and Windows XP installed to analyse the facial images.

B. Definition of Control Movements

As shown in Fig.2, the entire control strategy simulates the direction control from a joystick. Six control commands which are Go Forward (GF), Turn Left (TL), Turn Right (TR), Reduce Speed (RS), Stop (ST) and Go Backwards (GB) are used in this experiment as shown in the fourth column. And five facial control movements as shown in the first column are namely Left Eye Close (LEC) movement, Right Eye Close (REC) movement, Continuous Jaw Clenching (CJC) movement, Single Jaw Clenching (SJC) movement and Double Jaw Clenching (DJC) movement which are connected with these six control commands according to the classified EMG and image patterns (as shown in the second and third column in Fig. 2) abstracted from forehead EMG sensor and facial image camera.

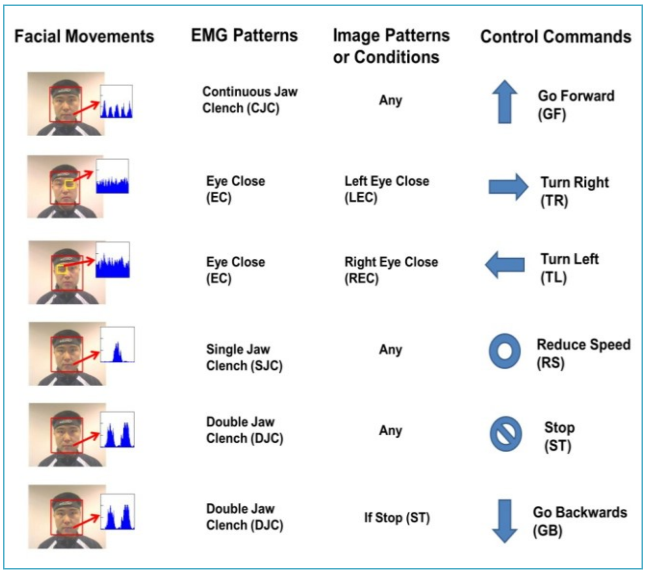


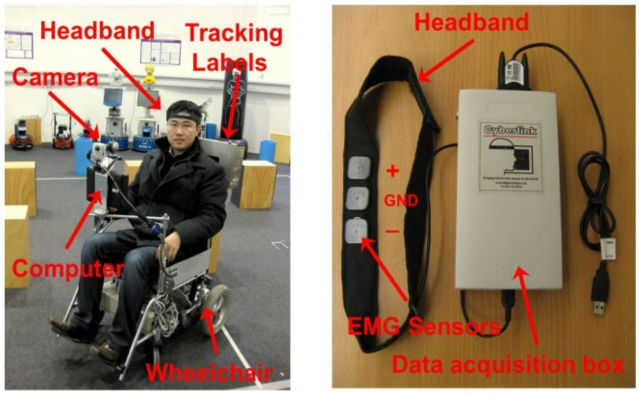
Fig. 2 Face movements used in the experiment and the corresponding wheelchair control commands

As depicted in Fig .2, the relation between control commands and wheelchair movements can be explained as follows:

1) If Left Eye Close (LEC) movement is detected then Turn Right (TR);
2) If Right Eye Close(REC) movement is detected then Turn Left (TL);
3) If Continuous Jaw Clenching (CJC)movement is detected then Go Forward (GF);
4) If Single Jaw Clenching (SJC)movement is detected then Reduce Speed (RS);
5) If Double Jaw Clenching (DJC)movement is detected then Stop (ST);
6) If Double Jaw Clenching (DJC) movement is detected and the wheelchair has stopped then Go Backwards (GB).

C. Human Machine Interface

Fig. 3 (A) shows the wheelchair system, in which the user wears an EMG signal acquisition device Cyberlink*TM* [14]. A Logitech S5500 web camera is mounted at the front of the wheelchair and pointing to the user's face. As shown in Fig. 3(B), the EMG sensing device consists of a data processing box and a wearable headband with three flat attachable electrodes used for detecting facial movements of a user.



