Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Vibrotactile Rendering of Head Gestures for Controlling Electric Wheelchair

Shafiq ur Réhman*, Bisser Raytchev†, Ikushi YODA†, Li Liu*
*Digital Media Lab., Tillämpad Fysik och Elektronik, Umeå Universitet, Sweden.
Email: {shafiq.urrehman, li.liu}@tfe.umu.se
†National Institute of Advanced Industrial Science and Technology (AIST), Japan.
Email: {i-yoda, b.raytchev}@aist.go.jp

***Abstract*—We have developed a head gesture controlled electric wheelchair system to aid persons with severe disabilities. Real-time range information obtained from a stereo camera is used to locate and segment the face images of the user from the sensed video. We use an Isomap based nonlinear manifold learning map of facial textures for head pose estimation. Our system is a non-contact vision system, making it much more convenient to use. The user is only required to gesture his/her head to command the wheelchair. To overcome problems with a non responding system, it is necessary to notify the user of the exact system state while the system is in use. In this paper, we explore the use of vibrotactile rendering of head gestures as feedback. Three different feedback systems are developed and tested, audio stimuli, vibrotactile stimuli and audio plus vibrotactile stimuli. We have performed user tests to study the usability of these three display methods. The usability studies show that the method using both audio plus vibrotactile response outperforms the other methods (i.e. audio stimuli, vibrotactile stimuli response).**

***Index Terms*—extended Isomap, Multidimensional Scaling (MDS), head gesture recognition, vibrotactile rendering, wheelchair system, usability.**

I. INTRODUCTION

Assistive devices help or allow persons with disabilities to perform activities that can not be performed or difficult to perform without the assistive devices. An assistive wheelchair system is currently being developed to aid disabled people who are unable to drive a standard powered wheelchair with joystick [1]. The target users are cerebral palsy or stroke patients (i.e. patients lacking fine motor control skills); they don't have ability to access and control the robotic wheelchair using conventional methods, such as joystick or chin stick. For these specific users, being able to move around freely is immensely important to their quality of life. Attempts have therefore been made to make things more convenient with a system that uses non-contact non-constraining image sensing instead [1]. Previously, researchers have studied voice controlled [2], eye controlled [3] and a head gesture controlled wheelchairs [4]. An overview of such systems can be found in [5].

Except for the tests in [1] there has been limited testing of these types of wheelchair systems. Most tests have been performed with somewhat restricted and stable conditions and/or with test subjects whose abilities do not match those of the intended users. Therefore, it cannot be assumed that they are suitable for real users in real environments. For efficiency of any assistive system, it is important to consider how a user will be able to become aware of the response or current state of the system before an action is performed by the assistive system. Regardless of the vision methods used in assistive wheelchairs for navigation and control, these systems can only be useful if they interact effectively with the users. The assistive system must have an adaptable user interface and while navigating in outdoor environment it must consider the cognitive workload of the users [6].

Our head gesture recognition consists of three levels: estimating the face orientation angle and face orientation direction in still images, and recognizing head gestures in moving images. Errors can occur due to incorrect recognition and it is extremely important that the user is informed of the selected action so that he/she can correct the error by repositioning his/her head. To do so the wheelchair needs to be equipped with a real-time feedback system. It is important to develop an intuitive state awareness or response for such system. Normally audio stimuli are used to inform the user of the current state i.e. through voice messages. But in an outdoor environment, it may not provide much help. It can be annoying as well as increase the cognitive workload of the users. To reduce physical and cognitive workload to operate the wheelchair it is important to develop an intuitive feedback method. In this paper, we have tested an assistive wheelchair response system based on three methods to inform the user about the current state of the system; using audio messages only, vibrotactile messages only and audio plus vibrotactile messages. We have investigated the following research questions:

- Which response method improves navigation performance more than the others?
- Does a 'vibrational plus audio stimuli' response method increase the effectiveness of wheelchair?
- Which stimuli interface response technique is efficient for the user to control the wheelchair system and reduce the cognitive workload?

