Integrated Vision-Based Robotic Arm Interface for Operators with Upper Limb Mobility Impairments

Hairong Jiang, Juan P. Wachs
School of Industrial Engineering
Purdue University
West Lafayette, USA
{jiang115, jpwachs}@purdue.edu

Bradley S. Duerstock
Weldon School of Biomedical Engineering and
School of Industrial Engineering
Purdue University
West Lafayette, USA
bsd@purdue.edu

Abstract— **An integrated, computer vision-based system was developed to operate a commercial wheelchair-mounted robotic manipulator (WMRM). In this paper, a gesture recognition interface system developed specifically for individuals with upper-level spinal cord injuries (SCIs) was combined with object tracking and face recognition systems to be an efficient, hands-free WMRM controller. In this test system, two Kinect cameras were used synergistically to perform a variety of simple object retrieval tasks. One camera was used to interpret the hand gestures to send as commands to control the WMRM and locate the operator's face for object positioning. The other sensor was used to automatically recognize different daily living objects for test subjects to select. The gesture recognition interface incorporated hand detection, tracking and recognition algorithms to obtain a high recognition accuracy of 97.5% for an eight-gesture lexicon. An object recognition module employing Speeded Up Robust Features (SURF) algorithm was performed and recognition results were sent as a command for "coarse positioning" of the robotic arm near the selected daily living object. Automatic face detection was also provided as a shortcut for the subjects to position the objects to the face by using a WMRM. Completion time tasks were conducted to compare manual (gestures only) and semi-manual (gestures, automatic face detection and object recognition) WMRM control modes. The use of automatic face and object detection significantly increased the completion times for retrieving a variety of daily living objects.**

Keywords—spinal cord injuries, gesture recognition, wheelchair-mounted robotic arm, object recognition

I. INTRODUCTION

Previous studies have been conducted to develop wheelchair-mounted robotic manipulators (WMRMs) that provide persons with upper extremity mobility impairments, such as persons with upper-level SCIs, greater autonomy and less reliance on others in retrieving and manipulating objects for activities of daily living (ADL) [1, 2, 3].

The development of WMRMs has been facilitated by the availability of commercial robotic arms emerging in the market. For instance, the Manus manipulator, produced by Exact Dynamics® is a 6 degree of freedom (DoF) robotic manipulator that can be re-programmed and mounted to a wheelchair system [4]. The JACO robotic arm developed by Kinova® is a light-weight robotic manipulator that is designed to be mounted to a motorized wheelchair to help people with upper limb impairments with ADL [5]. However, these commercially-available systems are designed to be controlled by traditional modalities (i.e. joystick), which may not be usable by operators with upper extremity motor impairments.

Prior investigations in human-computer interaction (HCI) for persons with upper extremity motor impairments or quadriplegics has resulted in alternate user input options. The greatest advances have occurred in personal computer (PC) control utilizing speech recognition, facial expression, eye tracking, and hand gesture recognition [6, 7]. However, these HCI modalities, which do not rely upon switch or joystick operation, have also been useful for controlling actuated assistive technology (AT) devices, such as driving intelligent wheelchairs. Alternate input modalities that do not require switch, button or joystick operation for directly or semi-autonomously controlling intelligent wheelchairs include speech recognition [8], gesture recognition [7], tongue movement [9], or electromyography (EMG) and electrooculography (EOG) [10].

These control modalities also have benefits for controlling robotic arms for WMRM systems, though the positioning of the robotic gripper in three-dimensional Cartesian space and prehensile manipulating of objects provide unique challenges. However, existing HCI modalities [3] as well as emerging brain computer interfaces (BCI) [11] and state-of-the-art computer vision systems have been shown to be capable controllers for WMRM systems [12, 13]. This latter work has shown that a camera mounted in the hand of the robotic manipulator provides an effective visual interface for WMRM control [2, 3, 14]. The vision-based system for the UCF-MANUS using a touchscreen interface was equivalent to other input modalities but significantly better than trackball operation [3].

