A hybrid human-machine interface for hands-free control of an intelligent wheelchair

Lai Wei and Huosheng Hu*

School of Computer Science and Electronic Engineering,
University of Essex,
Wivenhoe Park, Colchester CO4 3SQ, UK
E-mail: lwei@essex.ac.uk
E-mail: hhu@essex.ac.uk
*Corresponding author

Abstract: This paper presents a novel hybrid human-machine interface (HMI) designed for hands-free control of electric powered wheelchairs. Both forehead electromyography (EMG) signals and colour face image information are deployed to identify winking and jaw clenching movements of human face. Five winking and jaw clenching movement patterns are selected and classified, mapping into six control commands to drive an electric powered wheelchair in an indoor environment. Six subjects participated in the experiments and the experimental results show that the proposed control scheme have potential applicability to accommodate various individual cases and achieve good performance.

Keywords: human-machine interface; HMI; face detection; eye detection; electromyography; EMG; SVMs; electric wheelchair; computer vision; boosting; multi-modality.

Reference to this paper should be made as follows: Wei, L. and Hu, H. (2011) 'A hybrid human-machine interface for hands-free control of an intelligent wheelchair', *Int. J. Mechatronics and Automation*, Vol. 1, No. 2, pp.97–111.

Biographical notes: Lai Wei is currently a PhD student in Human-Centred Robotics Group at the University of Essex. He received his BEng in Measurement and Control from the Hebei University of Technology in 2004. His research interests include human computer interface, signal processing and pattern recognition, mobile robots, image processing and multi-sensor systems.

Huosheng Hu is a Professor in the School of Computer Science and Electronic Engineering at the University of Essex, leading the Human-Centred Robotics Group. His research interests include behaviour-based robotics, human-robot interaction, embedded systems, mechatronics, learning algorithms, pervasive computing and service robots. He has published over 300 papers in journals, books and conferences in these areas, and received a number of best paper awards. He is one of founding members of IEEE Robotics and Automation Society Technical Committee on Networked Robots, a Fellow of IET and InstMC, and a senior member of IEEE and ACM. He has been Chair or committee member for many international conferences such as IEEE ICRA, IEEE IROS, IEEE ROBIO, IEEE ICMA and IEEE ICIA conferences. He is currently the Editor-in-Chief for *International Journal of Automation and Computing*.

1 Introduction

For decades, a key issue that lies in science and engineering has been how human could interact with a computer or machine in a natural way. Until now, the most successful and widely used interfaces are keyboard and mouse through which users can operate a computer or control various devices that are connected to a computer. However, for people with sever physical disabilities such as spinal cord injury, quadriplegia, hemiplegia or amputation, keyboard and mouse are no longer adequate. It is necessary to develop novel human-machine interfaces (HMIs) for disabled and elderly people to use computers and robots for the better quality of life in the society.

Human face is a remarkable place for studying human machine interaction since a wide range of facial characters (emotion, voice, eye movements, mouth movements, etc.) can be potentially deployed for people interacting with a computer or a robot. The utilisation of visual face detection techniques such as AdaBoost with Haar-like features (Viola and Jones, 2001; Lienhart and Maydt, 2002) has formed a foundation work for analysing face expression or face movements. Many facial-based HMIs have recently developed for people to interact with computers and robots, including eye moment interface (Santis and Iacoviello, 2009; Colombo et al., 1995; Smith et al., 2003), head gesture interface (Jia et al., 2007; Hazlett and Benedek, 2007), lip movement interface

(Dupont and Luettin, 2000; Chen, 2001), etc. Apart from visual information, electromyography (EMG) has recently proven to be an effective tool for analysing human movements and human intentions through human muscle activities. Due to its effectiveness, a number of active research on EMG-based HMIs are carried out on different muscle groups of human body such as limb (Chu et al., 2007; Oskoei and Hu, 2008; Shenoy et al., 2008), neck (Barniv et al., 2005; Moon et al., 2005), shoulder (Kiguchi et al., 2008), face (Felzer and Freisleben, 2002; Wei and Hu, 2008), etc.

However, single modality of HMI, either facial image-based or EMG-based, may not work well since each of them observes only partial information of face movements. Vision systems may not function at all due to bad illumination conditions. EMG signals are susceptible to crosstalk noises from other muscles. Also, there are limited movement patterns on the face muscles that can be detected by EMG transducers. Therefore, it is necessary to study how to effectively combine facial EMG signals and face image information for handling uncertainty and changes that single modality cannot handle, which motivates to develop a novel hybrid HMI integrating both facial images and facial EMG signals to recognise human control intentions. The experiments are conducted by implementing this novel interface on a simulated robot and an electrical powered wheelchair to verify the feasibility.

The rest of this paper is organised as follows. Section 2 gives a brief introduction of the system architecture and experimental settings. Section 3 introduces the method for EMG feature extraction and classification and their application in this research. In Section 4, facial image processing is briefly explained, including AdaBoost and Haar-like features being used. Section 5 presents a hybrid control strategy for hands-free control of an intelligent wheelchair, which integrates facial image recognition and EMG pattern recognition. The experimental results from the implementation of the designed control strategy in both a simulated and an indoor environment are presented in Section 6 to verify the feasibility and performance of the proposed hybrid control scheme. Finally, a brief conclusion and further research prospective are given in Section 7.

2 Facial control movements

In this research, five facial movements are defined to control an electric powered wheelchair, namely left-eye-close (LEC), right-eye-close (REC), continuous-jaw-clenching (CJC), single-jaw-clenching (SJC) and double-jaw-clenching (DJC). They are from either jaw clenching (a chewing-like) movement or eye close (a winking-like) movement. The five control movements are mapped into six control commands to control orientation and velocity of a wheelchair.

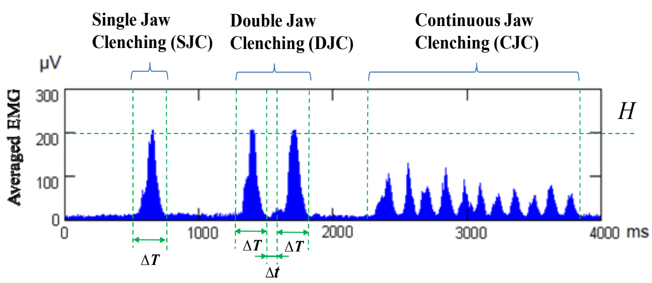
2.1 Jaw clenching (chewing-like) movement

2.1.1 SJC, DJC and CJC movement description

As shown in Figure 2, three jaw clenching movements are generated by a subject contracting masseter muscle and buccinators muscle with a jaw clenching or a chewing-like movement. Figure 1 shows averaged EMG waveforms (a transformed waveform from raw EMG signal data, detailed steps explained in Section 3) during three jaws clenching movements named as single jaw clenching movement (SJC), double jaw clenching movement (DJC) and continuous jaw clenching movement (CJC). These clenching movements are generated based on subjects holding up their upper and lower teeth, pushing the jaw against the upper teeth and making a chewing-like movement without separating the teeth. By doing this way, a subject could feel free to control the chewing speed and the level of tension in movement muscles.

CJC can be produced by a subject making alternative gentle and fast jaw clenching. As shown in Figure 1, this chewing-like movement can cause a series of short and regular EMG waves. Compared with CJC, SJC is a short and strong clenching movement, and DJC is a movement composed of two SJC in a fixed period of time. For making DJC, a subject needs to make two successive SJC in a quick manner. Since SJC involves major muscle contractions from masseter muscle, which is the strongest muscle of face, this short clenching movement will result in a spike shaped EMG waveform with a peak EMG amplitude value. We make use of this EMG feature and characterise SJC and DJC on an EMG waveform with maximum EMG amplitude and recognise them by a threshold-based strategy.

Figure 1 SJC, DJC, CJC patterns in averaged EMG waveform (see online version for colours)



2.1.2 SJC, DJC and CJC recognition

Felzer and Freisleben (2002) defined two '*contraction events*' and recognised them with adjustable thresholds applied on EMG amplitudes. Similar to their work, here we pick up SJC and DJC evens based on adjustable amplitude thresholds too. Figure 1 shows parameters by which SJC and DJC are identified.

SJC is recognised by timing how long the EMG waveform can make continuing overflows over a pre-set threshold value H in a fixed time ΔT. The detailed calculation process can be described as: if a continuous overflow in the waveform with a time span T that is longer than 30 milliseconds and shorter than 100 milliseconds is detected, based on T, we define an extension of 50 milliseconds before and after T as ΔT, and calculate the average of EMG amplitude within time ΔT. The result is compared with a pre-set threshold to avoid too much overflows in time ΔT. If the result is smaller than the pre-set threshold (600 μV in this experiment), then we say SJC is detected. Similarly, DJC event can be detected as finding two successive SJC evens in a fixed time span $2\Delta T + \Delta t$, as depicted in Figure 1.

