ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 11, March 2017

A Gesture based TV Control Interface for Visually Impaired: Initial Design and User Study

Author

R.Naveena, V.Vinitha (U.G.Scholars)

Department of Electronics and Communication Engineering,
V.R.S. College of Engineering & Technology,
Arasur,Villupuram-605602.

***Abstract*—We introduce our initial design of gesture interface for the TV control of visually impaired users, with it simple implementation and user study. Two bare-hand gesture sets were designed using simple linear and circular motions of arms. For the two gesture sets, a rule-based recognition system was developed using a Kinect sensor. In a user study, the linear and circular gesture sets on our recognition system scored similarly and showed equal or better subjective ratings compared to the commercial pointer-type gesture interface. The users were negative at pointer-type gestures and showed their concerns about large motions of linear gestures and difficult matching of circular gestures.**

Keywords—gesture recognition; depth camera; accessibility

I. INTRODUCTION

Currently, there are 285 million of blind or partially sighted persons in the world [1]. To guarantee their human rights, it is important to make media accessible for the visually impaired people. TV is a traditional and still the most popular media source over the all ages and regions. However, most of the TVsets can be controlled by a remote controller with some buttons and visually impaired people are suffering when they find the remote controller and press a button on it. This traditional interface have been an obstacle to the media access of visually impaired.

Recently, several alternative input methods such as voicerecognition and remote control using a smartphone have beenattempted with a use of voice guidance of operation and menu.

These methods are intuitive and provide more information to the visually impaired users than the traditional button based input method. However, they still need a certain type of remotecontroller (microphone or smartphone) held on the users' hand,which is hard to find and bother some to visually impaired users.

Gesture interface is getting the light with the development of motion sensors and its intuitiveness as a natural user interface [2][3]. Several research groups studied gesture interfaces for visually impaired [4][5]. Kane et al. studiedabout the drawing performance of visually impaired along withthe type of gestures [4]. Their results showed that visually impaired people prefer edge-based gestures and tapping gestures rather than shape and symbol type gestures. Guerreiroet al. suggested a novel character input method for visually impaired using directional gestures and showed better performance than the previous methods in the user study [5].Modzelewski investigated performance of graffiti gesture recognition algorithms for touchpad wearable on wrist.

Due to visually impaired users cannot have an adequate visual feedback about their posture, the previous studies were conducted mostly about the finger gesture input on touchscreen devices with existing natural or synthetic haptic cues, rather than gestures in the air without a physical contact of users' body and device.

In this paper, we suggest TV control interface based on bare-hand gestures designed for visually impaired. An approach using bare-hand gesture can easily extended to support control by multiple users and is free from difficulties occurred when the user find and try to make a proper contact with a remote controller.

First, we collected demands of visually impaired users about TV control via focused group interviews. Based on the demands, we designed two sets of TV control gestures for several frequently used functions. Then, we implemented a gesture recognition system. We used a depth camera to recognize bare-hand gestures. In a general TV watching environment, users are sitting in front of a display with a few meters distance. This condition meets well with the working range of commercial depth cameras. We used Kinect V1, which is affordable and enables easy tracking of users' body skeleton. We tested the usability of our gesture and recognition

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 11, March 2017

system compared to a gesture recognition system equipped on a commercial TV set. The results of user study showed subjective performance of the current system and raised important issues for the further developments.

II. GESTURE DESIGN

It is very hard to design a set of gestures with optimal usability. This study is our initial attempt to design a gesture interface for visually impaired, and we started from gathering users' demands. Then we derived several design criteria for the gesture interface and built two candidate sets of gestures cover frequently used TV control operations: channel/volume control, switching between electronic program guide (EPG) and

This research was funded by the MSIP (Ministry of Science, ICT & Future Planning), Korea, in the ICT R&D Program 2016

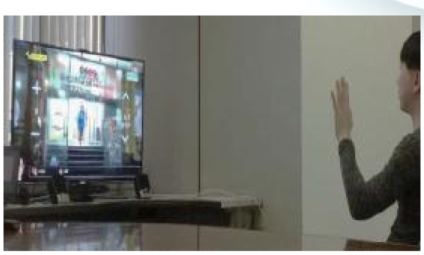
.