We developed an upper limb gesture recognition system to control a WMRM utilizing the JACO robotic arm. Hand and arm gestures are an intuitive communication form and provide an effective HCI modality. Gesture recognition does not require sensors or other contacts to the operator's body compared to many other HCI, such as EMG, EOG, tongue drives. Likewise, the user does not need to make contact with

NIH Director's Pathfinder Award, grant no. 1DP4GM096842-01

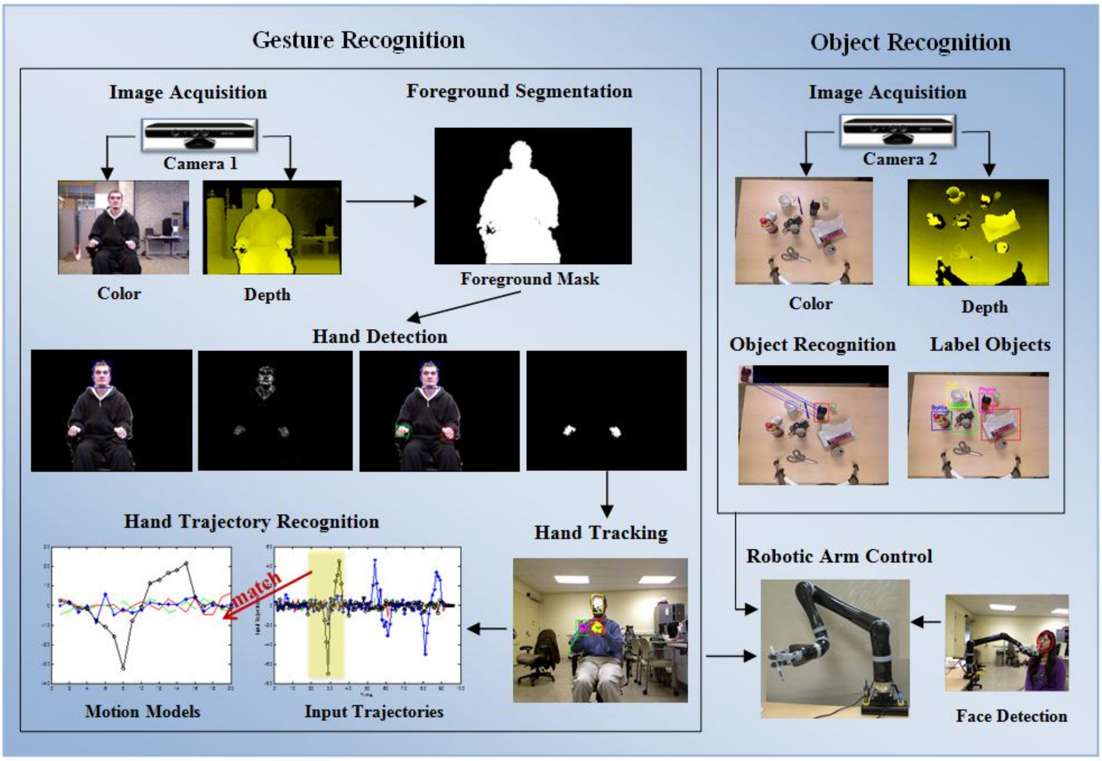


Fig. 1. System Architecture

buttons, joysticks, touchscreens, or sip and puff straws allowing free arm movement during AT device control [15]. Moreover, the lexicon of hand gestures for a gesture recognition-based interface can be customized to meet the requirements of the users for certain tasks. The works in [16] have shown that gestures are a simple and intuitive modality for robotic manipulator control.

In our previous studies [17, 18], a gesture recognition-based interface was designed and developed to allow individuals with upper-level SCIs to send commands for robotic control. In this paper, we combine this gesture recognition-based interface with face and object recognition modules for subjects to more efficiently retrieve daily living objects [19] in the environment. This study allows for further investigation of this vision-based WMRM controller.

II. System Architecture

The architecture of the proposed system is illustrated in Figure 1. Two Kinect® [20] video cameras were employed and served as inputs for the gesture recognition and object detection modules respectively. The results of these two modules were then passed as commands to the execution modules to control the JACO robotic arm (Kinova, Inc., Montréal, Canada). Briefly, these modules are described as follows:

A. *Gesture Recognition Module*

The video input from Kinect camera was processed in four stages using for gesture recognition based WMRM system control; foreground segmentation, hand detection, tracking, and hand trajectory recognition stage. Foreground segmentation was used to increase computational efficiency by reducing search range for hand detection and later stage process. The face and hands were detected from the foreground which provided an initialization region for hand tracking stage. The tracked trajectories were then segmented and compared to the pre-constructed motion models and classified them as certain gesture groups. The recognized gesture was then encoded and passed as command to control the WMRM.