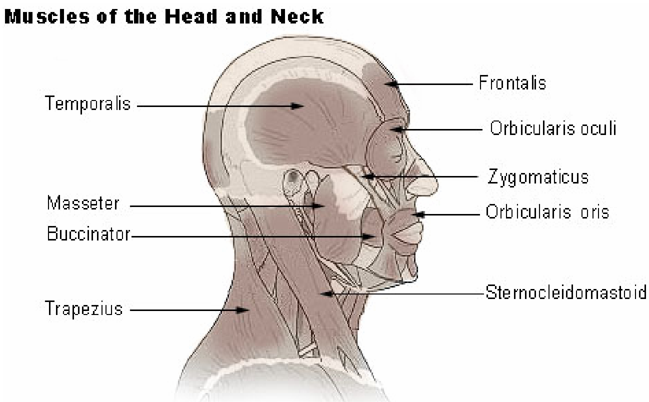
The waveforms generated by CJC movements (class CJ) contain a combination of more complex features in both time and frequency domains, and a SVM-based classification procedure is applied to feature data extracted from both time and frequency domains. The detailed procedure is discussed in Section 5.

2.2 *Eye close (winking) movement*

2.2.1 *LEC and REC movement description*

The eye close movements employed in our control strategy are called left eye close (LEC) and right eye close (REC). Figure 4 shows picture samples when a subject making LEC and REC movements that are recognised by Haar classifier based on face image data. Form Figure 4(b) and 4(c), we can see that the eye close movements used in this experiment are a kind of heavy (squeeze) eye close movements that can be generated by keeping one eye open and closing and squeezing the other eye. The squeezing movement involves contractions mainly from three muscle groups around eyes, which are frontalis, orbicularis oculi and tempolaris muscles as shown in anatomy diagram in Figure 2.

Figure 2 Anatomy diagram of human head and neck muscles (see online version for colours)



Source: SEER Training Modules (2008)

2.2.2 *LEC and REC recognition*

In this research, LEC and REC movements are identified by a hybrid strategy applied on both EMG features and face image features. As described above, the eye close movement designed contains a squeezing action. This squeezing action can fully activate the muscle around the closed eye and the tensed muscles can generate detectable EMG signals which can be easily detected by EMG electrodes, Figure 3 shows the averaged EMG waveform sampled during a subject is making LEC and REC movements. The closed and squeezed eye movements make LEC and REC distinguishable from normal eye close movements and can avoids noise caused by normal eye blinking in particular. The details about their classification are presented in Sections 5 and 6.

Figure 3 Average EMG waveform sampled during a subject making LEC and REC movements (see online version for colours)

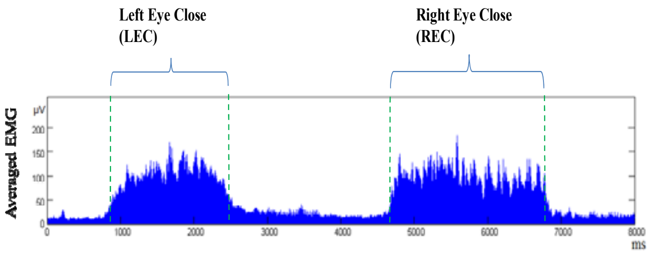
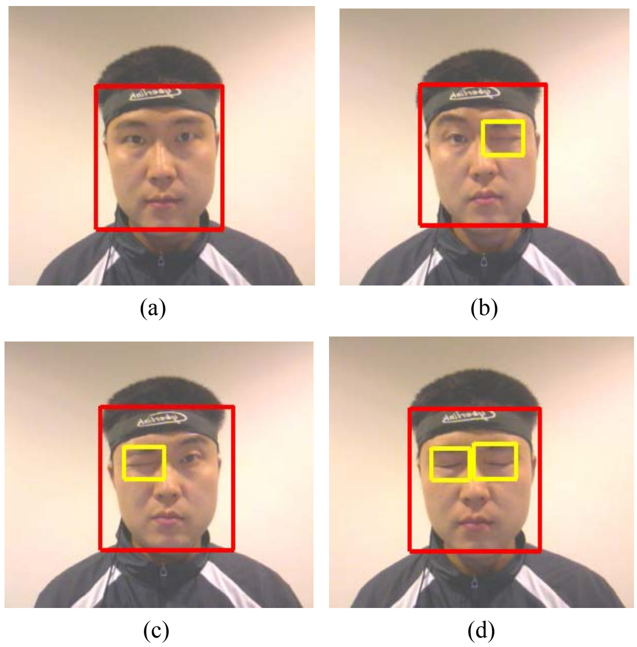


Figure 4 Eye close movements detected by a cascade of boosted classifier based on face image, (a) no eye close is detected (b) LEC is detected (c) REC is detected (d) both LEC and REC are detected (see online version for colours)

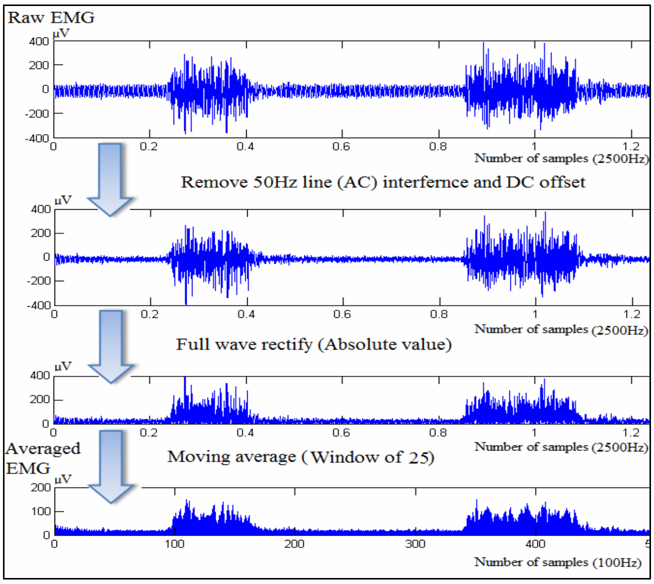


3 EMG feature extraction and classification

3.1 *Raw EMG pre-processing*

The EMG data used for the feature extraction purpose is called the averaged EMG; Figure 5 displays 3-step processes transforming raw EMG signal into an averaged EMG waveform. The first row shows a number of 2×10^4 raw EMG data sampled in eight seconds with a sampling rate of 2,500 Hz. These data are firstly processed by passing through a 50 Hz notch filter to remove inline 50 Hz AC noise, and secondly passes through a high pass filter to remove DC offsets. Finally, a clean EMG waveform with AC noise and DC offset removed is obtained in the second row of Figure 5. After that, the absolute value is calculated for all the processed sampling data in the second row, the result is called full wave rectified EMG data which is drawn in the third row.

Figure 5 Process of 3-step transformation from raw EMG data to averaged EMG waveform (see online version for colours)

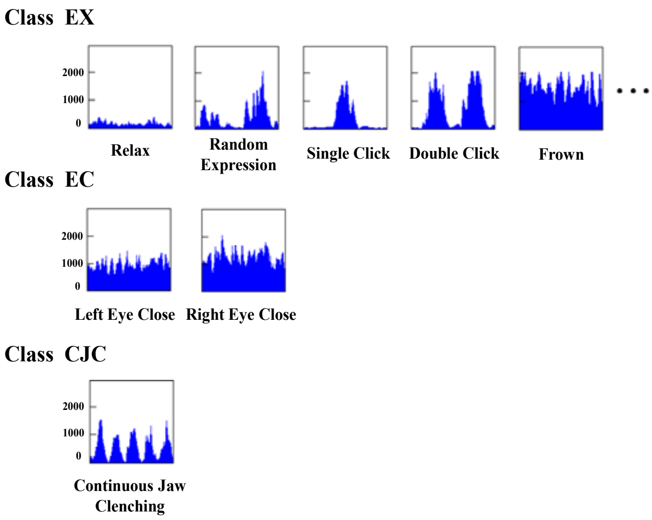


From the full wave rectified EMG data with a sampling rate of 2,500 Hz, for every 25 sampled data, an average value of these data is calculated and the result is kept as a new EMG sampling data, namely the averaged EMG data shown in the fourth row of Figure 5. In this process, the sampling rate is downgraded by 25 times to 100 Hz and the glitch noises are subsequently removed in this process. After steps all above, in the last row, an averaged EMG waveform with a 100 Hz sampling rate is shown and this is the unique form of EMG data used in this experiment to detect five designed control movements.

3.2 *EMG pattern combination*

For the five designed control movements, SJC and DJC patterns can be recognised by a threshold-based strategy, and also we can distinguish LEC from REC using face image features. Three movements LEC, REC and CJC can be divided into two groups in EMG feature classification. Since LEC and REC movements are distinguishable from each other with face image features, we combine LEC and REC patterns into one class namely EC in EMG feature classification strategy, and at the same time name another class containing only CJC pattern as class CJ. Therefore, as shown in Figure 6, the whole EMG waveform set is divided into three subset classes, namely EX, EC and CJ.