(A) (B)

Fig. 3 (A) A subject sitting in wheelchair and wearing Cyberlink Brainfingers^{TM} headband; (B) The Cyberlink Brainfingers^{TM} data acquisition box and the headband

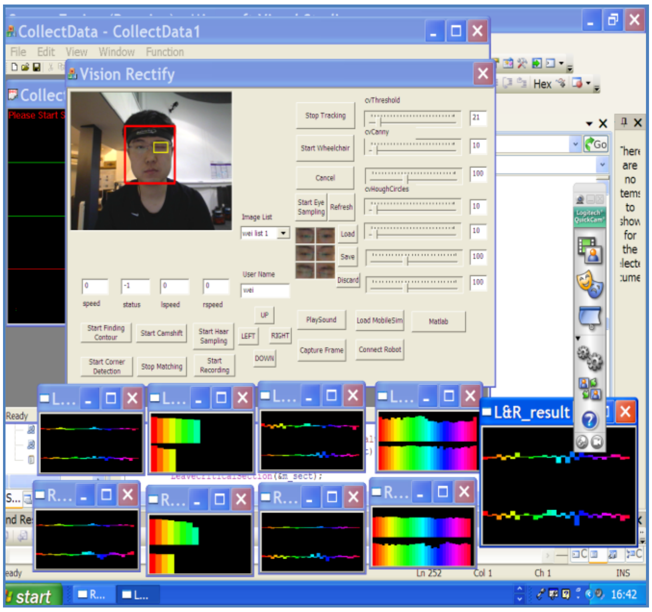


Fig. 4 Layout of the multi-modality human computer interface developed in the experiment

The electrodes are attached to human forehead by fastening the headband. From three electrodes, one channel EMG signal can be obtained with one positive electrode, one negative electrode and a reference electrode. The raw EMG amplified from the electrodes are then processed in data acquisition box and turned into digital averaged EMG data sequence with a sampling frequency of 100Hz

As shown in Fig. 4, the user interface consists of an interface that shows the classification results from user face movements. As shown in real-time user image steamed from the camera, an eye closing movement is successfully detected by the interface, and a yellow frame is shown to indicate the detected closing eye, and accordingly, at this time, a TR command is given to wheelchair motor controller and the wheelchair will turn right according to the control logic in Fig 2.

III. SIGNAL PROCESSING TECHNIQUES

As in this experiment, two modalities, i.e. EMG signal from user forehead and face image information from camera fixed in front of the wheelchair, are synthesized to analyze human facial movement features. Subsequently, two types of data, i.e. image pixel matrix data and EMG waveforms will be processed according to their own features. As the movement features in both modalities are steady and distinctive, we apply pattern recognition algorithms separately to each modality for feature and pattern classification purpose, and thereafter the classification results from both modalities are combined according to a combination of logic rules depicted in Fig 2. This procedure is depicted as a feature fusion in [2]. In the following paragraphs, we will briefly introduce the signal processing and classification procedures with the two types of modalities' data.

A. Forehead EMG data processing

In order to classify four EMG waveform patterns namely Continuous Jaw Clench (CJC), Eye Close (EC), Single Jaw Clench (SJC) and Double Jaw Clench (DJC) as shown in first and second column of Fig. 2 resulted from five control movements, a threshold based classifier is designed and a support vector machine (SVM) classifier [15] is trained. As explained in [3] [16], the threshold based classifier uses a threshold and time span triggering approach to detect patterns such as SJC and DJC.

For the other EMG patterns including CJC and EC, a SVM based classification approach is applied to classify features from both time and frequency domains after data segmentation of EMG waveform. Four features from time domain are mean absolute value (MAV), root mean square (RMS), waveform length (WL) and zero crossings (ZCs), and two from frequency domain are the mean and median of signal frequencies (FMN and FMD).

B. Face Image processing

As shown in Fig. 4, the interface shows a detected user face and the closed eye from image sequences from a camera. Here, the function of the image interface is to detect human frontal face while user is sitting in the wheelchair, and also, from the detected face area, the eye closing movements are also detected. This requires an image classifier that can detect closed left and right eyes. Two classifiers based on Haar-like features using an adaptive Boosting (AdaBoost) learning approach [17][18] are trained for detecting left-closed and right-closed eyes. The training is based on sampling closed-eye images from 5 subjects from different race groups under various illumination conditions, subject face orientations and distances relative to sampling camera.