B. *Object Recognition Module*

The goal of the object recognition module is to detect the different daily living objects and assign a unique identifier for each of these objects. A template was created for each object being recognized. These templates were compared to each frame in the video sequence to obtain the best matching object. The results were then encoded and passed as commands to position the robotic manipulator.

C. *Automatic Face Detection Module*

A face detector [21] was employed in this module to perform automatic face detection. The goal was to provide a shortcut for the subjects to position the objects to the front of the face by controlling the robotic arm.

D. *Execution Module*

The robotic arm was programmed as a wrapper using JACO API under C# environment which was then called by the main program. The JACO robotic arm was mounted to the seat frame of a motorized wheelchair. The robotic arm was controlled by the encoded commands from gesture recognition, automatic face detection and object recognition module

III. METHODOLOGY

A. Gesture Recognition-Based Interface

In this section, a brief introduction is provided for the gesture recognition-based interface (Figure 1. left column). A detailed description can be referred to [17, 18].

- Foreground Segmentation

Two steps were adopted in this stage to segment the human body as the foreground. In the first step, the depth information was acquired by a Kinect sensor with depth value *D(i, j)* for each pixel, where, *i* and *j* denote the horizontal and vertical coordinates of the pixel. Each frame was then thresholded by the depth value of each pixel. Two thresholds (T_{DH} and T_{DL}) were set to remove the pixels outside this range [18]. Only those pixels with a depth value between T_{DH} and T_{DL} were kept in a binary mask image. In the second step, the biggest region was extracted as the foreground and all the remaining blobs with a smaller area were discarded.

- Hand Detection and Tracking

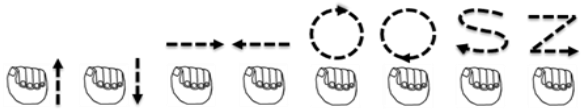
Skin color detection was conducted by employing two 3D histogram models. A face detector [21] was used to remove the face region and extract the remaining two largest blobs as the hand regions. The face and hands detection results were only used to provide an initialization region for hand tracking. A three dimensional particle filter framework was employed to track the hands through all the video sequence by incorporating both color and depth information. In addition, an interaction model using motion and spatial information was integrated to the particle filter framework to solve "false merge" (when the tracker loses the object being tracked and mistakenly focuses on a different object that has higher observation likelihood) and "false labeling" (when exchange of labels assigned to objects after interaction or occlusion occurs). These problems usually occur when hands cross or overlap each other [17], [18].

- Trajectory Recognition

An eight-gesture lexicon (Figure 2) was adopted for the gesture recognition based interface [17]. The acquired hand positions from the tracking stage were then formed as trajectories and compared with the motion models of each gesture in the lexicon. The motion models were created by using the training data collected from eight able-bodied and two subjects with quadriplegia by aligning using by dynamic time warping algorithm [22]. The CONDENSATION algorithm [23] was then used to recognize the input gesture trajectories. The state S at time t was extended to be used for two hand gestures as:

$$S_t = (\mu, \phi^i, \alpha^i, \rho^i) = (\mu, \phi^{right}, \phi^{left}, \alpha^{right}, \alpha^{left}, \rho^{right}, \rho^{left})$$

where, μ is the index of the motion models, ϕ is the current phase in the model, α is an amplitude scaling factor, ρ is a time dimension scaling factor, i equals to right hand, or left hand. Each classified gesture was then passed as commands to control the WMRM. As mentioned in [14], this gesture recognition based interface can provide a recognition accuracy of 95.8%.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 2. Gesture lexicon. (a) upward; (b) downward; (c) rightward; (d) leftward; (e) clockwise circle; (f) counter-clockwise circle; (g) figure S; (h) figure Z.

B. Object Recognition

An object recognition module was developed concurrently with the gesture recognition-based interface to provide more efficient operation for quadriplegic users in retrieving objects (Figure 1, right column). Each frame of the video sequences was captured by a Kinect camera. The distance of each pixel within an object from the depth sensor was mapped to intensity levels. Thus, the father the object is from the sensor, the higher the intensity is. An example of the color and depth frames is shown in Figure 3. In this figure, different daily living objects that a wheelchair user would be expected to often retrieve and bring to one's face were tested, including a box of tissues, cordless telephone, water bottle, coffee mug, and electric shaver. In addition, these objects vary significantly in shape, size, and weight for more exhaustive testing of object recognition and robotic arm manipulation.