Fig. 1 Pointer type hand gesture interface on commercial smart TV.

program view, four directional navigation, selection and cancellation.

A. Needs from Visually Impaired Users

We conducted three focused group interviews with visually impaired users about TV control interface. Six blind users and six partially sighted users participated in the first and the second interviews, respectively. Their main concerns about the current interfaces were the lack of auditory feedback which let them know the current status, and reliability of the gesture interface.

Eight blinds and nine partially sighted people participated in the third interview. The third interview was conducted after a short experience (about 15 mins.) of the gesture-based interface in a commercial smart TV set (Samsung UN46ES7000F, released in 2012). The smart TV set detects a user's palm from color images captured via an RGB camera attached on the top of display. The user can move a pointer to buttons on the screen by moving their palm on the plane parallel to the screen. Each button is corresponding to an operation of TV and the user can select it by closing and

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 11, March 2017

opening their palm (See Fig. 1.).

The participants were asked to do several tasks (channel/volume control, page scroll, smart menu switching, and run of an application) using the pointer-type gesture in smart TV set and reported their experiences about it. None of the blind participants could successively finish the tasks. They commonly complained about absence of voice feedback, fatigues on their arm, and small working area of gesture recognition. Four of the partially sighted participants failed to finish the all tasks. They also commented about the absence of voice feedback and small working area. Some participants showed their preference to gestures on touchpad and traditional remote controller.

B. Design Criteria

We could derive several criteria for the design of TV control gesture set.

- The gesture should be recognized robustly. The most visually impaired users feel embarrassed at unexpected operations and the absence of visual cue make them hard to perform gesture motions accurately. Therefore, the recognition system should be tolerable at large errors in motion, keeping the gesture distinguishable to the other gestures and meaningless motions.
- Sufficient feedback must be given to user. Visually impaired people mostly rely on information from senses of sound and touch. Haptic feedback is hard to be applied on bare-hand interface and their information transfer rate is relatively lower than auditory channel. Voice representation is preferred to transfer feedbacks for a user's gesture commands.
- Required physical load should be small when a user perform the gesture. A large, excessive, or long sequence of motion is not only inconvenient to user, but also can cause user's fatigue.
- The gesture should be easy to learn. The most of people are not familiar with gesture interfaces. To reduce the users' mental load, some metaphors familiar to users can be utilized to build an intuitive gesture set. In addition, complex gestures are difficult to be explained verbally for visually impaired users.
- Performance limits of the current technology should be considered. Though small finger motions are preferred

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 11, March 2017

for gestures by users, it is hard to track the finger motion reliably at TV watching distance (> 2 *m*) from an affordable depth camera. When the user's arm is overlapped with torso or the other arm, skeleton tracking can be unstable, too. [6] discussed about an eye blinking sensor. Nowadays heart attack patients are increasing day by day."Though it is tough to save the heart attack patients, we can increase the statistics of saving the life of patients & the life of others whom they are responsible for.

C. Linear Gesture Set

We designed the first candidate of a gesture set with simple linear motions of hands. A gesture trigger was introduced to discriminate the followed gesture motions from meaningless motions appeared while watching TV. Two-handed gestures can be more robust than one handed gestures. As shown in Fig. 3, a frontal stretching motion of both arm was selected to the trigger. The trigger allows recognition of followed gestures unless there are no gestures for a while.

The gesture for 'select' operation is designed as frontal stretch of left hand which is mimicking the motion when we press a button. The gesture for 'cancel' is a motion waving right hand sideways. The 'select' and 'cancel' operations have opposite meaning each other and their corresponding gestures use opposite hands for clear discrimination.