Boosting is a machine learning algorithm that constructs an accurate classifier with a series of regressive weak learners. A weak classifier (or learner) is a division of sample sets with recognition accuracy slightly over random. The training space is divided by those weak learners via a probabilistic deducing method, and the generalized performance with designed reasoning procedure can divide sample spaces according to those "rules", and comes to a deterministic result known as a strong classifier.

This learning method is similar to support vector machine which project nonlinear sample training sets into a hyperspace. While in AdaBoost, the projecting is substituted with a Haar feature projecting, which can best represent an image data with a set of ambient texture features, therefore, the image data which stored in forms of matrix pixel digits can be best represented in terms of feature searching, computing, and feature extracting.

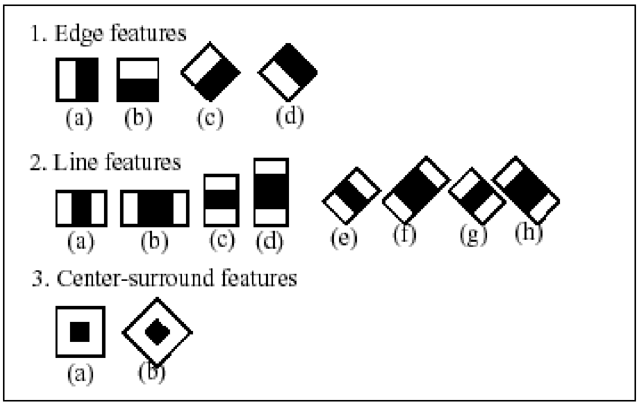


Fig.5 Feature prototypes of Haar-like and centre-surround features. Black areas have negative and white areas positive weights [18].

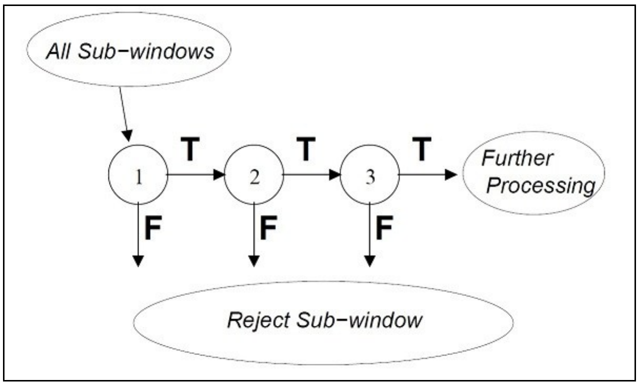


Fig.6 Structure of detection cascades, all target sub-windows are tested through cascade 1, 2, 3 and so on, and cascade 1 classifier trained with the most salient feature eliminates largest number of sub-window examples [17].

A cascade of classifiers trained from overlapped window area from the target image sample is trained with both eye samples. Before training, image samples are collected and transformed into a series data form namely integral image defined in [17], and thus Harr features which can be mathematically depicted with Haar wavelets, can be represented by subtracting equally divided axial symmetrical sub regions in an image area. Fig. 5 shows different forms of Haar-like features, here we can find that closed eye image can be best represented with longitudinal texture features, therefore distinctive features of a closed eye are mainly 1.(b), 2.(c) and 2.(d) in Fig.5.

The function of an AdaBoost image classifier is very similar to a two dimensional scalable bar code reader. AdaBoost combined the Boosting concept in a cascade structure, where for the convenience of regression, the algorithm aims to find Haar-like features with a hierarchical accepting and abandon rules (shown in Fig. 6), where the most ambient feature of Boosting classifier is tested, trained and applied to target image first, this approaches can reject negative image area without having to traverse all possible Haar feature based cascades. Each cascade in AdaBoost consists of a salient feature and the cascades are organized into a series structure according to their significance generalized with probabilistic deducing procedure in learning process.