(a) (b)

Fig. 3. (a) Color frame of test objects (b) Depth frame of test objects.

A Speeded Up Robust Features (SURF) algorithm was employed to recognize these daily living test objects [24]. A template with SURF features for each object was created before the object recognition process. Each frame captured by the Kinect camera was passed as input to the object recognition system. The SURF algorithm was then applied to each frame to acquire the features. These obtained features were then compared to the template features to get the best matching point pairs which were used to localize the objects (Figure 4). The label for each object was given to the matching object.

After localizing the objects, the robotic manipulator could be automatically directed to the position of the object. However, in this study we did not tackle the problem of how to grab objects for subjects to randomly choose. In terms of these constraints, the robotic arm needed to be fixed in a position where the object was not touched. The highest point of the object was extracted by computing the smallest value within the detected object region in the depth frame. This object recognition and localization process is called "coarse localization". In this paper, "fine localization" for object grasping and manipulating were accomplished by hand gesture recognition-based control.

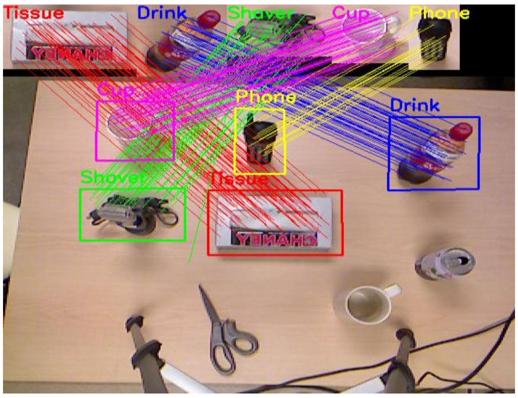


Fig. 4. Automatic Recognition of Daily Living Objects.

C. Robotic Manipulator Control Policies

The JACO robotic manipulator was mounted on the left side of the wheelchair (Figure 5(a)) to provide users with disabilities more capabilities to interact and manipulate the objects in the environment (Figure 5(b)). The JACO robotic arm was manufactured specifically to be mounted on wheelchairs to assist users in performing manipulation tasks. A C# wrapper was implemented using the resident JACO API to control the robotic manipulator. The JACO robotic manipulator has 6 degrees of freedom that were separated into three control modes: 3-D translation of the hand, wrist rotation, and finger grasping. During operation each mode had to be selected. Under translation and wrist control mode, three axes were controlled. Under finger control mode, two or three finger grasping could be selected. The eight-gesture lexicon in Figure 2 was used to control the system. A mapping between the gestures and the robotic control modes are shown in Table 1.

(a) (b)

Fig. 5. (a) JACO robotic arm (b) Object manipulation.

TABLE I. GESTURE CONTROLS FOR THE ROBOTIC ARM

Gesture	**JACO Arm Control Mode**		
	Translation (Directional hand motion)	***Wrist***	***Finger***
Upward	Up	Wrist rotation clockwise	--
Downward	Down	Wrist rotation counter-clockwise	--
Rightward	Right	Lateral orientation (index side)	Open three fingers
Leftward	Left	Lateral orientation (thumb side)	Close three fingers
Clockwise Circle	Forward	Vertical orientation (top side)	Open two fingers
Counter-clockwise Circle	Backward	Vertical orientation (botton side)	Close two fingers
S	Change mode (translation to wrist)	Change mode (wrist to translation)	Change mode (finger to wrist)
Z	Change mode (translation to finger)	Change mode (wrist to finger)	Change mode (finger to translation)

D. Integration of Vision-Based Systems

The robotic manipulator could be controlled through the integration of the gesture recognition-based interface, object recognition module, and automatic face detection. The gesture recognition controller was used to operate the robotic manipulator for fine localization. The object recognition and face detection modules were used for coarse localization of the robotic arm to the selected object and the user's face to provide more efficient robotic arm control. This flow chart of the proposed system is described in Figure 6.