II. HEAD GESTURE CONTROLLED WHEELCHAIR

We have developed a wheelchair system that can be controlled and steered with head gestures. A head gesture is a movement in a time-series arrangement of face orientations. Based on head gestures the operations of our wheelchair are defined as "forward", "right turn", "left turn", "turn on the

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

Fig. 1. Our electric wheelchair fitted with stereo vision camera (QVGA, 15fps), vibrotactile motors (5 coin-motors controlled by PWM), PC (CPU Pentium M 1.6 GHz, RAM: 1GB) and built in sound unit.

spot", "reverse", and "stop". Two other operations can also be performed, i.e., "turn on the spot" operation involves moving the left and right wheels so that they turn the chair around in a circle. The "stop" operation causes the wheelchair to cancel all commands and come to a halt. These operations commands are results of clinical trials in which an electric wheelchair was operated by a magnetic sensor [1]. In following the section, we consider how to characterize the head motion with two parameters, which will be used for estimation of the defined head gestures.

A. Head Gesture Estimation

Even though the ultimate goal is to estimate the face orientation direction with a stereo camera the electric wheelchair (Fig. 1) is also fitted with a previously developed magnetic sensor [1]. The magnetic information acts as a training signal, it is used as a learning procedure to estimate the angles from facial images. Using combinations of magnetic information and facial information extracted from the disparity information, we used a nonlinear manifold to learn the pan (horizontal angle ψ) and tilt (vertical angle θ) angles of the face orientation from the facial images. After the learning phase, the facial orientation angles are estimated using the stereo camera alone (without the magnetic sensor). The extended Isomap model is able to map high-dimensional input data points which are not in the training data set into the dimensionality-reduced space found by the model (also see [7]). To compute face orientation $\Theta = (\theta, \psi)$ from the input face samples it is assumed that there are N faces in the training data for head pose estimation, $\mathbf{X} = \{\mathbf{x}_1 - \bar{\mathbf{x}}, \dots, \mathbf{x}_N - \bar{\mathbf{x}}\}$, where $\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i$. Since the ambient geometry of view-varying face manifolds can be highly folded or curved in the high-dimensional input space, nonlinear manifold learning techniques have to be used here to represent views, i.e., we embed the face space $\mathbf{X}$ into a very compact space $\mathbf{R}$. This is achieved by the Isomap, which is a Multidimensional Scaling (MDS) operating on the pairwise geodesic distance matrix $\mathbf{D}_G$ built from the input data samples $\mathbf{x}_i$. More specifically, a neighborhood graph $\mathbf{G}$ is determined by the edge $\mathbf{x}_i\mathbf{x}_j$ only if $\mathbf{x}_j$ is one of the k nearest neighbors of $\mathbf{x}_i$. The shortest paths in $\mathbf{G}$ are computed for all the pairs of data points, e.g. by using Floyd's algorithm,

$$d_G(\mathbf{x}_i, \mathbf{x}_j) = min\{d_G(\mathbf{x}_i, \mathbf{x}_j), d_G(\mathbf{x}_i, \mathbf{x}_k) + d_G(\mathbf{x}_k, \mathbf{x}_j)\}$$

and the shortest paths between any two samples are presented in matrix $d_G = \{d_G(\mathbf{x}_i, \mathbf{x}_j)\}$. Then classical MDS is applied to $\mathbf{D}_G$ to find a lower-dimensional embedding of the data that best preserves the manifold's intrinsic geometry. This is achieved in two steps:

(1) the inner-product matrix $\mathbf{B} = \mathbf{HAH}^t$ is calculated, where $\mathbf{A} = \{-\frac{1}{2}d_G^2(\mathbf{x}_i, \mathbf{x}_j)\}$ and $\mathbf{H}$ is the centering matrix.

(2) the eigenvectors v_i, corresponding to the top positive l eigenvectors λ_i of $\mathbf{B}$ are found and the required $l - dimensional$ embedding is given by the matrix $\mathbf{R} = (\sqrt{\lambda_1}.\mathbf{v}_1, \dots, \sqrt{\lambda_l}.\mathbf{v}_l)^t$. Now, we can define a pose parameter map $\mathbf{F}$ relating view angles Θ to their corresponding embedded training samples $\mathbf{R}$ in a similar ways as

$$\Theta = \mathbf{FR} \tag{1}$$

where $\mathbf{F}$ can be learnt by

$$\mathbf{F}^t = \mathbf{V}_R \mathbf{W}_R^{-1} \mathbf{U}_R^t \Theta^t \tag{2}$$

