2014 CMBEC37 Conference
Vancouver, BC
May 21 – 23, 2014

A STUDY OF MULTI-MODAL INTERFACE IN ROBOTIC ASSISTIVE SYSTEMS

Carlos Diaz and Shahram Payandeh
Experimental Robotics and Imaging Laboratory, Simon Fraser University, Burnaby, B.C. Canada (email: carlosd@sfu.ca), (email: payandeh@sfu.ca).

INTRODUCTION

Several medical conditions can limit the ability of visually impaired people to obtain information about their environment. Such information is critical for an effective navigation. Remote sensing of objects could allow users to detect and avoid obstacles and enrich their perception.

One of the potential applications of a Multi-Modal Interface (MMI) is the use as a navigation aid for visually impaired people. Zöllner et al. [1] developed a navigation MMI for blind users using in a depth camera attached to a helmet to measure the distance of objects in front of the user. The system provided haptic feedback using a wearable belt. Other authors have developed techniques to remotely locate objects and people using 3D image processing for partially or totally blinded people using visual, auditory and tactile feedback [2-4].

Many MMI systems rely on hand gestures as an input. Several approaches have been proposed for hand gesture recognition. Lee used stereo vision for gesture detection [5] while Prisacariu [6] and Wang [7] developed color based segmentation techniques for bare hands and color gloves to deal with the background segmentation problem using video cameras.

In this paper we present the design of a multimodal user interface based on a 3D sensor to gather spatial information (Microsoft Kinect) from a scene, a haptic glove with vibrotactile feedback and a gesture recognition system to map the location, dimensions and shape of sample objects into the user hands (Figure 1). The user can perceive the shape, location and dimensions of the remote objects by moving the glove inside a scanning region. A marker detection camera provides the location and orientation of the user hand (glove) to map the corresponding tactile message. Additionally a gesture recognition subsystem was implemented providing the option to interact and control active elements such computer interfaces or automated devices.

SYSTEM OVERVIEW

The conceptual diagrams of our proposed MMI system concept in shown in Figure 1 The system has four main components: a) Bracelet Location Subsystem, b) Gesture Subsystem, c) Depth Imaging Subsystem and d) A Haptic Glove.

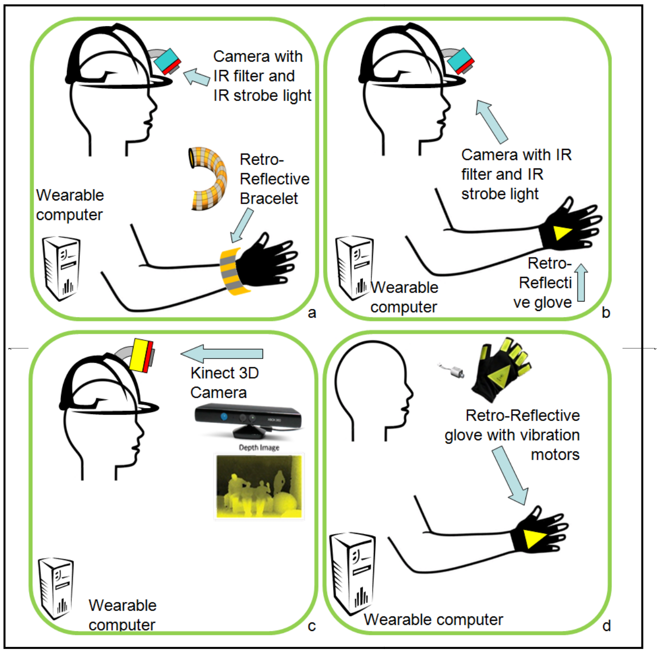

Camera with IR filter and IR strobe light
Retro-Reflective Bracelet
Wearable computer
a
Camera with IR filter and IR strobe light
Retro-Reflective glove
Wearable computer
b
Kinect 3D Camera
Depth Image
Wearable computer
c
Retro-Reflective glove with vibration motors
Wearable computer
d

Figure 1: Subsystem architecture: Bracelet Location subsystem (a), Gesture subsystem (b), Depth subsystem (c) and Haptic subsystem components (d).

Each module was designed and evaluated to collect metrics which can be used to measure the performance each sub-system will contribute to the overall performance. Figure 2 shows a diagram of the experimental setup.

A Kinect camera capture images of the sample objects placed on the platform. In the same figure it is shown the "object sampling region" under a green pyramid (green dotted line). The Kinect sensor covers the sampling region of the table within the field of view of the camera Objects placed on top of the sampling region can be measured and classified according to some basic geometrical shapes by the depth subsystem.