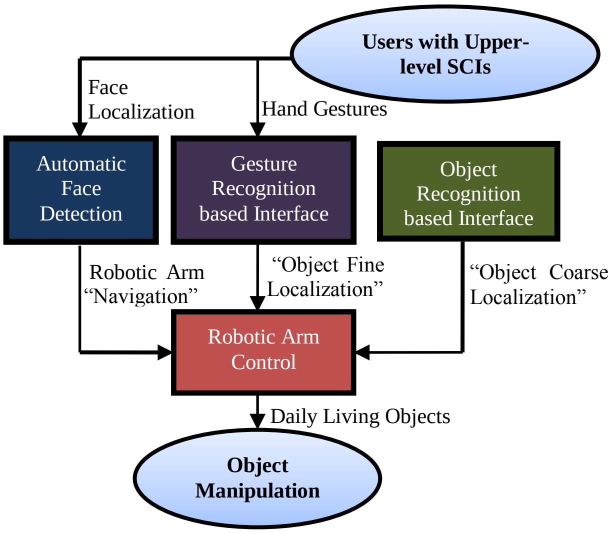


Fig. 6. Integrated computer vision system flow chart.

IV. EXPERIMENTS AND RESULTS

Preliminary experimentation was conducted with three able-bodied subjects to demonstrate the validity of the system. The Institutional Review Board (IRB) approval has been obtained to conduct this study. Although no subjects with upper-level spinal cord injuries were recruited in this experiment section, the gesture lexicon was constructed with three subjects with upper extremity mobility impairments and

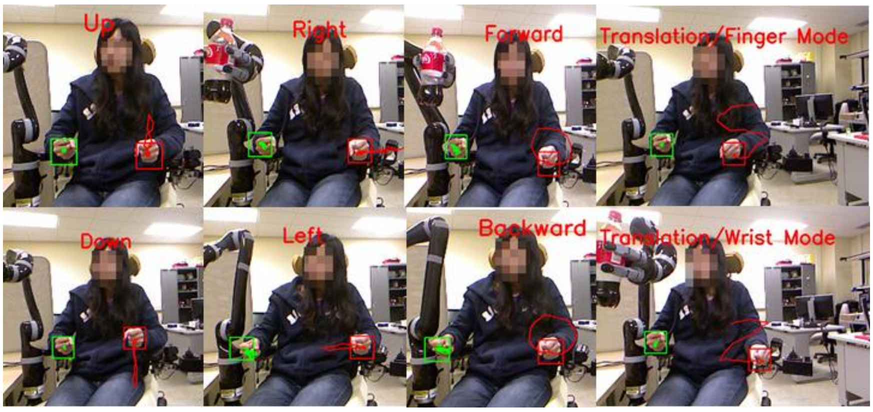


Fig. 7. Gesture recognition based robotic control sample results.

the gesture recognition based control system has already been evaluated with two subjects with quadriplegia [17, 18]. In this paper, we plan to test a more efficient means to operate a WMRM using the proposed optimized vision-based system. Five daily living objects (box of tissues, coffee mug, electric shaver, cordless phone and 16 ounce drink bottle) were selected as test targets to be manipulated by the vision-based system. These test objects were selected based on their variety of shapes, sizes, and weights. Two sets of performance experiments were compared. One is the "Manual" control experiment, which was to have subjects only use gestures to position the robotic manipulator to a test object, pick up the object, and position it in front of the face of the subject. The other experiment is "Semi-automatic" control, which was to perform object recognition to position the robotic arm to the top of the test objects and then use hand gestures to perform "fine positioning" and picking up the object. Then automatic face detection was used to position the object in front of the subject's face. The gesture lexicon in Figure 2 was used in this section for robotic arm control (Figure 7).

The average task completion time (mean with variance) for object grasping was compared (Figure 8). As expected, there was a significant difference between the average task completion time of "Semi-automatic" (176.9s) and "Manual" (287.4s) control.

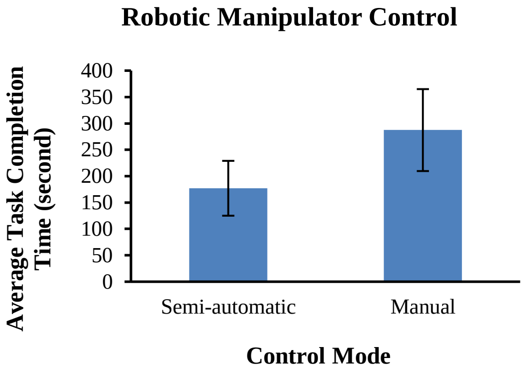


Fig. 8. Comparison between semi-automatic and manual robotic manipulator control modes, single factor ANOVA, $p<0.05$.