Channel and volume controls are the most frequently used functions while watching TV. We used a directional mapping to increment and decrement of channel and volume which is a similar way to the buttons on general remote controllers. First, we mapped channel control to the left side and volume control to the right side. Intuitively, increment and decrement corresponds to the upside and downside, respectively.

Consequently, we could express the commands to the four directional stretches of left arm (channel) and right arm (volume).

Fig. 2 (at the middle of the right column) is demonstrating these gesture mapping.

For the mode change between program watching and EPG, we used a metaphor of motions when we open and close a window, in designing the corresponding gestures. A stretching motion of both arms sideways was selected as the gesture to

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 11, March 2017

'open' the EPG, and the gesture set to 'close' the EPG was set

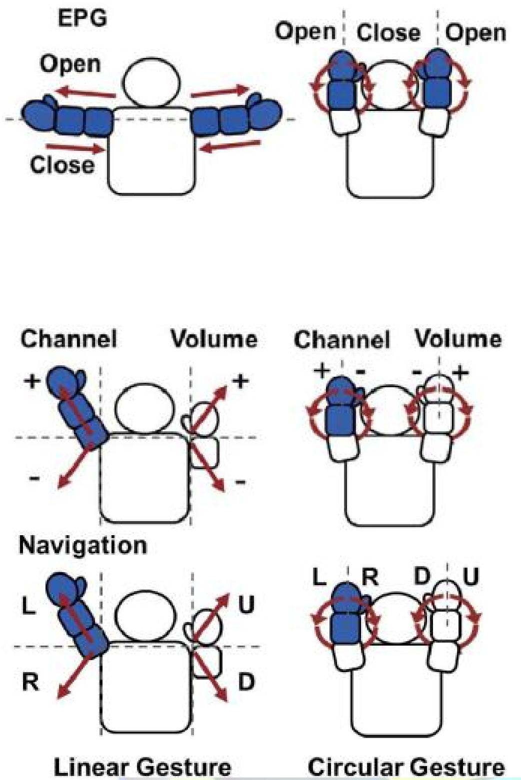


Fig. 2 Separately designed motions for the two types of gesture: linear and
circular (rear view). The blue colored parts are moved.

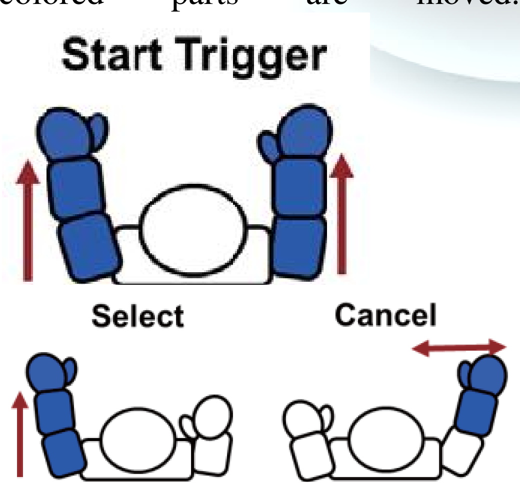


Fig. 3 Designed motions commonly used for the two types of gesture (top view). The blue colored parts are moved.

to the opposite motion: the user should gather hands to torso,
start from the finished pose of the 'open' gesture.

We reused the gestures for channel and volume control to
the gestures for the cursor navigation in EPG and menu. For the four directional navigation, left and right hands matched to
horizontal and vertical cursor movements, respectively. As
shown at the bottom of the right column in Fig. 2, stretching
gestures in left-up, left-down, right-up, and right-down
directions respectively correspond to the cursor movements of
left, right, up, and down.

D. Circular Gesture Set

The circular gesture set was designed as an alternative of
the linear gesture set with smaller and easily repeatable
motions. The circular set has different gestures for channel and
volume control, cursor navigation, and mode change between
program watching and EPG. Trigger, select, and cancel
gestures of the linear set are shared for the circular set.

