CHAPTER XX

Integrated Gesture Recognition Based Interface For People With Upper Extremity Mobility Impairments

Hairong Jiang, Bradley S. Duerstock, Juan P. Wachs

Purdue University
West Lafayette, USA
jiang115@purdue.edu
bsd@purdue.edu
jpwachs@purdue.edu

ABSTRACT

Gestures are of particular interest as a HCI modality for navigation because people already use gestures habitually to indicate directions. It only takes a user to learn few customized gestures for a given navigational task, as opposed to other technologies that require changing hardware components and lengthy procedures. We propose an integrated gesture recognition based interface for people with upper extremity mobility impairments to control a service robot. The following procedure was followed to construct the suggested system. Firstly, quadriplegics ranked a set of gestures using a Borg scale. This led to a number of principles for developing a gesture lexicon. Secondly, a particle filter method was used to recognize hands and represent a generalized model for hand motion based on its temporal trajectories. Finally, a CONDENSATION method was employed to classify the hand trajectories into different classes (commands) used, in turn, to control an actuated device-a robot. A validation experiment to control a service robot to negotiate obstacles in a controlled environment was conducted and results were reported.

Keywords: Borg scale, gesture recognition, particle filter, CONDENSATION

1 INTRODUCTION

Carrying out an independent and autonomous life is deemed a basic need for people with mobility impairments (Cooper, Rninger, and Spaeth, 2006). According to the 2009 US Census Bureau News, 3.3 million people who are 15 or older use a wheelchair. Another 10.2 million use an ambulatory aid such as a cane, crutches or walker and 11 million disabled people need personal assistance with everyday activities. For adults 65 and older over 40% are reported to have some form of disability (U.S. Census Bureau News, 2009). Thus, the demand for more innovative assistive technology (AT) development is needed by a large elderly population and those with movement impairments.

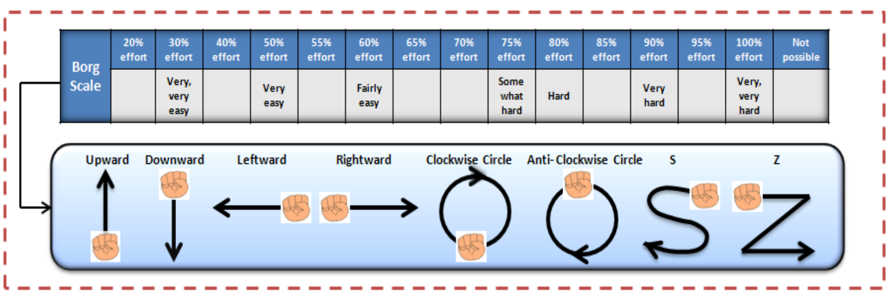
The rapid development of mobile distributed computing systems with effective-human-computer interfaces (HCI) is gaining more popularity (Jacko, 2011). Advanced HCI systems for people with mobility impairments, such as voice, facial and hand gesture based control have been developed where each modality acted alone for the control, or they were combined as multimodal interfaces (Moon, Lee, and Ryu, 2003). Such HCI systems have been used to control or operate wellness monitoring, caregiver assistance, in-home medical alert systems for elder care and intelligent wheelchairs (Scherer, Sax, and Vanbiervliet, 2005; Nguyen, Chahir, and Molina, 2010; Reale, Liu, and Yin, 2011). The use of hand gestures for navigation is an attractive alternative to otherwise cumbersome interfaces, such as joysticks, sip-and puff systems, and tongue controls (Huo and Ghovanloo, 2009). Gesture-based HCI have become popular because they are ergonomic and can be designed to meet individual's particular requirements. Gesture comes naturally to people and is a basic form for individuals to communicate with each other. While not every individual can use gestures, for those who are able to move their hands and upper arms to some degree, gesture-based HCI can be seen as an extremely promising alternative or complement to existing interface techniques.