where $\mathbf{V}_R \mathbf{W}_R \mathbf{U}_R^t$ is the SVD of $\mathbf{R}^t$. To compute the unknown view angles from test images $\mathbf{X}_T$ the key lies in how to embed the test samples in the subspace created by the model, i.e. to calculate $\mathbf{R}_T$ from $\mathbf{X}_T$. Fortunately, such a technique exists for MDS and can be applied to Isomap in a similar way:

$$\mathbf{R}_T = \frac{1}{2}((\mathbf{R}^t\mathbf{R})^{-1}\mathbf{R}^t)^t \mathbf{D}_T \tag{3}$$

where $\mathbf{D}_T = (d_1, \dots, d_M)$is given as

$$d_i = diag(\mathbf{R}^t\mathbf{R}) - d_{i,N} \tag{4}$$

and

$$d_{i,N} = (d_G^2(\mathbf{x}_{T(i)}, \mathbf{x}_1), \dots, d_G^2(\mathbf{x}_{T(i)}, \mathbf{x}_N))^t \tag{5}$$

Here $d_{i,N}$ is a column vector of the squared geodesic distance between the ith test sample $(i : 1, \dots, M)$ and each of the training samples. Now the unknown view angles Θ_T of the test samples in $\mathbf{X}_T$ can be calculated as

$$\Theta_T = \mathbf{FR}_T \tag{6}$$

Considering head gestures commands, the head orientation directions are classified into four types; *forward*, *right*, *left*, and *down* which can be obtained from the face orientation angle estimation results (see Fig. 2). From nonlinear manifold learning-based procedure, the pan (horizontal angle: ψ) and tilt (vertical angel: θ) angles of the face orientation can be directly obtained from high dimensional data. The face orientation parameters can be used to estimate the defined head gestures. Here we show how to classify the head orientation directions by using face orientation angle estimation values ψ and θ,

$$\delta(t) = \begin{cases} Forward & \text{if } L < \psi < R, B < \theta, \\ Right & \text{if } R \leq \psi, B < \theta, \\ Left & \text{if } \psi \leq L, B < \theta, \\ Down & \text{if } \theta \leq B \end{cases}$$

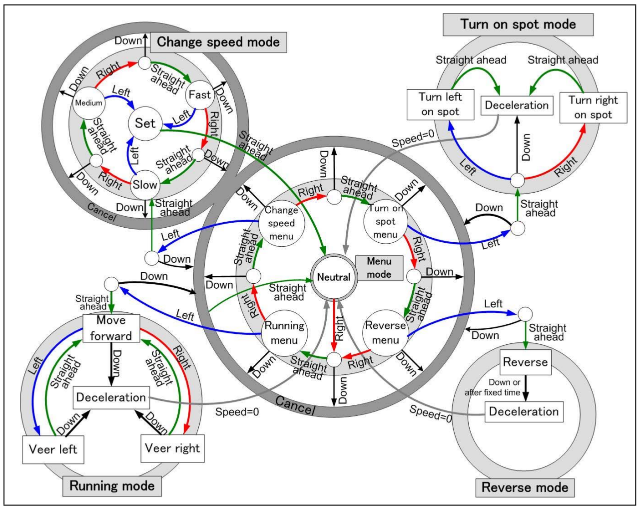

Fig. 2. State transition diagram for the wheelchair functionality. Each part is coded into vibrotactile and audio message to inform the state of the wheel chair system to its users.

where $\delta(t)$ is head/face orientation direction at time t and L, R, and B are the threshold values. In our project we have used view angles in the range $[-90°, \ldots, 90°]$ for pan and $[-45°, \ldots, 45°]$ for tilt, sampled at increments of $15°$.

III. DESIGNING USER CENTERED INTERFACE RESPONSE SYSTEM

In assistive systems the means of interaction provided by the interface must be efficient and effective for the human. To determine the response, we designed an experiment to test the performance of the subjects using (1) audio message response (2) vibrational message response (3) both audio and vibrational stimuli message to control and operate the wheelchair system. We considered a state machine (Fig. 2) in which operations constitute different states; the face orientation direction is represented by a string of symbols, and actions such as movements constitute the output. The machine state diagram comprises of five main components; main mode (in middle), running mode, change speed mode, turn-on-spot mode and reverse mode. To trigger any action by the wheelchair system, the user has to navigate through these modes (starting clockwise from main mode). The system informs the user about its current state by using the designed system response methods. For audio messages normal speakers are used to inform the user current state of the system during control and navigation. The second type of stimuli was more intuitive; we used vibrational message as system response, i.e., using vibrational stimuli on the back of the user. The vibrational stimuli were designed so that it could replace the audio messages and inform the user of system-state information. We used five vibration motors as tactors on the back of the chair to code the current state of the system (see Fig. 1 and Fig. 3). While mounting vibration motors on the wheelchair system all precautions were made so that the tactors must not contact the spinal cord or any other area which can be harmful to the users.