Figure 6 Grouping EMG patterns into classes for SVMs-based classification, movement patterns are displayed in 500 millisecond windows (see online version for colours)



Theoretically, EC class should include all the EMG patterns generated by LEC and REC movements, and CJC class should contain all EMG patterns produced by continuous jaw clenching movements. Class EX contains all the other existing EMG waveform patterns that are not included in classes EC and CJ. As shown in Figure 6, EX class contains a series of patterns such as patterns from SJC and DJC movements, patterns generated by random facial expressions, frowning, relaxing, etc.

3.3 *Training data sampling*

In the process of adopting EMG sample data sets for training a classifier, a well trained and experienced person is involved and considered as a standard subject who can perform control movements with little error, and can produce waveforms that can well represent the differences between waveform patterns in the feature space. In other word, a SVM classifier is trained here by sampling data from a standard subject, and then these standard movements can be taught and copied to different subjects later. This strategy is rational since humans have good adaptability by learning and imitating from others. The simulation results of this strategy are given in Section 6

In this research, training data are sampled according to the three divided classes displayed in Figure 6; the sample data are recorded by a standard subject doing different movements for each divided classes EX, EC or CJC.

- For class EX, data are sampled when the subject is relaxing, frowning, making single or double clicking movements, making random face expressions (such as talking, smiling or peering, etc.). All these actions are sampled within 160 seconds and generate 800 training vectors after data segmentation and feature extraction.
- For class EC, waveform data are recorded when the subject doing LEC and REC movements, and a total of

80 seconds data for each movement are recorded. The recorded data are mixed, making a training set that contains 800 training vectors after data segmentation and feature extraction.

- For class CJC, training data are sampled by the subject doing CJC movements for a total length of 56 seconds, which generates 280 training vectors after data segmentation and feature extraction.

During sampling training data, the subject will keep doing one movement each time for less than eight seconds and then relax their muscle; this is to prevent the subject from muscle fatigue.

3.4 EMG data segmentation and feature extraction

In order to train and apply the trained EMG classifier, EMG data is divided into data segments and each data segment is represented by some extracted features. In this experiment, EMG data is divided into 200 ms disjoint data segments and for each data segments, a series of six features from both time and frequency domains are extracted, four of which from time domain are mean absolute value (MAV), root mean square (RMS), waveform length (WL) and zero crossings (ZCs), and another two from frequency domain are the mean and median of signal frequencies (FMN and FMD).

Before training classifiers, by extracting features from each segment, a vector is constructed using six feature values and a number representing which classes the data segment is from. Figure 7 shows the procedure and function of data segmentation and feature extraction in the whole classification process that contains two processes, i.e., training and application. In both processes, the same technique of data extraction and segmentation are applied. While in the training process, a SVM classifier is trained using training data sampled from divided movements listed in Figure 6. In the application process, same data segmentation and feature extraction process are applied in real-time to live EMG data.

3.5 SVMs training and classification results

Support vector machines (SVMs) are a kernel-based classification method widely used for classification and regression in various applications. Compared with neural networks and other classification methods, it has a concise model structure and simpler testing parameters and a faster regression speed in training. The application of SVMs into classifying EMG data is widely introduced and has been proven to have better general performance in classifying multi-channel EMG data with complex features (Chang and Lin, 2001). In this paper, one channel EMG data with three classes are to be classified. As described in Figure 7, after data segmentation and feature extraction, the training data are represented as a vector composed of six extracted features and one label indicating which of the three classes the data are from.

After the data sampled from the designed movements, they are divided into segments with a class label, the extracting time and frequency domain features from each segment. Three classes of training vectors have been gathered. As described in Section 3.3, we have 800 training vectors labelled with class EX, 800 training vectors labelled with class EC and another 280 training vectors labelled with class CJ. For training a SVM classifier, RBF kernel is selected as the machine kernel as it can give general quick and reliable classification results. In training a SVM classifier with a training data set that contains a total number of 1,880 training vectors, a threefold cross validation strategy is applied to prevent over-fitting. In order to find the best combination of parameters in a RBF kernel, a grid search based on a combination of RBF kernel parameters (C, γ) pairs is conducted.

Figure 7 SVM classification process for EMG data

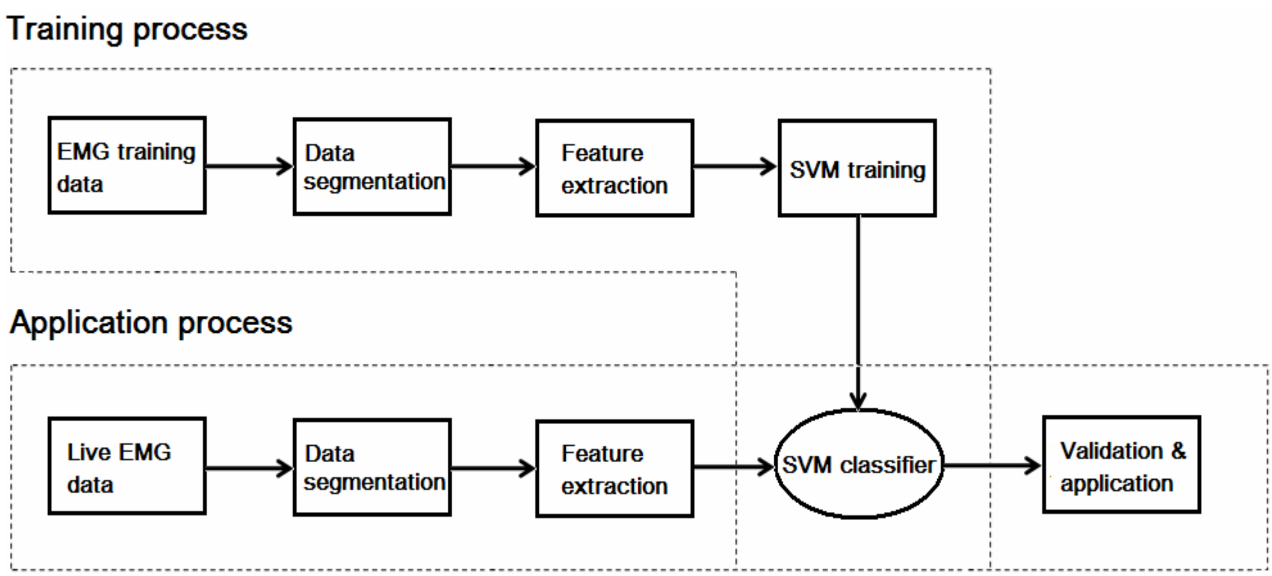


Figure 8 shows the grid search results of the three classes in a 3D graph. The vertical axis represents the classification accuracy value according to (C, γ) value pairs, in which C is a penalty parameter and γ is a value from SVM radial basis function (RBF) (Hsu et al., 2003). The bottom two axes are log(C) and log(γ). We have:

$$K\left(x_i, x_j\right) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right), \gamma > 0. \quad (1)$$

where $K(x_i, x_j) \equiv \phi(x_i)^T \phi(x_j)$ is called the kernel function.

Figure 8 SVM Grid search results over EMG training data (see online version for colours)

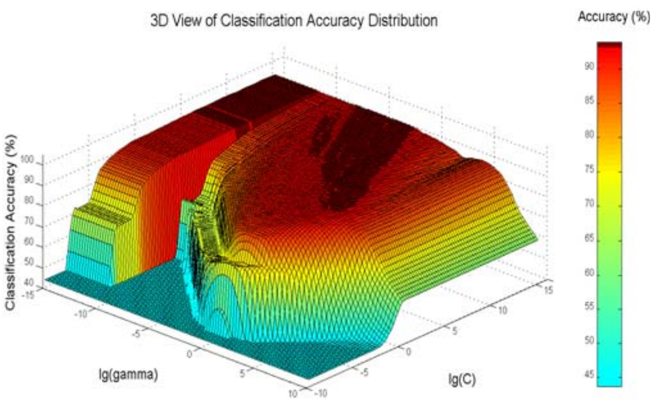


Figure 9 Top view of SVM grid search result over EMG training data (see online version for colours)

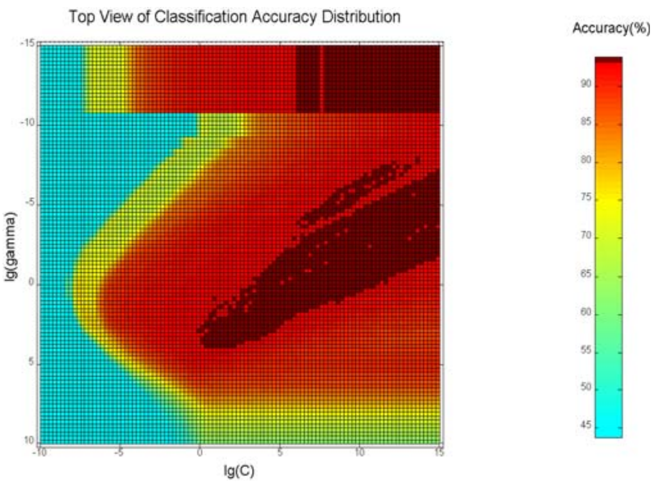


Figure 9 shows the grid search result with log(C) from –10 to 15, log(γ) from –15 to 10 with an incremental value of 0.25 for both log(C) and log(γ). The dark grey area shows log(C) and log(γ) combinations where classification accuracy are above 93%. The best training result can be found at a log(C) value of 13 and a log(γ) value of –0.5 where the maximum classification accuracy is 93.75%. From the training and cross validation result, we therefore apply this trained kernel with a classification accuracy of 93.75% as the SVM classifier for classifying real-time EMG data in application process depicted in Figure 6 and to help recognise the five designed movements in a designed recognition and control strategy illuminated in Section 5.