In actual training, six male subjects from five different ethics backgrounds (Italian, British, Chinese, Nigerian, Spanish and Iranian) are involved. From each subject with either left or right eye, an amount of around 200 positive images and 650 negative images are collected. And the positive or negative images of all six subjects are mixed together with respect to left eye or right eye, and grouped into either positive or negative image set for closed eye classifier training. The total amount of images used in left closed eye classifier training are 1475 positive images with image size range from 36 by 23 pixels to 65 by 44 pixels and 4286 negative images with an image size range from 28 by 24 pixels to 124 by 76 pixels. For the right closed eye training, the positive set contains 1318 image samples with a size range from 36 by 23 pixels to 72 by 48 pixels and the negative set have an collection of 3668 background images with an image size range from 38 by 24 pixels to 124 by 76 pixels.

IV. Experimental Results

Fig. 7 presents the experiment site used for this research, which consists of a combination of doorways, corridors, turning corners and docking places. In order to control the wheelchair going from the start point to the target area, the subject has to control the wheelchair to go through three corridors and make two major turnings with one left turning and one right turning each, and finally passes through a doorway to the target area for docking.

As shown in Fig. 7, the geometry size of the wheelchair is 650 millimetres (width) by 1000 millimetres (long), and the width for all the corridors and passages are 1200 millimetres. The overall size of the map is 5.6 by 4.1 metres, and the passage is separated evenly by two rows of six wooden boxes.

Fig. 8 shows that the subject controlled the wheelchair with a joystick, travelling through the designed experiment site from the start point to the target area for then times, following a planned route designed in Fig. 7. Ten trajectories labeled from “Trail 1” to “Trail 10” are drawn in different colors.

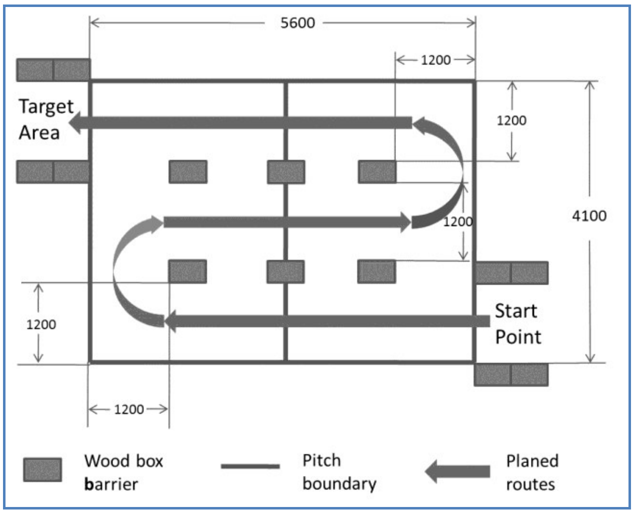


Fig.7 Diagram showing the dimensions of the experiment layout and the planed routes in the task

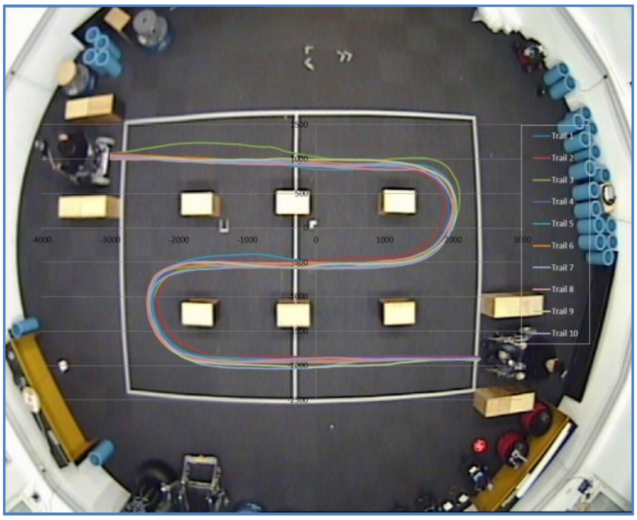


Fig. 8 Ten successive trajectories made by a subject controlling an intelligent wheelchair with joystick and following a planned route form start point to target area for ten times

While in Fig. 9, another ten trajectories result from the new control method are drawn, in which the subject controlled the wheelchair following the designed route in the map using different control methods. The trajectory of each trail is recorded and ten trajectories are drawn together in the map.