Average task completion time (mean with variance) for particular objects were also performed (Figure 9). There was no significant difference in task completion times among different objects. However, for the objects as cordless phone, since it needed to be grasped without touching the keyboard, the subjects may need more time to figure out a proper orientation to move towards it under "manually control". While under "Semi-automatic" control, the robotic arm was already located above the object, so the subjects only need to rotate the robotic arm with a few operations and then move the robotic arm down to grasp the cordless phone which cost them less time.

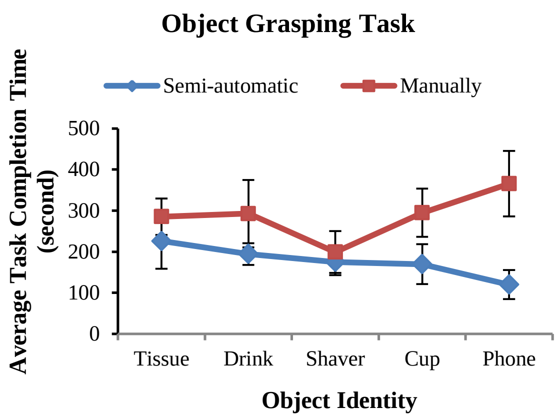


Fig. 9. Comparison among different object grasping tasks, single factor ANOVA, $p<0.05$.

Conclusions and Future Work

This paper demonstrates the feasibility and greater efficiency of a WMRM control system that implements an integrated gesture recognition-based interface incorporating both face and object recognition capabilities.

An eight-gesture lexicon was employed and mapped to the robotic control functions. The gesture recognition-based interface provides individuals with high level SCIs a noninvasive method to control a WMRM and interact with objects around them.

The object recognition module simplified the process of robotic arm navigation and reduced the task completion time for object grasping. The face detection module provided the subjects a shortcut to move the robotic arm towards the face instead of navigating it. It was shown that this "semi-automatic" mode saves users time and labor in performing common retrieval tasks, which would be the bulk of activity for a WMRM.

However, although it may further reduce the time for individuals with SCIs to control the robotic arm, it would also reduce the freedom for the subjects to interact with the environment. An optimal solution may be the "semi-automatic" control mode, which saves time and effort for the users and at the same time provide them with more flexibility in robotic arm control for object manipulation.

Future work will include; (1) recruiting more subjects, particularly those with upper-level SCIs, (2) integrating the whole system in a more efficient and practical design for practical use for wheelchair users, and (3) improving object recognition algorithm to allow the robotic arm to grasp objects according to its functionality.

ACKNOWLEDGMENT

This work was performed at the Institute for Accessible Science through the NIH Director's Pathfinder Award to Promote Diversity in the Scientific Workforce, funded by the American Recovery and Reinvestment Act and administered by the National Institute of General Medical Sciences (grant no. 1DP4GM096842-01). We are grateful for the assistance of Jamie Nolan from the Institute for Accessible Science and Mithun Jacob from the Intelligent System and Assistive Technology (ISAT) lab at Purdue University.

REFERENCES

[1] J. Amat, "Intelligent Wheelchairs and Assistant Robots," Autonomous Robotic Systems, Lecture Notes in Control and Information Sciences, vol. 236, 1998, pp. 211-221.

[2] C-S. Chung and R.A. Cooper, "Literature Review of Wheelchair-Mounted Robotic Manipulation: User Interface and End-User Evaluation," *In:* RENA 2012 Annual Conference, Baltimore, MD, USA.

[3] D. J. Kim, R. Hazlett-Knudsen, H. Culver-Godfrey, G. Rucks, T. Cunningham, D. Portee, and A. Behal, "How Autonomy Impacts Performance and Satisfaction: Results From a Study With Spinal Cord Injured Subjects Using an Assistive Robot," *J:* IEEE Transactions on Systems Man and Cybernetics Part A-Systems and Humans, vol. 42(1), 2012, pp. 2-14.

[4] H. Eftring, K. Boschian, "Technical results from manus user trials," *In:* Proceedings of the ICORR, 1999, pp. 136-141.

[5] V. Maheu, J. Frappier, P.S. Archambault, and F. Routhier, "Evaluation of the JACO robotic arm: Clinico-economic study for powered wheelchair users with upper-extremity disabilities," *In:* IEEE International Conference on Rehabilitation Robotics (ICORR), 2011, pp. 1-5.

[6] J. A. Jacko, "Human-Computer Interaction Design and Development Approaches," *In:* 14th HCI International Conference, 2011, pp. 169-180.