4 Facial image processing

AdaBoost is a machining learning algorithm that can learn a strong classifier by organising a set of weak classifiers in a cascade structures. Since Haar-like features (Viola and Jones, 2001) have good representation for comparing differences in images, Haar-like feature-based AdaBoost classifiers can deal with minor image rotation, arbitrary scaling and salient colour or texture changes. These features can well satisfy the requirement of this experiment as the closed eye classifier needed here is dedicated to a group of six testing subjects in an indoor environment.

However, it can be computational expensive when searching and calculating Haar-like features in a huge data set for classification. This is particularly true when searching in a pixelated colour image area. Since the best way to reduce computation load is to limit search area, a segmentation procedure is applied based on the face region detected and divided from the whole image. As a result of the segmentation, two specific AdaBoost classifiers for detecting left closed eye and right closed eye are trained and applied to their own segmentation areas.

4.1 Face image segmentation

To avoid unnecessary computing load in searching closed eyes from camera image sequences, the size of searching area is purposely tailored to be as small as possible. Figure 10 shows left eye region of interest (ROI) and right eye ROI which are divided based on relative position of left and right eye positions and general human face geometry. The thick rectangular is the result of face detection in an image with a resolution of 320 by 240 pixels; a frontal face AdaBoost classifier (Viola and Jones, 2001) is applied to the whole image with a minimum searching window of 40 by 40 pixels. This means that the classifier will not search image areas that are smaller than 40 by 40 pixels. Therefore the face image that is smaller than that size will be neglected. We define the height and width of the detected face window as h and w respectively.

In order to make an area that includes both eyes, we cut 20% of height from upper edge of the detected face frame, and another 40% of height from the lower edge of frame. The rectangle area left in the middle is an area that normally contains both eyes. In this area, we further cut 35% of either left or right edge of the area to make two regions of interest that by experience can enclose left eye or right eye. The image area bordered with dashed black lines is an area that contains left eye, and an area enclosed with dashed yellow borders is an experience area that contains right eye. For training closed eye classifiers and extract Haar-like features, these two areas are used as background area for sampling negative and positive training images. During applying the trained classifier for detecting closed eyes, these two regions of interest are used as areas for searching either left or right closed eye images.

Figure 10 Image segmentation of the ROI for left closed eye and right closed eye detection (see online version for colours)

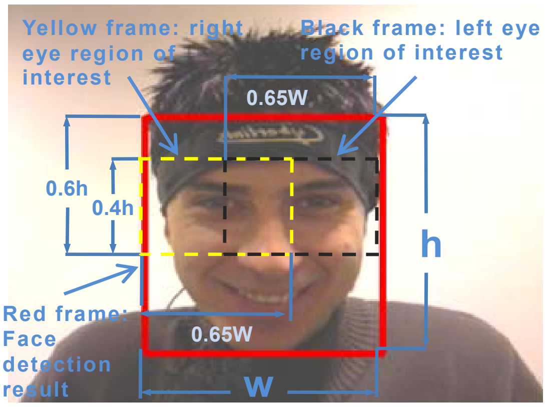
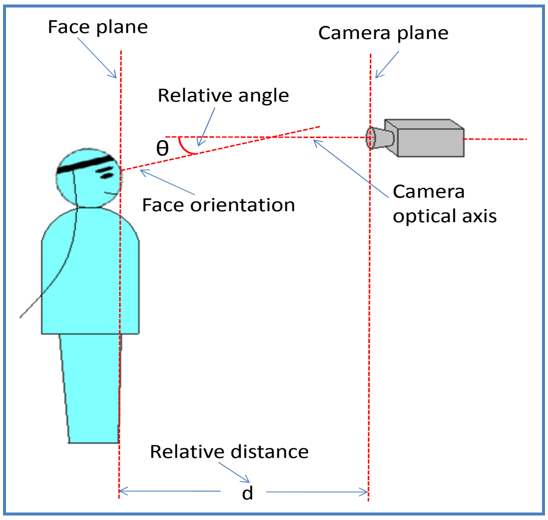


Figure 11 Adaboost training image sampling setup (see online version for colours)



4.2 *Sample training image collection*

For training a classifier with AdaBoost on Haar-like features, a positive image set and a negative image set are needed. The positive image set is a collection of image samples from a target that needs to be detected, and the negative image set can be arbitrary sample images that do not contain a target. In this experiment, two classifiers for detecting either left-closed eye or right-closed eye are trained. Take the left-closed eye training for example, the positive images are collected in left eye ROI when the subjects are closing their left eyes, and the left eye ROI are taken as negative images when subjects keeping their left eye open. We simplify the negative image sampling in application; the classifier is always applied within ROI which is from a fixed area as a result of the human face detector.

There are three factors that affect the detection of closed eyes in applications:

1 relative angle θ, which is the angle between subject face orientation and camera optical axis in Figure 11, and measures the angle of the eye image rotates

2 relative distance, which is the distance from subject face plane to camera plane, and reflects the proportion of zooming of eye images

3 illumination change, which is variations on light sources, luminous intensity, and the lighting angle are designed in image sampling process.

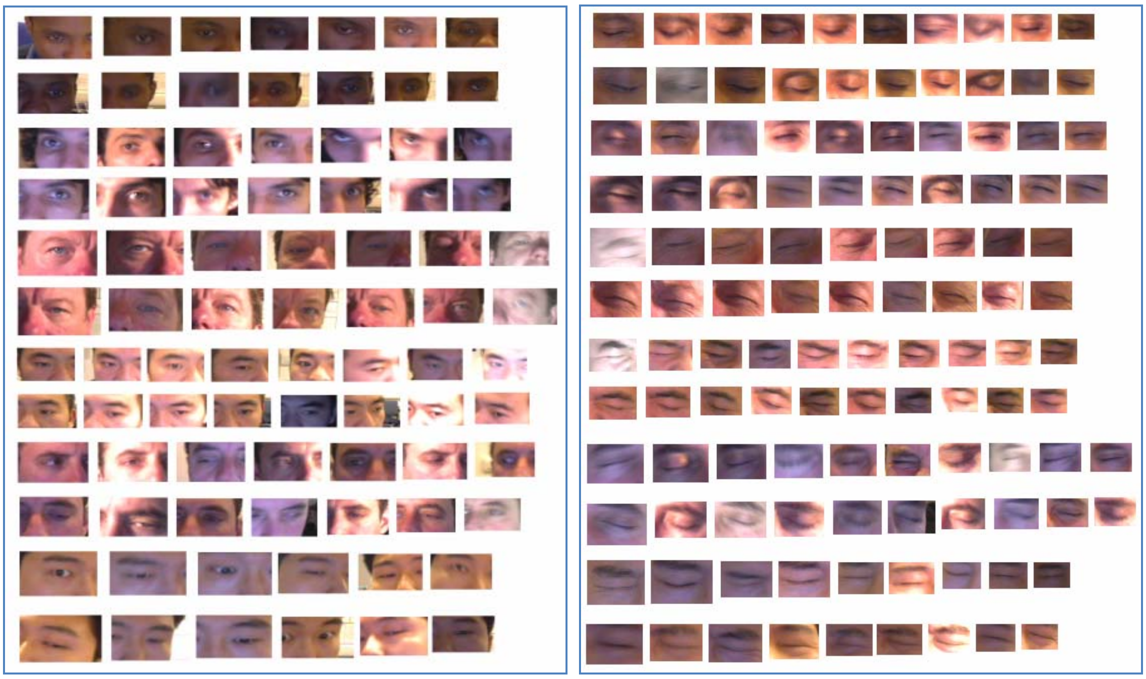
We take these three factors into account during training in the real-world environment for detecting closed eyes when a subject sits on a wheelchair.

In order to make the classifier adaptable, all the changes when a subject is on the wheelchair are synthesised and included in the training, so that these changes can be learned by the AdaBoost learning machine. A camera is fixed to a wheelchair and a subject sitting in the seat has a adjustable distance from 25 to 55 centimetres, while in sampling training images, we make distance changes from 25 to 80 centimetres. During the training, a subject is making open or closed eye movements in front of the sampling camera, at the same time, the subjects are asked to rotate their neck in a clock-wise direction to change the relative angle within 30 degrees when moving forth and back within 25 to 80 centimetres away from the camera. Simultaneously, the light conditions are changed from time to time, by turning on and off three of the six inset ceiling fluorescents each time alternatively, and by turning on and off three incandescent lamps in front of the subjects in turn. These combinations will result in a variety of illumination effects to subjects' face in terms of lighting angle and illumination density.