As can be seen from the recorded trajectories, the joystick control result is smooth and regular, especially the subject can maneuver well in the turning, and trajectories are smooth in following a corridor with the joystick. The difference between each trajectory is small, which means the joystick control is steady and easier for the subject to use and also less time was required in accomplishing designed task. It is clear that the joystick control outperforms the proposed control method.

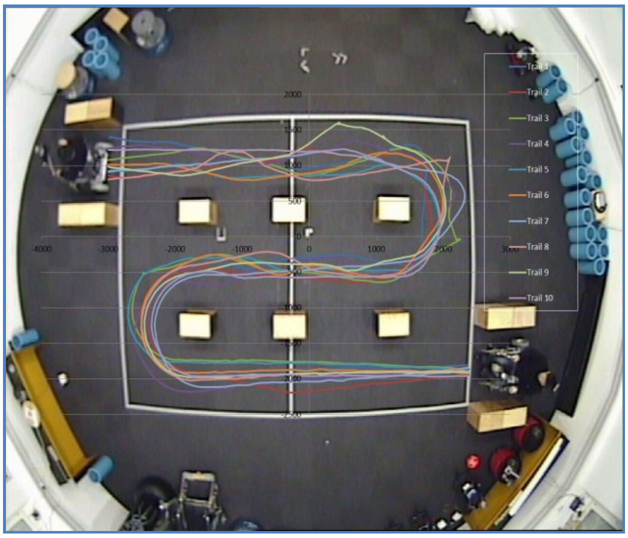


Fig. 9 Ten successive trajectories made by a subject controlling an intelligent wheelchair with EMG and vision based control method and following a planned route form start point to target area for ten times

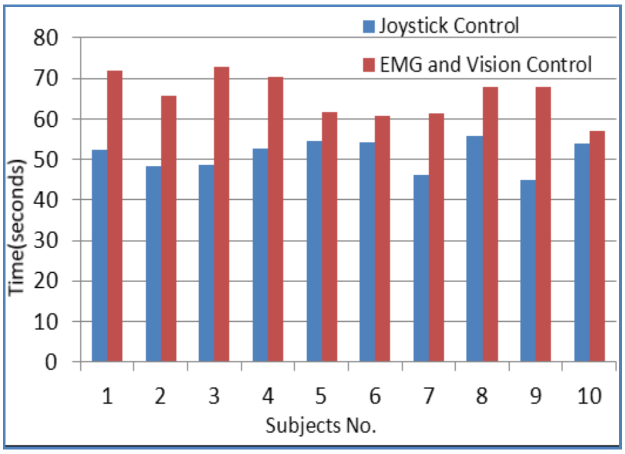


Fig. 10 Travel time comparison between joystick control and hybrid (EMG and vision) control

In Fig. 10, the difference of two control methods are analysed in terms of travelling time. As we can see, the time spent in the joystick control has a lower mean value, and a smaller variance compared with the hybrid (EMG and Vision) control method, which means the joystick control is faster and stable for the capable subject to use. The trajectories based on the joystick control are smoother than the trajectories achieved by the proposed hands-free control method. Also subject feel the performance of the new control method can be improved by further training and adaptation from user. Therefore, further improvement of the designed control method is required so that it can be used reliability by disabled and elderly people who have difficulty to use their hands.

V. Conclusion and Future Work

In this paper, a novel multi-modality human-machine interface is proposed for elderly and disabled people to achieve hands-free control of electric powered wheelchairs. The developed HMI integrates both facial image information and facial EMG signals to recognize human intentions for the wheelchair control. The mechanism combining multi-modal sensory data from both facial EMG signals and face image information has been investigated for handling uncertainty and changes in the experiment. The designed control system is implemented on a wheelchair operated in a corridor based indoor environment with docking stations, walls, turnings, corners and obstacles. Experimental results show the feasibility and great potentials of the proposed multi-modal control method, and prove to be a good alternative control interface to traditional joystick.

In the future, different combinations of control methods will be further investigated to form various control strategies in order to have a more complete comparison. Also, experiments on a larger number of subjects will be implemented in the real-world in order to further carry out performance evaluations between single and multi-modality control methods.