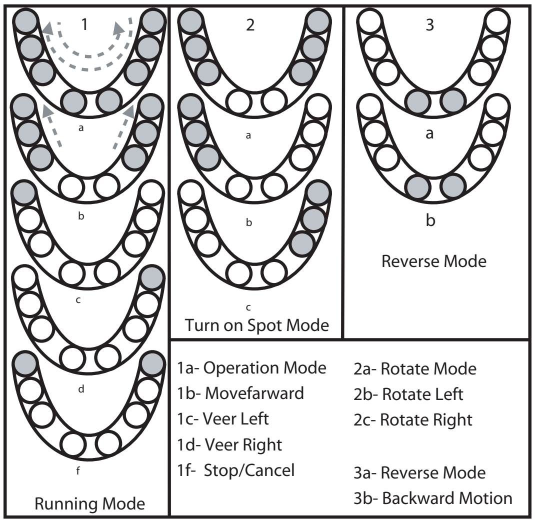

Fig. 3. Tactor arrangement on the back of the wheelchair, which displays vibration stimuli messages corresponding to the state machine components, namely main (1a, 2a, 3a and 1f), running mode (1b, 1c, 1d, 1e), turn-on-spot mode (2b, 2c), reverse mode (3b).

A. *Vibrotactile Rendering of Head Gestures*

To achieve vibrotactile rendering, we developed and attached a stand alone vibrotactile module (see Fig. 1). The vibrotactile module consisted of two components; a Printed Circuit Board (PCB) and vibrator motors. The PCB, containing a micro controller (Atmel ATmega16 L), was programmed to generate Pulse Width Modulation (PWM) at one of the pins and send signals to the vibration motors. The PCB was connected with the onboard computer of a wheelchair. The system response messages were coded into vibrotactile signals and then sent to the PCB. By adjusting the PCB's output to switch it on or off, the vibration motor fitted on the back of the wheelchair could rotate at certain locations with designed frequencies and intensities (see vibration patterns Fig. 3). The tactors were low-cost coin motors used in pagers and cellular phone. The vibration frequencies were selected in a range from $100Hz$ to $1kHz$.

In our system, a vibrotactile signal was controlled by two parameters of vibration; frequency and magnitude. The gesture type was coded into magnitude of vibrotactile stimuli, whereas the intensity was coded into the frequency of the vibrotactile pattern. The magnitude was determined by relative duration ratio of on-off signals. The vibration patterns were carefully designed so that long durations would not make users irritated. Vibration intensities were used to control detection threshold problem, similar to volume control in audio stimuli. The vibrotactile patterns associated with the state transition diagram are depicted in Fig. 3. To reduce possible confusion we presented vibrotactile patterns only for four machine states: *main*, *running mode*, *turn-on-spot mode* and *reverse mode* (Fig 3).

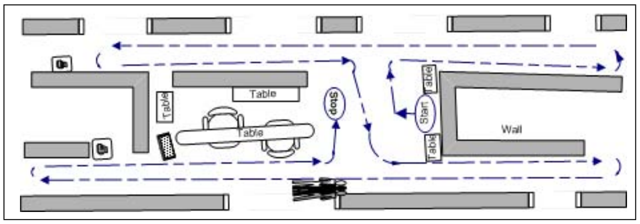

Fig. 4. A diagram of test course used for current experiments. All subjects were asked to navigate using the designed wheelchair response methods.

B. Participant and Procedure

Thirteen able-bodied subjects (10 male and 3 female), ranging in age from 18 to 43, participated in our experiments. All of them had previous experience with audio and vibrational stimuli (i.e. mobile phone vibration). Neither of the subjects had any back problem nor disability. The motivations of experiment were explained to all the subjects. At the beginning of an experiment, the subjects were shown the wheelchair and briefly informed about the purpose of the experiments. Stereo vision sensors and vibrational sensors (used in assisted control) were pointed out and safety measures, such as the power button, were discussed.