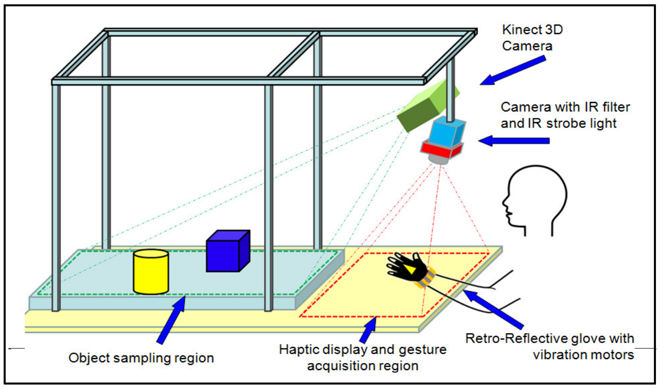
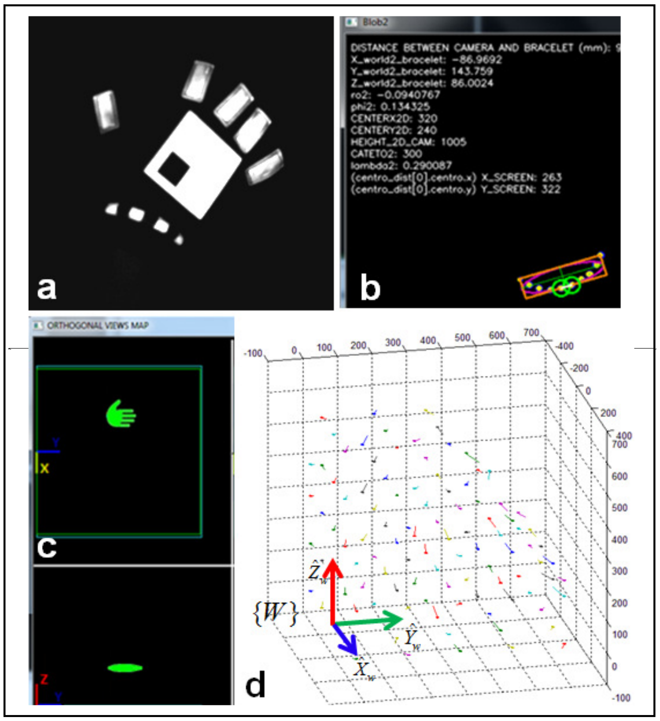

Kinect 3D Camera
Camera with IR filter and IR strobe light
Object sampling region
Haptic display and gesture acquisition region
Retro-Reflective glove with vibration motors

Figure 2: The experimental setup.

Connected below the Kinect sensor is a grayscale 2D camera with an infrared LED ringlight to provide coaxial illumination. The field of view of this camera covers the "haptic display and gesture region". Within this region, shown in Figure 2 as a red-dotted pyramid, the user is able to present different gestures or hand poses, the gesture subsystem is able to identify the gesture displayed from the vocabulary of gestures to trigger events in the computer platform.

With the data provided by the 3D and 2D subsystems, it is possible to build a map of haptic stimuli. This map of vibrotactile sensations is presented to the user as feedback data with the activation of the tactons on the glove. Each motor can be activated and modulated independently according to the required information to be displayed. By moving the hand, in a scanning motion along the haptic region, it is possible to display haptic messages corresponding to different areas of the sampling region. Data measured from objects placed on the sample platform can be encoded as tactile stimuli on the haptic display area.

BRACELET LOCATION SUBSYSTEM



```
DISTANCE BETWEEN CAMERA AND BRACELET (mm): 9
X_world2_bracelet: -86.9692
Y_world2_bracelet: 143.759
Z_world2_bracelet: 86.0024
ro2: -0.0940767
phi2: 0.134325
CENTERX2D: 320
CENTERY2D: 240
HEIGHT_2D_CAM: 1005
CATETO2: 300
lambda2: 0.290087
(centro_dist[0].centro.x) X_SCREEN: 263
(centro_dist[0].centro.y) Y_SCREEN: 322
```

Figure 3: Bracelet and glove under IR illumination (a). Distance between markers measurements (b). Bracelet is used to estimate the location of the wrist of the user in 3D coordinates (c). Plot of the error in computed 3D location Vs. ground truth of the bracelet (d).

Segments of reflective tape on the bracelet with a constant separation provide an invariant feature to compute the distance between the bracelet and the camera using a pinhole camera model. With this information it is possible to calculate the position of the bracelet (and the wrist of the user) in a 3D space representation.

To improve the precision of the measurements, it was required to estimate distortion parameters in the camera-lens assembly. Camera calibration algorithm [8, 9] was used to rectify images reducing lens distortion effects in the two cameras used.

In a raw image, reflective elements corresponding to the glove and bracelet are present (Figure 3.a). The location routine starts with the identification of bracelet markers. These elements are isolated based on combined properties of individual blobs and blobs pairs.

Once the "best candidates" from the bracelet are selected, the distance in pixels between the markers close to the center of the bracelet is measured and evaluated using the pinhole camera model equation (Figure 3.b).