Acknowledgments

This research has been financially supported by EU Interreg IV A 2 Mers Seas Zeeën Cross-border Cooperation Programme – SYSIASS: Autonomous and Intelligent Healthcare System. More details can be found from the project website http://www.sysiass.eu/. We would also like to thank Robin Dowling for the technical support during our experiments.

References

[1]. T. Felzer and B. Freisleben, "HaWCoS: The 'Hands-free' Wheelchair Control System," Proceedings of ACM SIGACCESS Conference on Computers and Accessibility, pp. 127–134, ACM Press, 2002

[2]. Moon, M. Lee, J. Chu, and M. Mun, "Wearable EMG-based HCI for electric-powered wheelchair users with motor disabilities," Proc. of IEEE Int. Conf. on Robotics and Automation, pp. 2649–2654, 2005.

[3]. M. Asghari Oskoei and H. Hu, "Support Vector Machine-based Classification Scheme for Myoelectric Control Applied to Upper Limb", IEEE Transactions on Biomedical Engineering, vol. 55, no. 8, pp. 1956-1965, Aug. 2008..

[4]. B. Rebsamen, E. Burdet, C. Guan, H. Zhang, C.L. Teo, Q. Zeng, C. Laugier, Jr. Ang and H. Marcelo, "Controlling a Wheelchair Indoors Using Thought," IEEE Intelligent Systems, Vol. 22, No. 2, pp. 18–24, 2007.B. Simpson, et al, "Title of paper goes here if known," unpublished.

[5]. R. Millan, F. Renkens, J. Mourino and W. Gerstner, "Noninvasive brain-actuated control of a mobile robot by human EEG," IEEE Transactions on Biomedical Engineering, Vol.51, No.6, pp.1026-1033, June 2004.

[6]. I. Iturrate, J. M. Antelis, A. Kübler and J. Minguez, "A noninvasive brain-actuated wheelchair based on a P300 neuro-physiological protocol and automated navigation," IEEE Transactions on Robotics, Vol.25 No.3, pp.614-627, June 2009.

[7]. D. Santis and D. Iacoviello, "Robust real time eye tracking for computer interface for disabled people," Computer Methods and Programs in Biomedicine, Oct. 2009

[8]. P. Jia, H. Hu, T. Lu and K. Yuan, "Head gesture recognition for hands-free control of an intelligent wheelchair," Journal of Industrial Robot, Vol. 34, No. 1, pp. 60–68, 2007.

[9]. J. Colombo, A. Del Bimbo and S. De Magistris, "Human-computer interaction based on eye movement tracking," International Workshop on Computer Architectures for Machine Perception, pp. 258, 1995.

[10]. P. Smith, M. Shah and N. da Vitoria Lobo, "Determining driver visual attention with one camera," IEEE Transactions on Intelligent Transportation Systems, vol.4, no.4, pp. 205-218, Dec. 2003.

[11]. R. Sharma, V. Pavlovic and T. Huang, "Toward multimodal human-computer interface," Proc. of the IEEE, Vol.86, No. 5, pp. 853-869, 1998.

[12]. T. Jaimes and N. Sebe, "Multimodal human-computer interaction: A survey," Computer Vision and Image Understanding, Vol. 108, No. 1-2, pp. 116-134, October 2007.

[13]. S. Oviatt, "User-Centered Modeling for Spoken Language and Multimodal Interfaces," IEEE MultiMedia, Vol. 3, No. 4, pp. 26-35, Dec. 1996.

[14]. Cyberlink Brainfingers, Hands-free Computer Access Solution < http://www.brainfingers.com/>

[15]. C.C. Chang and C.J. Lin, LIBSVM: a library for support vector machines, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[16]. L. Wei and H. Hu, "EMG and Visual based HMI for hands-free control of an intelligent wheelchair," Proc. of the 8th World Congress on Intelligent Control & Automation (WCICA), pp. 1027-1032, Jinan, China, July 2010.

[17]. P. Viola and M.J. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001.

[18]. R. Lienhart and J. Maydt, "An Extended Set of Haar-like Features for Rapid Object Detection," IEEE ICIP 2002, Vol. 1, pp. 900-903, Sep. 2002.