After that three response methods were explained to the subjects. The subjects were asked to be seated in the wheelchair and the user interface (i.e. stereovision camera system) was connected to the wheelchair. Once the subject was comfortable with the interface, the session entered a practice phase in which the subject first tried all the response method (i.e. audio, vibration, audio plus vibration). The subject usually spent about one minute per response-method practicing before expressing an understanding of each message method. The test course was designed for user tests; it included obstacles like several office tables, ranks and chairs, turns to the left and to the right, rotation to right and left, and forward/backwards movement. A diagram of the test course used during the experiment is given in Fig. 4.

The test phase consisted of a complete navigation of the test course, using the three defined system response methods. Six subjects started test session with vibration only assisted control and the others started with audio message only based control first; while in the final session, the audio plus vibrational messages were used to traverse the test course by all the subjects. The total test session time for each subject was approximately 30 minutes for all the three experiments using different stimuli.

IV. USABILITY RESULTS

After each test session the subjects were asked to fill in a questionnaire. The researcher recorded the time taken by the subject to traverse the test course. At the completion of the test, the subjects were also asked to rank the designed response methods for each task on a scale from 1 (worst) to 8 (best). The numbers of head-gesture commands given by the subjects to perform the given task were also recorded.

A. User Satisfaction

A measure of user satisfaction can be made if the evaluation team's measure is based on observations of user attitudes towards the system. Thus, it is possible to measure user attitudes using a questionnaire, e.g., "*Is this application interesting? 1=very boring (negative) to 8 =very interesting (positive).*" We used likert style questionnaire (1-8); Fig. 5 shows the user responses for the designed user centered interface. Each label on the x-axis represents 'questions', i.e.

- Rating: How do you rate the user centered response system based on "vibration plus audio" stimuli?
- Usefulness: Effectiveness of the user centered interface.
- Control: Is this response system helpful for system control?
- Trainability: Is the training helpful?
- Comfort: Is this application comfortable to use?
- AInterest: Are you interested in assistive systems?
- Willingness: Are you willing to buy such a product?
- Easiness: Are vibration patterns easy to use?

Experimental results indicated a high easiness for our designed user centered interface, mean score 7.23 (see Fig. 5), i.e. difficulty-easiness of the designed vibration signal interpreting the state transition diagram for the user . Participants gave 'Control' an average score of 6.2 indicating that the controlling wheelchair system was quite effective using the designed response system (i.e. vibration plus audio message). Similarly an average rating of 6.23 was given by the users to consider the application useful.

B. Efficiency

Efficiency regards how much effort that is required in order to accomplish a specific task. An efficient system ought to require as little effort as possible. In the experiments, we used the reaction time of a user and user questionnaire response as an indicator of efficiency. Each participant was asked to provide likert style scores as response to three designed user centered interface response methods. Fig. 6 shows the efficiency in the designed vibration plus audio interface response was improved as compared to the other methods (i.e. 30 to 40%).

C. Effectiveness

Effectiveness of the designed user interface response methods can be based on the error rate of system and cognitive work load of the user. The error rate calculated can be system failure, which was zero as all the participants completed the task while using the three designed interface response methods. To calculate cognitive workload of the user, evaluation criteria from [6] were employed, i.e., the workload was estimated as the number of commands given by the user during the run divided by the run time (i.e. the average command per second). During the experiments the distance covered by each participant (i.e. 6200 mm) and speed of the system were kept fixed. Table I represents the user's performance recorded during the experiments. The average percentage decrease in input commands (i.e. head gestures) per user from "vibration only" to "vibrotactile plus audio" was 15.03%. Similarly using

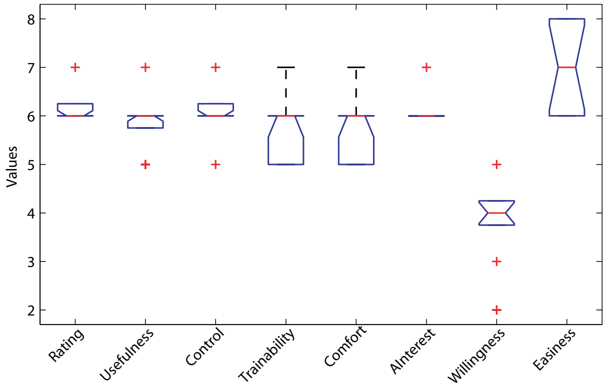

Fig. 5. Mean Questionnaire Score for vibration plus audio response.