For the designed task, 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 1,475 positive images with image size range from 36 by 23 pixels to 65 by 44 pixels and 4,286 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 1,318 image samples with a size range from 36 by 23 pixels to 72 by 48 pixels and the negative set have an collection of 3,668 background images with an image size range from 38 by 24 pixels to 124 by 76 pixels. A selection of the sampled negative images of six subjects is shown in Figure 12(a), some of which were sampled in ROI. These negative images may contain some background scenery of the subject since the face detector will be affected by face rotation of subjects. Here the rotation is confined within 30 degrees. The sampled block images in Figure 12(b) shows a part of the collection of positive sampling images from six subjects under different illumination conditions, relative distances and relative angles.

Figure 12 (a) Negative training images captured from six subjects under different illumination conditions, face orientations and distances (b) positive training images captured under different illumination conditions, face orientations and distances (see online version for colours)



(a) (b)

For training AdaBoost learning machines based on Haar-like features extracted from positive and negative images for both eyes, a left-closed eye classifier with 20 cascades was implemented, and the classification accuracy validated among sample training images is 91.75%. For the right-closed eye, the classifier training process was stopped at 19 cascades with a validation accuracy of 92.2%. These two classifiers are therefore applied to their own left or right eye ROI with a minimum searching window of 20 by 20 pixels.

In real-world experiments, the closed eye detector can give a very good performance among the six subjects under different illumination conditions, and can adjust to changes when subjects are moving away or towards the camera in the wheelchair. Also, it can accommodate small rotation with a relative angle within 30 degrees when subjects are turning their heads. The bigger head rotations should be prevented as these rotations can result in the failure of the frontal face detector and the close-eye detector.

5 hybrid recognition-based control strategy

5.1 *Hybrid recognition*

After the patterns of five control movements from both image and EMG are classified, a recognition strategy is applied to combine the classified patterns and recognise the control movements. There are five defined control movements: SJC, DJC, CJC, REC and LEC. SJC and DJC can be recognised by a threshold-based strategy depicted in Section 2.2. Within other three movements, LEC and REC can be distinguished from image information, and their patterns are combined into one EMG class EC. CJC movements can be uniquely recognised from EMG patterns namely CJ. Therefore, the rules for recognising five control movements with their EMG or image patterns can be summarised as follows:

a if (image) LEC pattern is detected in face image and any EMG pattern in class EC is detected then LEC movement is detected, as shown in Figure 13

b if (image) REC pattern is detected in face image and any EMG pattern in class EC is detected then REC movement is detected as shown in Figure 13

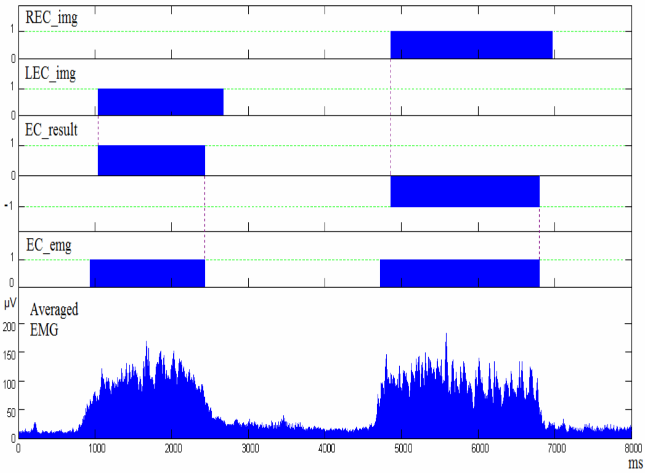
c if any EMG pattern in class CJ is detected then CJC movement is detected as shown in Figure 6

d if the length of time that EMG waveform overflows over fixed a threshold (H) within time ΔT exceeds 30 milliseconds and the average value over EMG amplitude within time ΔT is less than a fixed threshold then SJC movement is detected (threshold-based strategy), as shown in Figure 1

e if two successive SJC movements are detected within a fixed time span ($2\Delta T + \Delta t$) then DJC movement is detected, as shown in Figure 13.

The detailed computing processes for recognising LEC and REC movements are shown in Figure 13. Recognition results of three patterns from LEC and REC movements: two image patterns indicating LEC and REC and one EMG pattern EC are represented as REC_img, LEC_img and EC_emg in the diagram. REC_img, LEC_img and EC_emg are flags whose value is set to '1' when a corresponding pattern is detected and remains '0' when the pattern is not detected. EC_result shows the recognition result of LEC and REC movements, in which EC_result is represented by three conditions: '1', '–1' and '0'. EC_result equals to '1' means LEC is detected, EC_result equals to '-1' means REC is detected, and '0' means no control movements is detected.

Figure 13 Integration of visual pattern recognition and EMG classification (see online version for colours)



The last row of the diagram shows the actual averaged EMG waveform when the subject is making ELC and REC movements. The small grids are the segmentation result in the actual signal processing process, and the EMG signals are processed based on 200 ms segmentations. For each segmentation section, the features of the data section is extracted and classified by the SVM classifier. EC_emg is the SVM classification result of EMG EC class; '1' means the eye close pattern in EMG waveforms is detected. EC_result is a synthesised final result of left and REC pattern recognition based on both EMG classification and image detection results.

5.2 *Control rules*

So far, we have picked up five recognisable facial movements which are SJC, DJC and CJC from jaw movements, as well as LEC and REC from eye movements. To apply these five designated facial movement patterns to the wheelchair control, we designed a rule-based control method comprising six control commands: go forward (GF), turn left (TL), turn right (TR), reduce speed (RS), stop (ST) and go backwards (GB). The logic connections between five facial movements and six executable control commands can be stated as follows:

a if LEC movement is detected then TR

b if REC movement is detected then TL

c if CJC movement is detected then GF

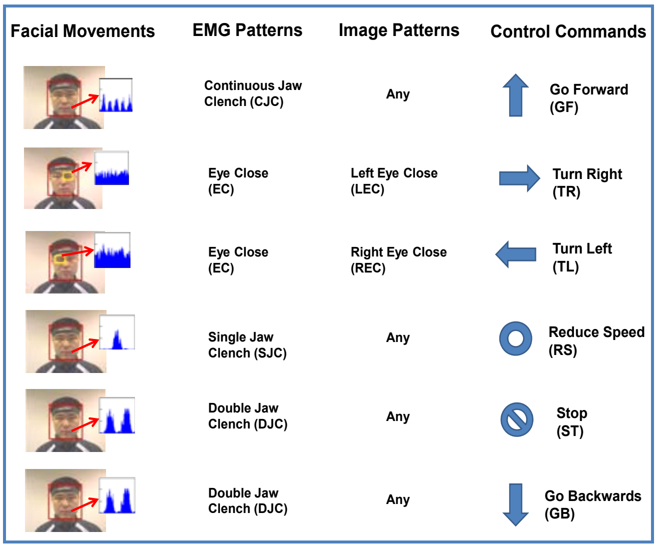
d if SJC movement is detected then RS

e if DJC movement is detected then ST

f If DJC movement is detected and the wheelchair has stopped then GB.

The relations among facial movements, recognised images and EMG patterns and control commands are synthesised in Figure 14. The 1st column presents six facial movements in which the bottom two movements are same, and the picture shows the recognition result in the application. Some detailed muscle movements (eye movements or jaw movements) where relative EMG patterns come from are indicated by arrows, and the EMG waveform is displayed in a section window of 500 ms which are used in Figure 6. The face detection and closed eyes detection results are displayed with red frames and yellow frames respectively.

Figure 14 Logic connection between facial movements, control commands and recognisable patterns (see online version for colours)



The 2nd column lists six EMG patterns used for recognising the six facial movements in the 1st column. The 3re column contains two image patterns used for recognising the facial images in collaborative with the EMG patterns in the 2nd column. The 'any' condition in the 3rd column means the image patterns is irrelevant to the recognition result of control movements, which means the value of the image pattern can be discretional at that time. Finally, the 4th column gives a list of six control commands generated from EMG and image patterns form the 2nd and 3rd columns.

5.3 *Control commands and control strategy*

In this experiment, instead of using joystick, six control commands, namely GF, TL, TR, RS, ST and GB, are applied to replace the function of a joystick for controlling a simulated robot or a real wheelchair. Since both the simulated robot and the electric wheelchair are differentially-driven vehicles, they can be controlled by the velocities of both left and right wheels.

We accelerate the robot or wheelchair by adding a fixed acceleration value to both driving wheels. The simulated robot can response the control commands very fast, but the electric powered wheelchair is unable to response the acceleration demand since it needs time to inertial from the load of the subject and wheelchair frame. A maximum speed of the wheelchair is set so that it will neglect any acceleration commands at this speed. The wheelchair can turn on a spot and follow a curve trajectory. For rotating on the spot, it will rotate at a fixed angular velocity, and for following a curve, it will have an angular velocity for the rotation, which is realised by simply reduce the velocity of one wheel accordingly.