2 RELATED WORK

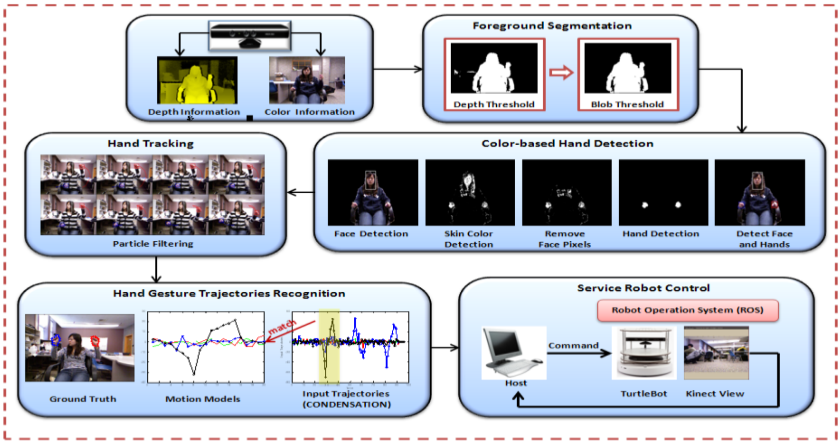
There exists a number of works incorporating hand gestures through interfaces to control wheelchair, mobile robot and some other home devices. A hand-gesture based wheelchair navigation system was designed which tracked the hand in real time by using the combination of particle filtering and mean shift method (Shan, Wei, and Tan, 2004). However, it only took into account one hand tracking and only four gestures to control the wheelchair. An intelligent wheelchair control system based on hand gesture recognition was proposed, in which static hand gestures were recognized by employing Haar-like feature detector (Zhang, Zhang, and Luo, 2011). The commands used to control the wheelchair were generated by locating the static gesture in different regions in a window for each frame in the video sequence. The main problem with static hand gestures based control is that the users have to hold their hands in a fixed position for a period of time, which is exhausting when upper extremity impairments exist. A robotic smart house interface was designed to assist people with movement disabilities based on hand gesture, voice and body movements (Park, Bien, and Lee, 2007). The proposed system used ceiling-mounted CCD cameras to observe and recognize the user's hand gestures and used these gestures to control robotic systems and other home-installed devices. But in their system, only pointing gestures were considered. A hand-gesture based control interface for navigating a car-robot was introduced in (Wu, Su, and Wang, 2010). They adopted the dynamic time warping algorithm to classify hand trajectories. Only six commands were considered for car-robot navigation. A natural interaction framework for programming a mobile robot with gestures was developed using two low-cost small mobile robots available on the market (Uribe, Alves, and Rosario, 2011). A comparison between a joystick, wiimote and kinect based interfaces was made in the paper. However, the preliminary gestures selected for controlling the mobile robot required the user to hold their hands over their body for a period of time which made them feel tired during the experiments.

3 SYSTEM ARCHITECTURE

The architecture of this system is illustrated in Figure 1. First, a gesture lexicon was constructed. Subjects with upper extremity impairments were interviewed and a Borg scale (Borg, 1982) rankings were collected to set the guidelines for gestures selection. Detailed description and analysis is shown in sections 4 and 7. After data analysis, eight dynamic gestures were selected to be the components of the gesture lexicon. The hand gesture recognition system includes four parts: foreground segmentation, detection, tracking, and trajectories recognition. A detailed description of the system is shown in sections 5 and 7.