We designed one-hand circular motion for the

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 11, March 2017

channel/volume control and cursor navigation. The operation of TV was differed by hand in motion (left or right) and the direction of rotation (inward or outward). As defined in the linear set, left and right hands correspond to the channel and volume control, respectively. Outward rotation (counterclockwise on left hand, and clockwise on right hand) was matched to the increment of channel or volume and inward rotation was matched to the decrement. For the cursor navigation, left and right hands were assigned with the horizontal and vertical movements again. An outward rotation was mapped to the leftward or upward movements, and an inward rotation mapped to the rightward or downward movements.

A two-handed gesture was assigned on the mode change between program watching and EPG. Outward rotations of left and right hands were mapped to the 'open' operation for EPG and inward rotations mapped to the 'close' operation, similarly with those of the linear gesture set.

We depicted the gesture motions in the circular set on the right column of Fig. 2.

III. IMPLEMENTATION

We implemented a windows PC based system which can recognize the designed linear and circular gesture sets. Kinect for windows sensor (model V1, Microsoft) was selected as a depth camera due to its affordable price for personal use and well-structured software development environments with builtin skeleton tracking from the captured depth images.

Overall process for gesture recognition is as follows. First, the Kinect camera captures user's motion and generate depth images at 30 frames/s. The skeleton structure of the user is derived from the depth image. A motion segment for each of arm is found from the sequence of nodes on the skeleton. A set of motion segments are compared to the rule-based gesture detector. If a gesture is detected, a corresponding control command is transferred to the TV receiver and a voice feedback is given to the user.

A. Skeleton Tracking

We used the built-in skeleton tracking function in Kinect SDK V1.8 for C++. The SDK supports two different modes for skeleton tracking. A skeleton has twenty nodes for whole body in standard mode and ten nodes for upper body in seated mode.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 11, March 2017

Since we assumed the user is seating while watching TV, we used the seated mode and utilized nodes on two arms (each node on hand, wrist, elbow, and shoulder and a head

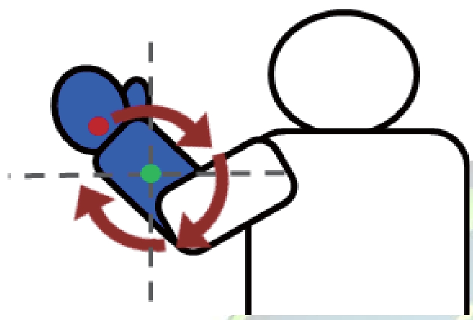


Fig. 4 A center point and quadrants for circular movement detection.

TABLE I. TASKS IN USER STUDY

Order

Gesture Set Type

Pointer Linear Circular

1 Increase channel by 3.

2 Decrease channel by 3.

3 Increase volume by 3 levels.

4 Decrease volume by 3 levels.

5 Switch to 'Smart home'. Switch to EPG.

6 Scroll left and right. Move right 4 times, and down 3 times.

7 Select an app icon. Select and cancel.

8 - Switch to watching mode.

user's frontal direction can be rotated along the user's posture. Therefore, we normalized the frontal direction of the skeleton for easier decision of gesture motion.

B. Gesture Recognition

Using the normalized coordinates of user's body, gesture commands are detected from user's motion. We used different detectors for the linear and circular gestures.

For the linear gesture set, we calculated elbow angle and motion vector of wrist. We used dot product of upper arm and lower arm to calculate the elbow angle. We can define a coordinate of an arm joint at n-th frame as $p_{i,j}(n)$: i is direction of the arm (0: left, 1: right), and j is index for arm joint from the torso (0: shoulder, 1: elbow, 2: wrist). The dot product of unit directional vectors for upper arm and lower arm, $dot_i(n)$, can be derived as in (1).) p (n)