'GF' is a command that will add a fixed acceleration value to the current velocity of both wheels. CJC is a persistent movement and a subject can accelerate the wheelchair speed incrementally by keep doing CJC movements. For every 200 milliseconds, a CJC movement is recognised, thus GF commands will be sent continuously until the wheelchair reaches to a required speed. In this process, the subject does not know how much GF commands are generated but he or she can obtain the feedback from the current speed of the wheelchair. Likewise, LEC and REC are persistent movements and used to control the rotation of the robot or wheelchair. Here the left-eye-closing movement LEC is mapped to the control command TR and right-eye-closing movement REC is used for generating the control command TL.

We generate right eye closing movements to control the wheelchair to TL and left eye closing movements to control the wheelchair to TR. This is because when the wheelchair is turning left; keeping the left eye open will enable the subject to observe the wheelchair turning easily. In case that an obstacle appears, the subject can stop the turning by opening his/her right eye. While for realising rotating on the spot or into a curve trajectory, the general rule is to change the wheel velocities accordingly. Taking TL for example, if the wheelchair is required to TL, the velocity of its left wheel will be reduced to half while the velocity of its right wheel will be kept unchanged. If the wheelchair is required to turn on the spot, both wheels should have same velocity, but in different turning directions.

The recognition time for LEC, REC and CJC movements are 200 milliseconds. But for a SJC movement, the recognition time ΔT depicted in Section 2.1 can be as short as 130 milliseconds. Comparing with LEC, REC and CJC movements, a SJC movement is quicker and easier to produce. Therefore, SJC is generated as the control movement for sending 'RS' command in which the wheelchair speed will reduce to a half each time when this command is received. By using this command a subject can quickly maintain the robot to a safe speed to prepare for turning or avoiding collision. 'ST' and GB are two control commands that both are resulted from DJC movements. More specifically, ST is a command that the wheelchair will reduce its speed to zero at the fastest deceleration, and GB is sent on the occasion that the wheelchair has stopped in which case the wheelchair will go backwards at a safe speed that is the half of the maximum speed.

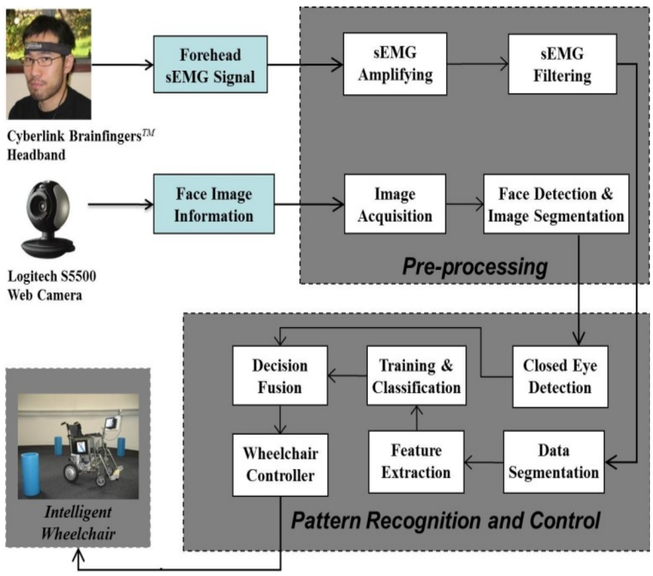
It should be mentioned that a DJC movement is the result of two SJC movements. Before recognising a DJC movement, a SJC movement is recognised earlier. That means that in the actual control, a 'RS' command is sent before every 'ST' and 'GB' commands. For ST commands, a reduce of speed can help the wheelchair to stop smoothly, and for GB commands, since the command can only be active when the robot has stopped, a RS command will keep the velocity of the wheelchair to be zero which means it has stopped.

6 Simulation and real-world experiment results

6.1 System architecture

Figure 15 illustrates the detailed procedure of the system. As can be seen, the system architecture consists of three main parts. The first part is data acquisition devices such as Brainfinger Cyberlink^{TM} and Logitech^{TM} S5500 web camera. The second part is the human computer interface which is responsible for training and classifying facial movement features, mapping selected facial movements into wheelchair control commands. It has two sub blocks, namely data pre-processing block and pattern recognition-control block. The 3rd part is an executable device that can be either a wheelchair or a simulated robot, as shown in Figure 16(a).

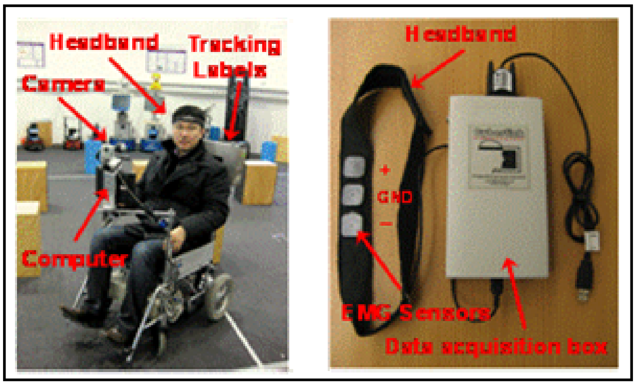
Figure 15 Block diagram of the proposed system architecture (see online version for colours)



6.2 Data acquisition

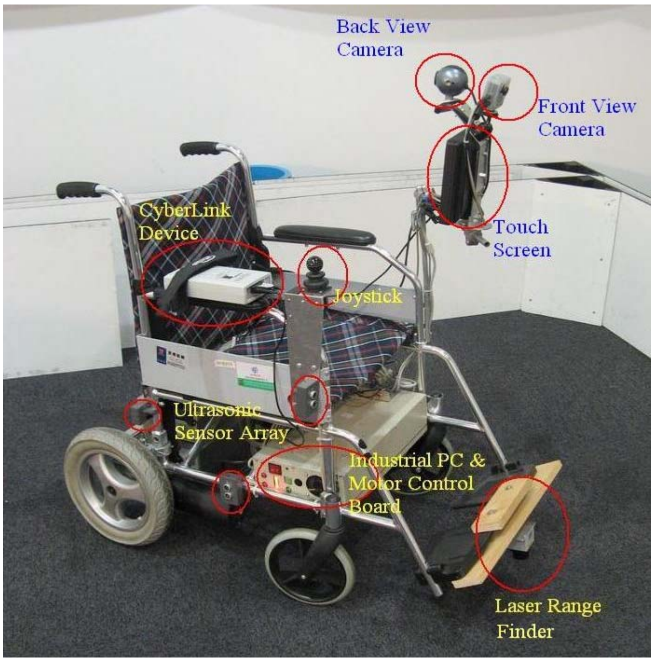
As shown in Figure 16(a), forehead surface EMG signals and facial image information are collected from subjects by attaching wearable skin surface EMG (sEMG) sensors and fixing a camera in front of the wheelchair pointing to user's face. Figure 16(b) gives a close look of the EMG device which consists of a wearable headband and a data acquisition box (DAB). One channel EMG signal is obtained from forehead by the subject wearing a headband attached with three flat EMG electrodes. The DAB contains a series of signal processing algorithms which can turn the amplified raw EMG signals (RES) into averaged EMG waveform with a sampling rate of 100 Hz. The detailed processes of converting raw EMG signals to averaged EMG waveforms are addressed in Section 5.

Figure 16 (a) A subject sitting in wheelchair and wearing cyberlink brainfingers™ headband (b) the cyberlink brainfingers™ data acquisition box and the headband (see online version for colours)



(a) (b)

Figure 17 Intelligent electrical powered wheelchair platform used in the experiment (see online version for colours)



6.3 *Intelligent wheelchair platform*

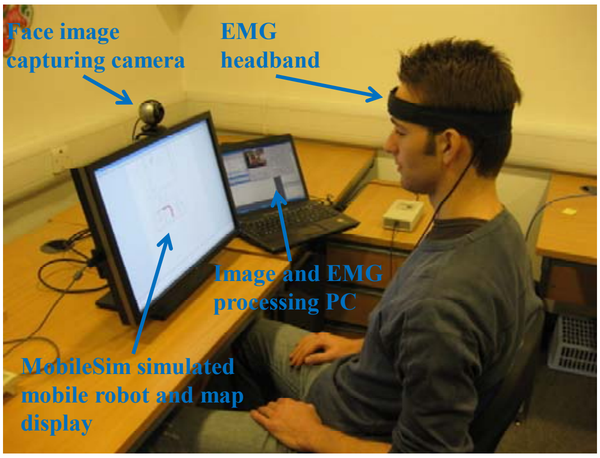
Figure 17 shows the picture of our intelligent wheelchair, namely RoboChair which was built in 2006 and has the following components:

- 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 two 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
- AMD 1.6 G dual core laptop with 1 GB physical memory, integrated graphic card and Windows XP installed to analyse the facial images.