(a) Gesture lexicon



(b) Gesture recognition

Figure 1 System overview

4 GESTURE LEXICON

When developing a gesture-based HCI for persons with upper extremity mobility impairments, the first step is to design a "gesture lexicon". The gesture lexicon is a set of gestures that was determined from quadriplegic subjects using Borg scale metrics. The procedure for designing this lexicon involved first selecting a subset of gestures. This could be accomplished through technology-based approach or human-based approach (Cassell, 1998). The technology-based approach aims to select a set of gestures that can be easily recognized and classified by the system. The biggest problem for the technology-based approach is that gestures selected by this method are not user-centered, which may be difficult to relate to functions or difficult to perform, particularly for persons with upper limb mobility impairments. The human-based approach proposes to construct the gesture lexicon based on studying how people (who are the potential users) interact with each other. In our research, the target population is users with mobility impairments. In such a scenario, it is mandatory that the gestures will be ergonomic (easy to perform). To achieve this goal the human-based approach will be used for constructing the gesture lexicon.

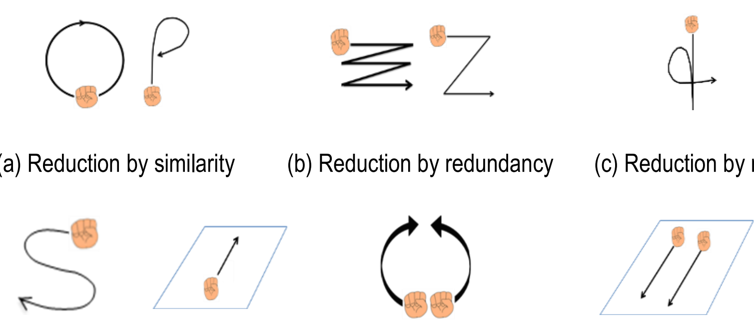
First, a preliminary lexicon including both dynamic and static gestures were shown and demonstrated through a video presentation to subjects with upper extremity impairments and they ranked these gestures employing the Borg scale. The dynamic gestures are hand and upper arm movements by which hand trajectories are generated. The static gestures are postures involving fixed hand and upper limb positions, but not incorporating sophisticated finger variation. Originally, the preliminary gesture lexicon is constructed with 50 dynamic gestures and 15 static gestures. Since the number of gestures that a user can remember is very limited, we limited the number of gestures. The reduced gesture lexicon includes 30 dynamic gestures and 10 static gestures. The gestures were reduced according to the following principles:

(1) Similarity (Proctor and Zandt, 2008): if the trajectory of one gesture has very few differences from another gesture or has many common elements with another gesture, one of the gestures can be deleted. One example is the clockwise circle gesture and the P gesture. Both of their trajectories first go up and then go down as shown by Figure 2(a). The clockwise circle gesture was selected because it was more symmetric.

(2) Redundancy (Yee, 2009): if the trajectory of one gesture contains part of the trajectory of another, this gesture is a redundant gesture of the other. For example: as shown by Figure 2(b) the start gesture (the left one) trajectory contains a Z gesture trajectory. Arbitrary the Z gesture was selected.

(3) Minimize memory load (Nielsen, 1992): reduce gestures that are hard to remember, i.e. the sum gesture as shown by Figure 2(c).

After reduction, the preliminary gesture lexicon includes single hand dynamic gestures in both vertical plane and horizontal plane, two hand gestures in both vertical plane and horizontal plane (as shown in Figure 2(d)) and static gestures.



(a) Reduction by similarity (b) Reduction by redundancy (c) Reduction by memory

(d) Preliminary gesture lexicon components

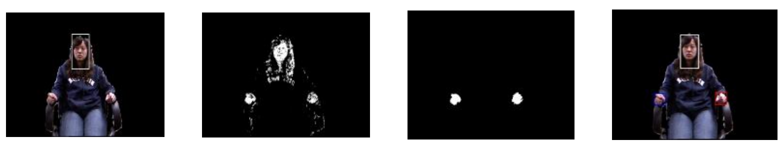
Figure 2 Gesture lexicon

5 GESTURE RECOGNITION

The gestures in the lexicon were recognized by the system and the commands corresponding to each gesture were sent to service robot. Four procedures were followed to achieve the gesture recognition. Firstly, the whole human body will be treated as the foreground and segmented from the background (as shown by Figure 3). Secondly, the hands are detected by using face detection and skin color histogram model. The results are shown as in Figure 4. A particle filter method was subsequently used to recognize the hands and represent human hand motion as temporal trajectories (as shown in Figure 5). Finally, dynamic velocity motion models for the gestures in the lexicon were constructed and CONDENSATION-based trajectory recognition method was used to classify the hand trajectories into different classes.