If $dot_i(n)$ is larger than a predefined upper threshold (0.9, about 25.7° • in angle), the upper and lower arm are close to parallel and we considered the arm is 'stretched'. If the dot product is smaller than the lower threshold (0.8, about 36.8°), the arm is considered to be 'folded'. When the user's arm posture is changed between 'folded' and 'stretched', we checked motion vector of the wrist using the distance between two shoulders, $|p_{0,0}(n) – p_{1,0}(n)|$, as a unit

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 11, March 2017

length $l(n)$. If $|pi,2(n) – pi,2(n\text{-}3)| < 0.1\ l(n)$, the arm is considered in 'pause' state, and we find the latest motion vector, $|pi,2(n) – pi,2(n\text{-}k)| > 0.2\ l(n)$ where $k < 100$. When the motion vector is successively found, we find direction of the motion vector by finding the maximum dot product with ten directional unit vectors (eight *xy*-planar vectors and two *z*-axial vectors).

For the circular gesture set, we checked wrist position to find the period of a rotation. We find smallest k with a constraint of $|pi,2(n) – pi,2(n\text{-}k)| < 0.1\ l(n)$, where $20 < k < 100$.

A center point was estimated as an average of wrist positions from *n-k* to *n.* The wrist positions are categorized into four quadrants on *xy*-plane using the center point (See Fig. 4). A rotary hand motion is detected when the wrist positions moved along the four quadrants sequentially in clockwise or counterclockwise direction.

We keep the detected movement of a hand for ten frames (about 333 ms). If the movements on both side detected in this interval can be regarded as a two-handed gesture, otherwise, they regarded as two individual one-handed gestures.

C. Interfaces

The recognized gesture command is converted into a keyboard event to control the TV on the PC. A voice feedback is given to the user to tell the successful recognition of the gesture. The voice is synthesized using text-to-speech (TTS) engine built in Microsoft Windows 7.

IV. USER STUDY

A short-term user study was conducted to test feasibility of the designed gesture sets compared to the commercial pointertype gesture interface.

A. Participants and Apparatus

Sixteen visually impaired people (eight blinds and eight partially sighted persons; nine males and seven females) participated in the experiment. A smart TV (Samsung UN46ES7000F) was used for a display and commercial pointer-type gesture interface. The linear and circular gesture sets are recognized on a laptop (Sony VPC-SA27GK) operated with Microsoft Windows 7. A USB HDTV receiver dongle is used with Windows media center to receive and show the TV and EPG. The TV display was placed on a table with a Kinect sensor. A chair is located at 2 *m* distance in front of the TV

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 11, March 2017

display and the Kinect sensor.

B. Experimental Design

We used a within-subjects design of three gesture sets: pointer, linear, and circular. The sixteen participants experienced each of the three gesture sets in random order. The evaluation was conducted to measure five subjective factors brought from NASA-TLX [7]: mental demand ("How mentally demanding was the task?"), physical demand ("How physically demanding was the task?"), own performance ("How successful were you in accomplishing your level of performance?"), effort ("How hard did you have to work to accomplish your level of performance?", and frustration level ("How stressed, and irritated were you?").

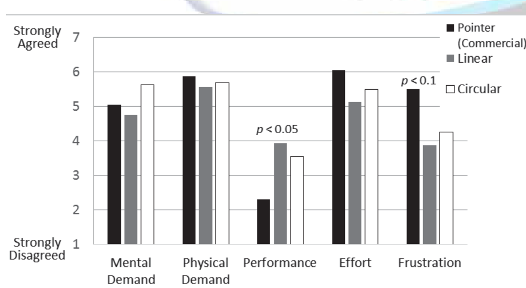


Fig. 5 Averaged subjective ratings of the three gesture interfaces.

C. Procedures

First, a participant sat on the chair and listened about the experiment. For a gesture set, the participant learned the gestures and exercised for 5 mins. Then, the participant performed given tasks shown in Table 1. The participant rated the gesture interface for five factors (verbally explained in Korean) and gave subjective feeling verbally. The participants rated the interface for each factor using a Likert-type 7-scale (1: strongly disagreed, 4: normal, 7: strongly agreed). After a 3-min. break, a session for the next gesture set started. The experiment took about 40-50 mins. for a participant.