6.4 *Simulated experimental results*

To test the designed control strategy, a simulation was carried out. As shown in Figure 18, a laptop PC (AMD Turion 64 × 2 TL50 CPU and 1GB memory) with an integrated graphic card running on 32 bit Windows XP operation system was deployed. A commercial robot simulator, MobileSim (Shenoy et al., 2008), was running on the PC and displayed on an 18 inches LED screen with a webcam on its top. A subject will wear the cyberlink brainfingers™ headband in the front of the webcam for experiments in which a simulated mobile robot was running in a simulated office on the monitor screen under the control of the proposed hybrid HMI. The camera is used to capture the front face image of the subject during control. At the same time, the EMG headband is put on by the subject to obtain facial EMG signals from subject's forehead.

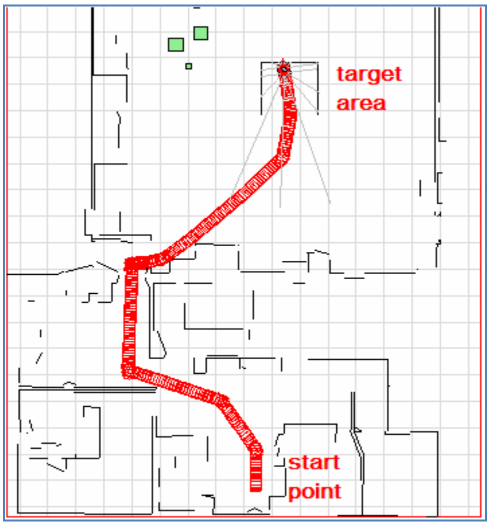
Figure 18 Software simulation environment setup (see online version for colours)



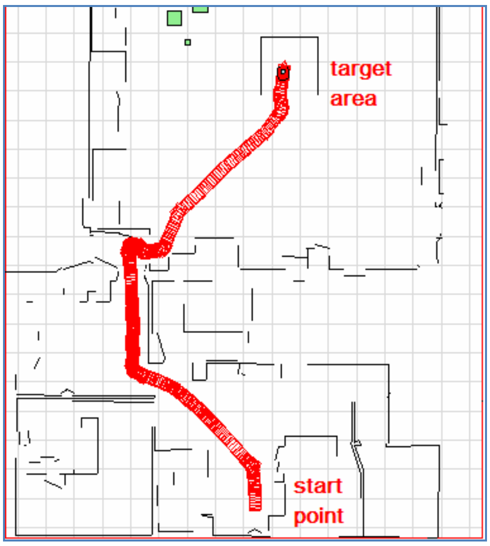
The six male subjects participated in training AdaBoost closed eye classifier in this experiment in order to test the applicability of the proposed control method. In the process of combining the EMG and image control strategies to the subjects, we found that the proposed strategy did not work well for three of the six subjects. In which two subjects felt difficulty in generating one eye open and one eye close movements, and one subject reported difficulty in producing CJC movements by not being able to make fast clenching movements. Although there are still possibilities that these three subjects can learn to generate designed facial movements through further training. However in this experiment, we eliminated these three subjects and use the rest three subjects to operate the simulated robot in the map in an office environment.

Figure 19 shows three trajectories generated by the three subjects controlling the robot travelling from the start point to the target area. The size of the map is 30.5 metres (long) by 15.5 metres (width). The result of the subject is evaluated by the travelling time. In this experiment, subject A is an experienced subject who has knowledge of the map and the control method, while Subjects B and C are new users that have no experience about the control method and the map. Before controlling the actual robot, the inexperienced subjects are taught about the relation between facial movements and the control strategy by controlling the robot in a short training for several minutes. The training ends as long as the trained subjects can remember the facial movements for individual control commands.

Figure 19 Robot trails from a start point to a target area in a simulated indoor environment controlled by three subjects, (a) Subject A (2'30') (b) Subject B (4'50') (c) Subject C (3'12') (see online version for colours)

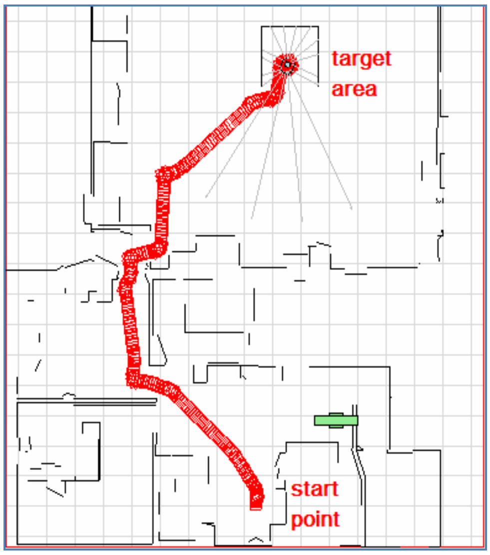


(a)



(b)

Figure 19 Robot trails from a start point to a target area in a simulated indoor environment controlled by three subjects, (a) Subject A (2'30') (b) Subject B (4'50') (c) Subject C (3'12') (continued) (see online version for colours)



(c)

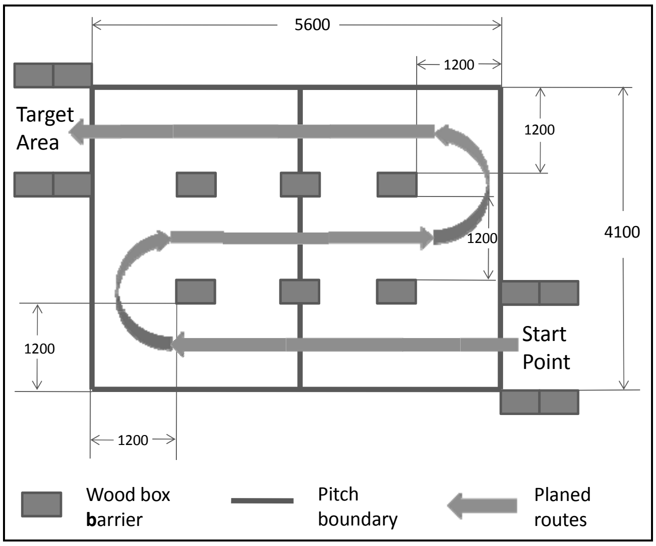
As can be seen from the experimental results shown in Figure 19, all three subjects were asked to control a wheelchair travelling from a state point to a target area. Subject A was the experienced user and took two and half minutes to reach the target. Subject B took four minutes 50 seconds to reach the goal and Subject C took three minutes 12 seconds to finish the travelling. From the trajectory we can see that the inexperienced subjects feel difficulty in turning the robot in the narrow space and the time was spend by manoeuvring out the narrow space. The experienced subject had a clear turning at corners and made less rotation during control and mostly he keeps the robot travelling straight ahead.

6.5 *Real world experimental results*

After the applicability of the proposed control strategy is tested by the simulated experiment, we now choose an experienced subject who can handle the control strategy with the best performance to do real experiments on the real wheelchair. As shown in Figure 20, an experiment map is designed and 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 the map, the geometry size of the wheelchair is 650 millimetres (width) by 1,000 millimetres (long), and the width for all the corridors and passages are 1,200 millimetres. The overall size of the map is 5.6 by 4.1 metres, and the passage is separated evenly by two rows of wood boxes.

Figure 20 Diagram showing the parameters of the experiment map setup and planed routes



The wheelchair system configuration is shown in Figure 16(a). A subject is sitting in the wheelchair with a camera fixed in front for tracking the face, and EMG signals are acquired by a headband sensor device. The configuration of the wheelchair and the onboard computer is described in Section 6.3. In this experiment, we select a best performed subject from the simulation experiment to be the subject in this experiment. This subject can deploy the new control methods to achieve the best performance, which is compared with the joystick control. We hope the difference between control methods can be revealed thorough experiments in terms of time spent on driving the wheelchair from the start point to the target area.

Figure 21 shows the subject controlled the wheelchair with a joystick travelling through the designed map from the start point to the target area for ten times, following a planned route designed in Figure 20. Ten trajectories labelled from ‘Trail 1’ to ‘Trail 10’ are drawn. While in Figure 22, 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 EMG and visual 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 is very 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 ten trajectories 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.

Figure 23 shows the difference of travelling time for two control methods that are analysed statistically in Figure 24. In Figure 24(a), the differences between two control methods in 20 experiments is drawn with the mean and variance in terms of the travelling time from the start point to the target area. 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 much 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. Therefore, the further improvement of the designed control method is required so that it can be used by disabled and elderly people who have difficulty to use their hands. One-way variance analysis (ANOVA) table for analysing sample difference between joystick and hybrid (EMG and vision) control methods is shown in Figure 24(a).