Using the known feature, and the focal length, the pinhole model provides an estimate of the distance between the wrist of the user and the camera. The system is able to compute the 3D position of the bracelet with an average error of 22.07mm and a standard deviation of 11.55mm with respect to the World frame. Figure 3.d shows the error vectors for the bracelet position in space.

GESTURE SUBSYSTEM

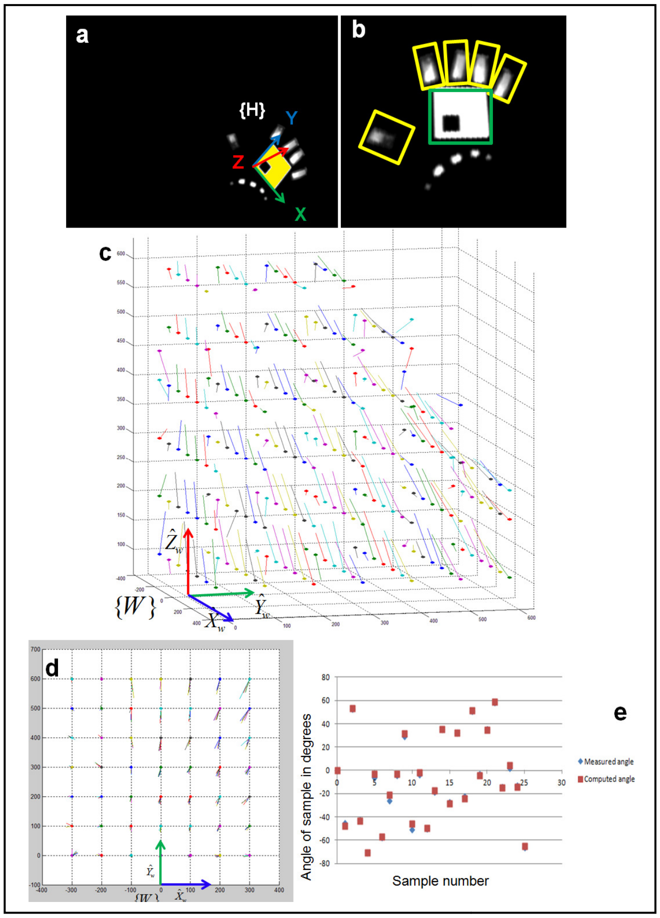


Figure 4: Gesture recognition system. A square pattern on the hand is used to apply perspective correction (a) and (b). Error in position and orientation between computed values and ground truth (c) (d) and (e).

A pattern of reflective tape fixed to the glove is used to compute the position and orientation of the hand in 3D space. Assuming that the observed pattern is under perspective deformation it is possible to recover the original square shape. Figure 4.a show the process of assigning a frame to the "candidate" square region. To compute the perspective transformation, the four corners of a trapezoidal figure are used.

Perspective correction is applied to the hand image (using the inverse perspective transformation) to obtain a normalized image in size, rotation and perspective. A fixed mask or regions of interests is applied over the normalized image to count the number of fingers displayed in the gesture (Figure 4.b).

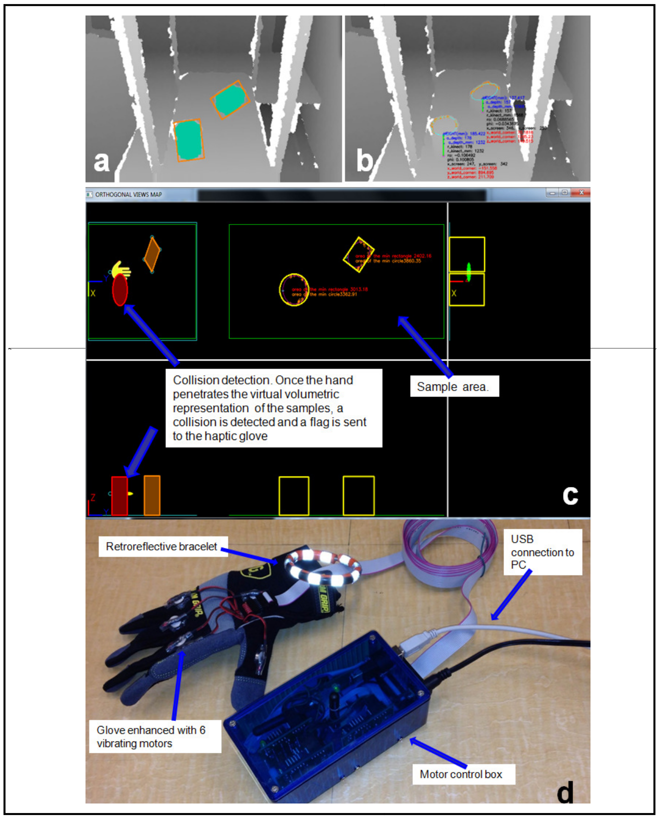

a
b
ORTHOGONAL VIEWS MAP
Collision detection. Once the hand penetrates the virtual volumetric representation of the samples, a collision is detected and a flag is sent to the haptic glove
Sample area
c
Retroreflective bracelet
USB connection to PC
Glove enhanced with 6 vibrating motors
Motor control box
d