(a)Depth Image (b) Depth threshold mask (c) Foreground mask

Figure 3 Foreground segmentation



(a)Face Detection (b) Skin color detection (c) Hand extraction (d) Localization

Figure 4 Face and hand detection

Figure 5 Face and hand tracking

6 SERVICE ROBOT CONTROL

The service robot controlled by the gestures recognition system is the TurtleBot™ robot from Willow garage® (as shown in Figure 6), which includes a mobile base, a Kinect 3D sensor and a netbook with robot operation system (ROS) on a linux environment. TurtleBot was controlled through gesture commands. Two modes were used to control TurtleBot: discrete mode and continuous mode. In discrete mode, the robot moves every time that a command is issued, otherwise it stays still. While in the continuous mode, the robot responds to a given command, until the stop command is issued. To switch between these two modes one distinctive gesture is used.

Figure 6 Turtlebot robot

7 EXPERIMENTS

7.1 Gesture lexicon construction

One female with hemi-media in her left arm and two male subjects with upper extremity impairments were interviewed for data collection. One of the male subjects was a Cervical-4/5 quadriplegic, whose left arm had less movement than his right arm, which performed most of the gestures but no hand movement. During evaluation, both dynamic and static gestures were ranked by each subject from 20% effort to not possible according to the Borg scale. The 15-point Borg scale was chosen because it is more sensitive to the variation of effort subjects spent on each gesture. The subject was asked to perform each gesture, to describe the limitations and physical stress they experience, and then to rank the effort needed to perform the gesture on the Borg scale. Once the subjects finished all the 16 single hand dynamic gestures, 14 two hand dynamic gestures and 10 static gestures, their answers were summarized in a data sheet to help us develop the guidelines for hand movements and postures.

According to the scored rankings on the Borg scale, eight dynamic gestures were selected as the optimal candidates for the gesture lexicon. They are upward, Downward, Rightward, Leftward, Z, Clockwise Circle, Counter Clock Circle, and S gesture in vertical plane (as shown in Figure 1(a)). The effort required to perform the selected dynamic gestures, discarded dynamic gestures, static gestures and two hand dynamic gestures are shown by the histogram in Figure 7. From the histogram, it can be seen that the selected dynamic gestures required the least effort, on average.

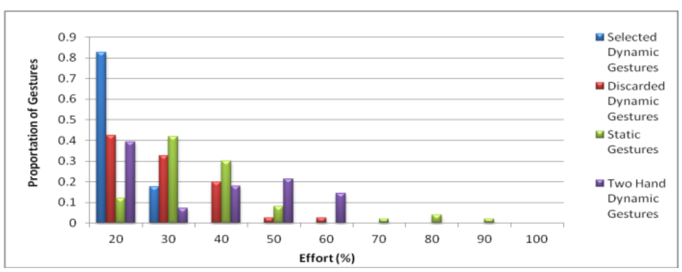


Figure 7 Comparison between selected and discarded gesture groups

ANOVA (Analysis of Variance) and T-test results were given to prove that there was significance difference between the different groups in terms of the effort required to perform the gestures. The mean for the selected dynamic gesture population is μ_1, for the discarded dynamic gesture population is μ_2, for the discarded static gesture population is μ_3 and for the two hands dynamic gesture population is μ_4. The hypotheses were that H_0 states that there is no significant difference among the effort of the selected dynamic, discarded dynamic, static and two hand dynamic gesture population, while H_1 states that there is significant difference among them. The significance level α is set to 0.05. The hypotheses tested are:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 \quad H_1: \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$$