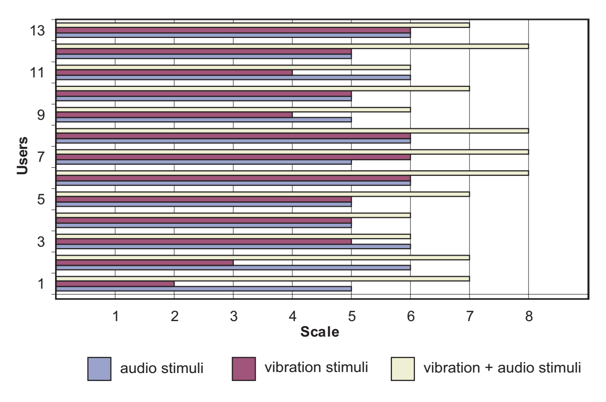

Fig. 6. Efficiency of designed user centered response system.

"vibrotactile plus audio" technique, about 28% less input was required when compared to the effort it takes using "audio only" as interface response method.

V. CONCLUDING REMARKS

TABLE I
EFFECTIVENESS OF THREE USER INTERFACE-RESPONSE METHODS: NUMBER OF COMMANDS (HEAD GESTURES)AND AMOUNT OF TIME (SEC.) SPENT BY THE USER WHILE TRAVERSING THE TEST COURSE USING "AUDIO", "VIBROTACTILE" AND "VIBROTACTILE & AUDIO" INTERFACE RESPONSE METHODS.

User ID	Audio	Vib.	Vib. + Audio
S1	284 (531)	231 (417)	180 (370)
S2	287 (540)	228 (400)	200 (375)
S3	226 (459)	167 (389)	210 (390)
S4	189 (422)	212 (390)	138 (364)
S5	148 (543)	172 (453)	100 (439)
S6	180 (494)	210 (416)	190 (350)
S7	222 (491)	200 (353)	213 (300)
S8	309 (566)	226 (540)	150 (400)
S9	200 (422)	204 (390)	160 (360)
S10	220 (418)	160 (352)	140 (360)
S11	168 (318)	171 (330)	159 (299)
S12	220 (418)	160 (353)	140 (340)
S13	250 (417)	171 (330)	159 (320)

Three types of system response methods are tested and experimental results indicate that the user preferred to use the 'audio plus vibration' message system; and the effectiveness of the interface response system increased by using the designed audio-vibrational messages. When travelling straight (i.e. forward mode) veer left and veer right was very much benefited by vibrational message and the same was true for reverse mode. Whereas to navigate in the main menu (i.e. 1a, 2a, 3a from Fig. 2), most users preferred audio-messages as compared to vibration only stimuli method. It can be due to the fact that native language messages are easier to remember than the designed vibrational messages. Vibrotactile displays provide intuitive methods for human-machine interaction. It is less distractive than audio-visual interaction, and more informative in noisy environments. Tactile and other haptic displays can be more informative regardless of native language and age. We have two subjects with native languages other than the audio messages language (i.e. Japanese); they found tactile response messages more informative than the audio ones. User studies indicate an improvement in the performance of the head gesture based control (i.e. stereo vision module). Increase in the effectiveness of the designed user centered interface response system based on vibrotactile plus audio stimuli is also observed.

REFERENCES

[1] T. Inoue, H. Hirose, and Y. Sudoh, "Development of head gesture interface for electric wheelchair with severe cerebral palsy," *Biomechanism*, vol. 12, pp. 303–313, 2004.

[2] R. Simpson and S. Levine, "Voice control of a powered wheelchair," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 10, pp. 118–122, 2002.

[3] Y. Matsumoto, T. Ino, , and T. Ogasawara, "Development of intelligent wheelchair system with face and gaze based interface," in *IEEE Workshop on Robot and Human Communication*, 2001, pp. 262–267.

[4] 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, 2007.

[5] R. C. Simpson, "Smart wheelchairs: A literature review," *Journal of Rehabilitation Research and Development*, vol. 42, pp. 423–438, 2005.

[6] Holly A. Yanco, *Shared User-Computer Control of a Robotic Wheelchair System*, Ph.D. thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, USA, 2000.

[7] B. Raytchev, I. Yoda, and K. Sakaue, "Head pose estimation by nonlinear manifold learning," in *International Conference on Pattern Recognition*, 2004, vol. 14, pp. 462–466.