[7] M. Reale, P. Liu, and L. J. Yin, "Using eye gaze, head pose and facial expression for person-alized non-player character interaction," *In:* IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, IEEE Press, 2011, pp. 13-18.

[8] M. Nishimori, T. Saitoh, and R. Konishi, "Voice controlled intelligent wheelchair". *In:* SICE Annual Conference, 2007, pp. 336-340.

[9] R. Vaidyanathan, B. Chung, L. Gupta, H. Kook, S. Kota, and J. D. West, "Tongue-movement communication and control concept for hands-free human-machine interfaces," *J:* IEEE Transactions on Systems Man and Cybernetics Part a-Systems and Humans, 2007, vol. 37(4), pp. 533-546.

[10] I. Moon, M. Lee, J. Ryu, and M. Mun, "Intelligent Robotic Wheelchair with EMG-, Ges-ture-, and Voice-based Interfaces," *In:* International Conference on Intelligent Robots and Systems, IEEE Press, 2003, pp. 3453-3458.

[11] M. Palankar, K. J. De Laurentis, R. Alqasemi, E. Veras, R. Dubey, Y. Arbel, and E. Donchin, "Control of a 9-Dof Wheelchair-mounted robotic arm system using a P300 Brain Computer Interface: Initial experiments", *In:* IEEE International Conference on Robotics and Biomimetics, 2008, pp. 348-353.

[12] K. Tsui, and H. Yanco. "Simplifying wheelchair mounted robotic arm control with a visual interface," *In:* AAAI Spring Symposium on Multidisciplinary Collaboration for Socially Assistive Robotics, 2007.

[13] D. J. Kim, R. Lovelett, and A. Behal, "An Empirical Study with Simulated ADL Tasks using a Vision-Guided Assistive Robot Arm," 11th IEEE International Conference on Rehabilitation Robotics, 2009.

[14] K. M. Tsui, D. J. Kim, A. Behal, D. Kontak, and H. A. Yanco, "I want that: Human-in-the-loop control of a wheelchair-mounted robotic arm," Applied Bionics and Biomechanics, vol. 8(1), 2011, pp. 127-147.

[15] M. Hashimoto, K. Takahashi, and M. Shimada, "Wheelchair control using an EOG- and EMG-based gesture interface," *In:* IEEE International Conference on Advanced Intelligent Mechatronics, 2009, pp. 1212-1217.

[16] J. Wachs, H. Stern, and Y. Edan, "Cluster Labeling and Parameter Estimation for the Automated Setup of a Hand-Gesture Recognition System," *J:* IEEE Transactions on Systems, Man, and Cybernetics: Part A, vol. 35, 2005, pp.932-944.

[17] H. Jiang, B. S. Duerstock, and J. P. Wachs, "Integrated gesture recognition based interface for people with upper extremity mobility impairments," Advances in Human Factors and Ergonomics, vol. 8, 2012, pp. 546-554.

[18] H. Jiang, J. P. Wachs, and B. S. Duerstock, "Facilitated Gesture Recognition Based Interfaces for People with Upper Extremity Physical Impairments," *In Proc.* Pattern Recognition, Image Analysis, Computer Vision, and Applications. Lecture Notes in Computer Science, 2012, pp. 228-235.

[19] H. Pirsiavash, and D. Ramanan, "Detecting activities of daily living in first-person camera views," *In:* IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2012, pp: 2847-2854.

[20] P. M. Yanik, J. Manganelli, J. Merino, A. L Threatt, J. O. Brooks, K. E. Green, and I. D. Walker, "Use of kinect depth data and Growing Neural Gas for gesture based robot control," *In:* International Conference on Pervasive Computing Technologies for Healthcare, 2012, pp. 283-290.

[21] P. Viola, and M. Jones, "Rapid object detection using a boosted cascade of simple features," *In*: International Conference on Computer Vision and Pattern Recognition, 2001, pp. 511-518.

[22] J. Aach, and G. M. Church, "Alignment gene expression time series with time warping algorithms", *J:* Bioinformatics, Oxford University Press, 2001, vol. 17, pp. 495-508.

[23] M. J. Black, and A. D. Jepson, "A probabilistic framework for matching temporal trajectories: CONDENSATION-based recognition of gesture and expressions," *In:* European Conference on Computer Vision, 1998, pp. 909-924.

[24] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "SURF: Speeded Up Robust Features," *J:* Computer Vision and Image Understanding, vol. 110, 2008, pp. 346-359.