P value found was 8.73E-09, which was less than the significance level 0.05. The null hypothesis was rejected, which means there was significant difference among the effort of the selected dynamic, discarded dynamic, static and two hand dynamic gesture populations. Also, it was found that there was significant difference in the effort required between the selected dynamic gestures and the other three gesture groups by applying t-test to each of the two populations. By choosing the null hypothesis as H_{10} and the alternative hypothesis as H_{11}:

$$H_{10}: \mu_1 = \mu_i \quad H_{11}: \mu_1 \neq \mu_i$$

where j = 2, 3, 4. The p value found were 8.19E-05, 2.61E-09 and 3.15E-05, which were less than the significance level 0.05. These indicated that there were significances of effort between the selected dynamic gestures and the other three gesture groups.

7.2 Heuristics

According to (Kortum, 2008), and the scored rankings on the Borg scale, the following heuristics were found to guide the lexicon design process.

(1) Select gestures that do not strain the muscles.
(2) Select gestures that do not require much outward elbow joint extension.
(3) Select gestures that do not require much outward shoulder joint extension.
(4) Select gestures that avoid outer positions.
(5) Select dynamic gestures instead of static gestures.
(6) Select vertical plane gestures where hands' extension is avoided.
(7) Relaxed neutral position is in the middle between outer positions.
(8) Select gestures that do not require wrist joint extension caused by hand rotation.

7.3 Gesture recognition and robot control

The eight-gesture lexicon for the system was tested by nine users, which resulted in a recognition accuracy of about 90.36% on average for all the gestures. The Turtlebot robot was controlled to deliver instruments from place A to B. Figure 8 shows a sequence of the Kinect view from the Turtlebot. The map for the lab and the trajectories of the robot for both discrete mode control (red solid line) and continuous mode control (blue dash line) are shown in Figure 9.

Figure 8 Turtlebot robot kinect view

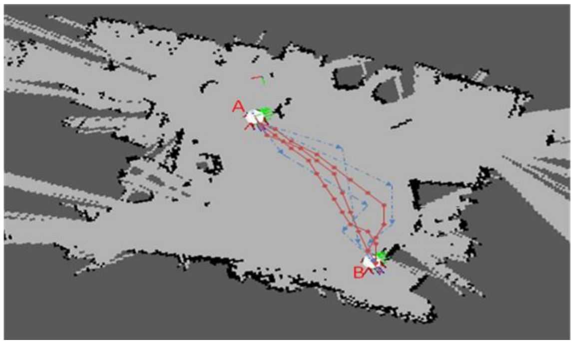


Figure 9 Map of the lab and the robot trajectories

8 CONCLUSIONS

In this paper, a hand gesture lexicon was designed for people with upper extremity mobility impairments. It was shown to require low physical effort based on subjective rankings made on a Borg scale. The heuristics for selecting appropriate gesture lexicons was outlined considering physical and ergonomic constraints. It was found that it is better to select dynamic vertical plane gestures instead of static gestures. In addition, the gestures within the lexicon must not require much outward wrist, elbow or shoulder joint extension. A gesture-based recognition system utilizing this eight-gesture lexicon was tested by nine users and resulted in a recognition accuracy of about 90.36%. The lexicon was validated in a task involving a service robot, which was controlled in a lab environment to deliver laboratory instruments. Two of the gestures in the lexicon were not always recognized due to the fact that the motion models used for those gestures were similar to other gestures in the lexicon. Future work includes developing more robust recognition algorithms based on Bayesian belief networks.

ACKNOWLEDGMENTS

This work was partially funded by the National Institutes of Health through the NIH Director's Pathfinder Award to Promote Diversity in the Scientific Workforce, grant number DP4-GM096842-01.

REFERENCES

Baudel, T. and Beaudoin-lafon, M.. 1993. Charade: remote control of objects using free-hand gestures. *Communications of the ACM* 36: 28-35.