Figure 21 Ten successive trajectories made by a normal subject controlling an intelligent wheelchair with joystick and following a planned route form start point to target area for ten times (see online version for colours)

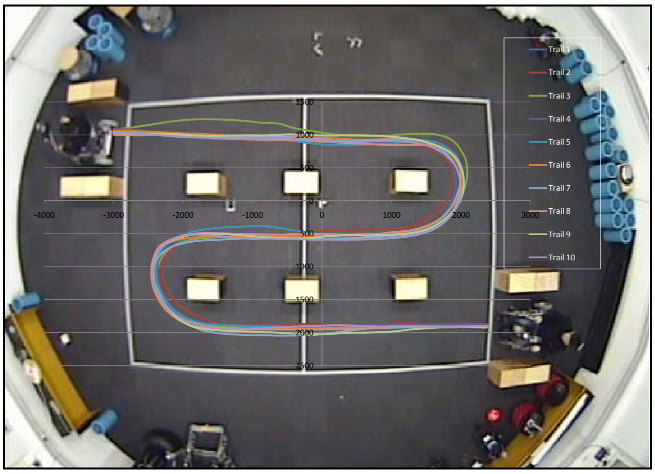


Figure 22 Ten successive trajectories made by a normal subject controlling an intelligent wheelchair with EMG and vision-based control method and following a planned route form start point to target area for ten times (see online version for colours)

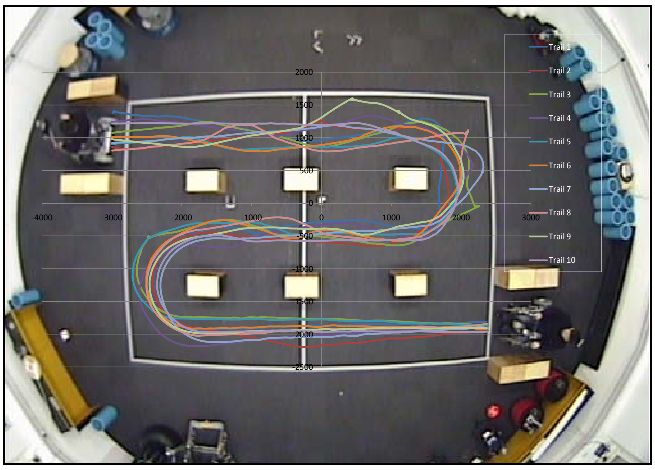


Figure 23 Travel time comparison between joystick control and hybrid (EMG and vision) control (see online version for colours)

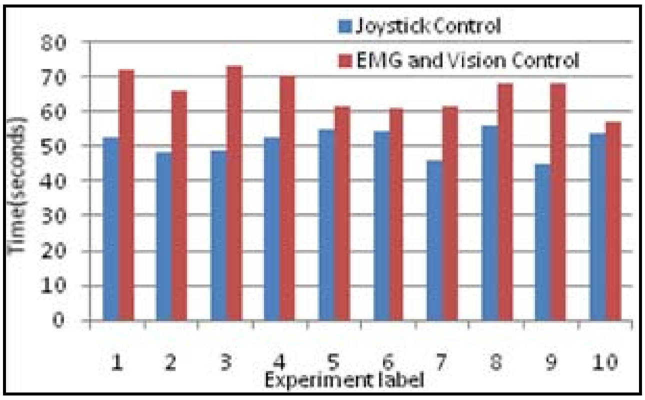
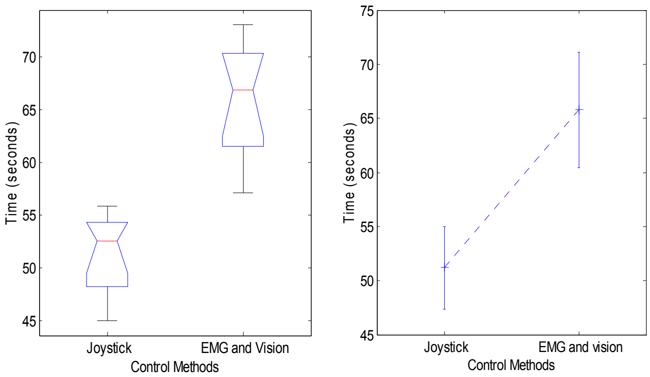


Figure 24 (a) Mean and variance analysis on time difference for joystick and hybrid (EMG and Vision) control methods (b) one-way variance analysis (ANOVA) table for analysing sample difference between joystick and hybrid (EMG and vision) control methods (see online version for colours)



(a) (b)

7 Conclusions and future work

In this paper, a novel hybrid HMI is proposed for hands-free control of an electric powered wheelchair. The developed HMI integrates both facial image information and facial EMG signals to recognise 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. The experiments are carried out by implementing the designed control system firstly onto a simulated robot operated in the MobileSim simulator and secondly on an electric powered wheelchair in an indoor environment. Experimental results show the feasibility and performance of the proposed novel control method, which proves to be a good alternative control interface to traditional joystick control.

In the future, different combinations of muscle groups on the face and facial expressions will be investigated to form various control strategies in order to satisfy more individual needs since different subjects can have different EMG signals or image features, adaptive strategies are required for the real-world application. A separate training of EMG detection and face recognition on each individual subject will be conducted in order to improve the recognition accuracy of human control intensions. Also, real experiments on a large number of subjects will be implemented in the real-world in order to further carry out performance evaluations of this hybrid control method.

References

Barniv, Y., Aguilar, M. and Hasanbelliu, E. (2005) 'Using EMG to anticipate head motion for virtual-environment applications', *IEEE Trans. on Biomedical Engineering*, June, Vol. 52, No. 6, pp.1078–1093.

Chang, C.C. and Lin, C.J. (2001) LIBSVM: A Library for Support Vector Machines, Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Chen, T. (2001) 'Audiovisual speech processing', *IEEE Signal Processing Magazine*, January, Vol. 18, No. 1, pp.9–21.

Chu, J.U., Moon, I., Lee, Y.J., Kim, S.K. and Mun, M.S. (2007) 'A supervised feature-projection-based real-time EMG pattern recognition for multifunction myoelectric hand control', *IEEE/ASME Transactions on Mechatronics*, June, Vol. 12, No. 3, pp.282–290.

Colombo, C., Del Bimbo, A. and De Magistris, S. (1995) 'Human-computer interaction based on eye movement tracking', *International Workshop on Computer Architectures for Machine Perception,* p.258.

Dupont, S. and Luettin, J. (2000) 'Audio-visual speech modeling for continuous speech recognition', *IEEE Transactions on Multimedia*, September, Vol. 2, No. 3, pp.141–151.

Felzer, T. and Freisleben, B. (2002) 'HaWCoS: the 'hands-free' wheelchair control system', *Proceedings of ACM SIGACCESS Conference on Computers and Accessibility*, ACM Press, pp.127–134.

Hazlett, R.L. and Benedek, J. (2007) 'Measuring emotional valence to understand the user's experience of software', *International Journal of Human-Computer Studies*, April, Vol. 65, No. 4, pp.306–314.

Hsu, C.W., Chang, C.C. and Lin, C.J. (2003) 'A practical guide to support vector classification', Technical report, Taipei.

Jia, P., Hu, H., Lu, T. and Yuan, K. (2007) 'Head gesture recognition for hands-free control of an intelligent wheelchair', *Journal of Industrial Robot*, Vol. 34, No. 1, pp.60–68.

Kiguchi, K., Rahman, M.H., Sasaki, M. and Teramoto, K. (2008) 'Development of a 3DOF mobile exoskeleton robot for human upper-limb motion assist', *Journal of Robotics and Autonomous Systems*, August, Vol. 56, No. 8, pp.678–691.

Lienhart, R. and Maydt, J. (2002) 'An extended set of haar-like features for rapid object detection', *IEEE ICIP 2002*, September, Vol. 1, pp.900–903.

Moon, I., Lee, M., Chu, J. and Mun, M. (2005) 'Wearable EMG-based HCI for electric-powered wheelchair users with motor disabilities', *Proc. of IEEE Int. Conf. on Robotics and Automation*, pp. 2649–2654.

Oskoei, M.A. and Hu, H. (2008) 'Support vector machine-based classification scheme for myoelectric control applied to upper limb', *IEEE Transactions on Biomedical Engineering*, August, Vol. 55, No. 8, pp.1956–1965.

Santis, A.D. and Iacoviello, D. (2009) 'Robust real time eye tracking for computer interface for disabled people', *Journal of Computer Methods and Programs in Biomedicine*, October.

SEER Training Modules (2008) *Cancer Registration & Surveillance Modules*, US National Institutes of Health, 20/4/2008, available at http://training.seer.cancer.gov/anatomy/muscular/groups/head_neck.html.

Shenoy, P., K. J. Miller, B. Crawford, R.P.N. Rao, 'Online EMG Control of a Robotic Prosthesis', *Journal of Biomedical Engineering*, Vol. 55, No. 3, pp.1128-1135, March 2008.

Smith, P., Shah, M. and da Vitoria Lobo, N. (2003) 'Determining driver visual attention with one camera', *IEEE Transactions on Intelligent Transportation Systems*, December, Vol. 4, No. 4, pp.205–218.

Viola, P. and Jones, M.J. (2001) 'Rapid object detection using a boosted cascade of simple features', *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.

Wei, L. and Hu, H. (2008) 'Use of forehead bio-signals for controlling an intelligent wheelchair', *Proceedings of IEEE Int. Conf. on Robotics and Biomimetics*, Bangkok, Thailand, 14–17 December, pp.108–113.