Figure 5: The depth image of sample objects (a)(b). Objects are mapped in an orthogonal representation (c). Tactile representation of the sampled object (in orange and red) (c). Vibrating elements in the glove (d)

Different hand poses are combined in time to assemble dynamic gestures. The "Double click" of a mouse can be implemented by a rapid sequence of close-hand and open-hand transition.

Experiments using a pool of images previously recorded reported an average error in 3D location of 45.42 mm with a standard deviation of 24.69mm. Location error is plotted in Figure 4.c. Additionally, average error in pitch, yaw and roll angles were 11.4, 3.42 and 10.1 degrees respectively. The gesture recognition rate was 87.3%

DEPTH IMAGING SUBSYSTEM AND HAPTIC INTERFACE

Depth images from the 3D camera are processed to find the dimensions and position of the samples on the table. Figure 5.a shows the contours of 2 objects using background subtraction. The found contours are analyzed in detail to find the higher points that describe the height of the objects

Objects are modeled as a volume with constant section along its height. They are mapped in an orthogonal frame representation with respect to the World frame {W} (Figure 5.c). A tactile representation of the sampled objects (in orange and red) (Figure 5.c) is displayed to the user activating the tactons in the glove. When the user's hand "penetrates" the volume corresponding to one particular sample, the motors are activated indicating such condition.

The performance of the depth system was tested using 8 different objects with different height and section. The average error in the determination of the height of the objects is 27.71 mm. The average width error is 25.11mm and the depth average error is 31.17 mm. (measured in local object coordinates). The position of objects, measured from the World frame of the system reported an average error of 43.27mm in the Y axis and 35.90 and 29.48 in X and Z axis of {W}.

DISCUSSIONS AND CONCLUSIONS

Both bracelet and gesture systems provide information about the position of the user hand in the 3D space. It was shown that the error of the former system was smaller than the latter (22.0 Vs. 45.42mm). In spite of this, the gesture detection approach provides a useful tool to compute the orientation of the hand frame in space and normalized hand recognition in size, orientation and perspective.

The gesture detection failed in cases where the pitch or roll of the hand is out of the -65, +65 degrees range. Under this condition the reflectivity gain of the markers is reduced over 5 decibels making it difficult to find the hand pattern. Yaw angle can be computed with a small error in any case where the hand pattern is found.

Users could distinguish between samples at different locations based on the haptic feedback. Distance perception was reduced when samples were close to each other (under 10cm). Objects with different heights were identified properly. The success of recognition is higher when samples have more than 6cm of difference in height. Shape recognition was not determinant in object identification

REFERENCES

[1] M. Zöllner, S. Huber, H. Jetter and H. Reiterer. "NAVI–A proof-of-concept of a mobile navigational aid for visually impaired based on the Microsoft kinect," in *Human-Computer Interaction – INTERACT 2011*, P. Campos, N. Graham, J. Jorge, N. Nunes, P. Palanque and M. Winckler, Eds. 2011, . DOI: 10.1007/978-3-642-23768-3_88.

[2] V. Filipe, F. Fernandes, H. Fernandes, A. Sousa, H. Paredes and J. Barroso. Blind navigation support system based on microsoft kinect. *Procedia Computer Science 14*pp. 94-101. 2012.

[3] S. L. Hicks, I. Wilson, L. Muhammed, J. Worsfold, S. M. Downes and C. Kennard. A depth-based head-mounted visual display to aid navigation in partially sighted individuals. *PloS One 8(7),* pp. e67695. 2013.

[4] V. Khambadkar and E. Folmer. GIST: A gestural interface for remote nonvisual spatial perception. Presented at Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology. 2013.

[5] M. Lee, M. Billinghurst, W. Baek, R. Green and W. Woo. A usability study of multimodal input in an augmented reality environment. *Virtual Reality 17(4),* pp. 293-305. 2013.

[6] V. A. Prisacariu and I. Reid. 3D hand tracking for human computer interaction. *Image Vision Comput. 30(3),* pp. 236-250. 2012.

[7] R. Y. Wang and J. Popović. Real-time hand-tracking with a color glove. Presented at ACM Transactions on Graphics (TOG). 2009.

[8] C. B. Duane. Close-range camera calibration. *Photogrammetric Engineering 37(8),* pp. 855-866. 1971.

[9] Z. Zhang. A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions On 22(11),* pp. 1330-1334. 2000.