Borg, G. 1982. Psychophysical bases of perceived exertion. *Medicine & Science in Sports & Exercise* 14(5): 377-383.

Cassell, J.. 1998. A Framework For Gesture Generation and Interpretation. *Computer Vision in Human-Machine Interaction, New York: Cambridge University Press* 191-215.

Cooper, R.A., Rninger, M.L. and Spaeth, D.M., et al. 2006. Engineering better wheelchairs to enhance community participation. *IEEE Transaction n Neural Systems and Rehabilitation Engineering* 14(4): 438-455.

Huo, X. and Ghovanloo, M.. 2009. Using Unconstrained Tongue Motion as an Alternative Control Mechanism for Wheeled Mobility. *IEEE Transaction on Biomedical Engineering* 56(6): 1719-1726.

Jacko, J.A.. 2011. Human-Computer Interaction Design and Development Approaches. *14th HCI International Conference* 169-180.

Kortum, P.. 2008. HCI Beyond the GUI: Design for Haptic, Speech, Olfactory, and Other Nontraditional Inferences. *P. Kortum, ed., Morgan Kaufmann* 75-106.

Moon, I., Lee, M. and Ryu, J., et al. 2003. Intelligent Robotic Wheelchair with EMG-, Gesture-, and Voice-based Interfaces. *Proceedings of the 2003 IEEE International Conference on Intelligent Robots and Systems* 3453-3458.

Nguyen, B.L., Chahir, Y. and Molina, M., et al. 2010. Eye Gaze Tracking with Free Head Movement using a single camera. *Proceedings of the Symposium on Information and Communication Technology.*

Nielsen, J.. 1992. The usability engineering life cycle. *IEEE Computer* 25(3): 12-22.

Park, K.H., Bien, Z. and Lee, J.J., et al. 2007. Robotic smart house to assist people with movement disabilities. *Auton Robot* 22:183-198.

Proctor, R.W. and Zandt, T.V.. 2008. *Human Factors in Simple and Complex Systems*.

Reale, M., Liu, P. and Yin. L.J.. 2011. Using eye gaze, head pose and facial expression for personalized non-player character interaction. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* 13-18.

Scherer, M.J., Sax, C. and Vanbiervliet, A., et al. 2005. Predictors of assistive technology use: The importance of personal and psychosocial factors. *Disability and Rehabilitation* 27(21): 1321-1331.

Shan, C.F., Wei, Y.C. and Tan, T.N., et al. 2004. Real time hand tracking by combining particle filtering and mean shift. *Sixth IEEE International Conference on Automatic Face and Gesture Recognition* 669- 674.

Suma, E.A., Lange, B. and Rizzo, A., et al. 2011. FAAST: The Flexible Action and Articulated Skeleton Toolkit. *IEEE Virtual Reality*. 247-248.

Uribe, A., Alves, S., Rosario, J.M., et al. 2011. Mobile robotic teleoperation using gesture-based human interfaces. *IEEE Colombian Conference on Automatic Control and Industrial Applications (LARC).* 1-6.

U.S. Census Bureau, [Online]. Available: http://www.census.gov/hhes/www/disability/over-view.html.

Wu, XH., Su, MC., Wang, PC.. 2010, A Hand-Gesture-Based Control Interface for a Car-Robot. *IEEE International Conference on Intelligent Robots and Systems.*

Yee, W.. 2009. Potential Limitations of Multi-touch Gesture Vocabulary: Differentiation, Adoption, Fatigue. *Human-Computer Interaction, Part II, HCI* 291-300.

Zhang, Y., Zhang, J. and Luo, Y.. 2011. A Novel Intelligent Wheelchair Control System Based On Hand Gesture Recognition. *Proceedings of the IEEE International Conference on Complex Medical Engineering*

Zimmerman, T., Lanier, J. and Blanchard, C., et al. 1987. A Hand Gesture Interface Device. *In Proceedings of CHI 87 and GI, ACM* 189-192.