D. Results and Discussion

Subjective ratings for the five factors were averaged for participants and represented in Fig. 5. The three types of interfaces have similar average scores in mental demand, physical demand, and effort. An ANOVA test conducted for each rating factor showed a significant difference in performance ($F(2, 45) = 4.00$, $p = 0.0252$), and marginal significance in frustration ($F(2, 45) = 2.88$, $p = 0.0663$). Tukey multiple comparison tests showed significantly larger

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 4, Special Issue 11, March 2017

performance of the linear and circular sets than the pointertype interface, and less frustration of the linear set than the pointer-type.

The subjective rating showed equal or better performance of our linear and circular gesture sets than that of the commercial pointer-type gesture. However, overall performance of the gesture interfaces were not so good. The participants rated that the gesture requires mentally and physically higher load than the normal, with performances below or similar to normal. The participants needed their effort and frustrated while they were using the gesture interfaces. Differences between linear and circular set were not significant in all factors.

The participants commonly reported that pointer-type gesture is inappropriate to visually impaired users (8 of 16) and hard to locate the cursor on an icon (8 of 16). For the linear set, some participants concerned about large gesture motions (4 of 16). Many participants had difficulties in memorizing and matching the circular gestures to the corresponding operations (8 of 16).

In design of the gestures, we expected the circular set can reduce users' physical loads and will show better usability than that of the linear set. However, this was not shown in the results of user study. In this study, we measured the usability for a short time of experience. Therefore the participants might not be sufficiently familiar with the gestures and could not feel the significant difference in a short task sequence.

V. CONCLUSION

We designed gesture set for the TV control of visually impaired people. Two sets of bare-hand gestures were designed using linear motions and circular motions for frequently used control functions on TV. The designed gesture sets showed equal or better subjective ratings than those of the commercial pointer-type TV gesture interface.

This study is our initial attempt on gesture interface for visually impaired, and there is a far way to go. The further research should include: 1) effective voice feedback method, 2) smaller physical load, 3) more intuitive design, and 4) better robustness. We hope that our study can give better understanding and some idea to design user interfaces for visually impaired people.

ACKNOWLEDGMENT

The authors thank to Wansik Kang and Eunha Lee at Korea Blind Union for their help in this study.

REFERENCES

[1] Webpage: World Blind Union, “White Cane Day 2014,” http://www.worldblindunion.org/English/news/Pages/White-Cane-Day-2014.asp.

[2] I. Guyon, V. Athitsos, P. Jangyodsuk, H. J. Escalante, and B. Hamner. “Results and analysis of the ChaLearn gesture challenge 2012,” In Advances in Depth Image Analysis and Applications, pp. 186-204. Springer Berlin Heidelberg, 2013.

[3] D. Ionescu, B. Ionescu, C. Gadea, and S. Islam, “An intelligent gesture interface for controlling TV sets and set-top boxes,” In Proceedings of IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI), pp. 159-164. 2011.

[4] S. K. Kane, J. O. Wobbrock, and R. E. Ladner, “Usable gestures for blind people: understanding preference and performance,” In Proceedings of the CHI, pp. 413-422. 2011.

[5] T. Guerreiro, P. Lagoá, H. Nicolau, D. Gonçalves, and J. A. Jorge, “From tapping to touching: Making touch screens accessible to blind users,” IEEE MultiMedia, 15(4), pp. 48-50. 2008.

[6] Christo Ananth, S.Shafiqa Shalaysha, M.Vaishnavi, J.Sasi Rabiyathul Sabena, A.P.L.Sangeetha, M.Santhi, “Realtime Monitoring Of Cardiac Patients At Distance Using Tarang Communication”, International Journal of Innovative Research in Engineering & Science (IJIRES), Volume 9, Issue 3,September 2014,pp-15-20