A Real-Time Hand Gesture Interface for a Medical Image Guided System

Juan Wachs[1], Helman Stern[1], Yael Edan[1], Michael Gillam[2], Craig Feied[2], Mark Smith[2], Jon Handler[2]

[1]Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, Be'er-Sheva, Israel, 84105, {helman, yael,juan}@bgu.ac.il.
[2]Institute for Medical Informatics, Washington Hospital Center, 110 Irving Street, NW, Washington, DC, 20010, {feied,smith,handler,gillam}@medstar.net

Abstract - *In this paper, we designed a gesture interface for users, such as doctors/surgeons, to browse medical images in a sterile medical environment. A vision-based gesture capture system interprets user's gestures in real-time to manipulate the objects for an image and data visualization environment. Dynamic navigation gestures are translated to commands based on their relative positions on the screen. The gesture system relies on real-time robust tracking of the user's hand based on color-motion cues. A state machine switches between other gestures such as zoom and rotate. A prototype of the gesture interface was implemented in a sterile medical data-browser environment.*

Keywords: *hand gesture recognition, medical databases, browsing, image visualization, sterile interface*

1 Introduction

Computer information technology is increasingly penetrating into the hospital domain. It is important that such technology be used in a safe manner to avoid serious mistakes leading to possible fatal incidents. Keyboards and mice are today's principle method of human – computer interaction. Unfortunately, it has been found that a common method of spreading infection from one person to another involves computer keyboards and mice in intensive care units (ICUs) used by doctors and nurses [1]. Kiosks using touch screens [2] introduced recently into hospitals, to provide patient information, bring no guarantee to stop the spread of bacteria (such as an outbreak of SARS). When an epidemic crisis erupts access to information is absolutely critical, and kiosk users may forego the washing of hands in the interest of speed.

By the early 1990's scientists, surgeons and other experts were beginning to draw together state of the art technologies to develop comprehensive frameless image-guidance systems for surgery, such as the StealthStation [3]. This is a free-hand stereotactic pointing device, which transmits its position via attached light emitting diodes (LEDs), and converts this position in to the corresponding location in the image space of a high-performance computer monitor. Also, touch-screens are a popular means of interaction. As in traditional POS (point of sale) environments, one style of touch screen does not work in all healthcare environments. In a hospital, different departments will insist on different touch screen characteristics. Medical offices want large screens, with large buttons, to help reduce training time [4]. In a setting like an operating room [OR], touch screen displays must be sealed to prevent the buildup of

contaminants, and should also have smooth surfaces for easy cleaning with common cleaning solutions.

Many of these deficiencies may be overcome by introducing a more natural human computer interaction (HCI) mode into the hospital environment. The basis of human-human communication is speech and gesture including facial expression, hand and body gestures and eye gaze. Some of this concepts where exploited in systems for improving medical procedures and systems. In FAce MOUSe [5], a surgeon can control the motion of the laparoscope by simply making the appropriate face gesture, without hand or foot switches or voice input. Current research to incorporate hand gestures into doctor-computer interfaces appeared in Graetzel et al. [6]. They developed a computer vision system that enables surgeons to perform standard mouse functions (pointer movement and button presses) with hand gestures. Zeng et al. [7] by tracking finger position are able to gather quantitative data about the breast palpation process for further analysis. Other systems [8] suggests a teleoperated robotic arm using hand gestures for multipurpose tasks. Another aspect of gestures is their capability to aid handicapped people by offering a natural and alternative form of interface. Wheelchairs, as mobility aids, have been enhanced as robotic/intelligent vehicles able to recognize the user's commands indicated by hand gestures [9]. The Gesture Pendant [10] is a wearable gesture recognition system that can be used to control home devices and provides additional functionality as a medical diagnostic tool. Staying Alive [11] is a virtual reality imagery and relaxation tool that allows cancer patients to navigate through a virtual scene using 18 T'ai Chi gestures. A tele-rehabilitation system [12] for kinesthetic therapy (treatment of patients with arm motion coordination disorders) was proposed using patient force-feedback gestures. Force-feedback was used, as well, in [13] to guide the adaptation of a teachable interface for individuals with severe dysarthric speech. A haptic glove was used to rehabilitate post-stroke patients in the chronic phase in [14]. These patients receive therapy while being immersed in a virtual environment (VE). In this paper we explore only the use of hand gestures, which can in the future be further enhanced by other modalities. Gesture capture is vision based and used to manipulate windows and objects, especially images, within a graphical user interface (GUI).

We propose a doctor-computer interaction system based on effective methods for analyzing and recognizing gestures in sterile dynamic environments such as operation rooms. Much of the research on real-time gesture recognition has focused exclusively on dynamic or static gestures. In our work, we consider hand motion and posture simultaneously. This allows for much richer and realistic gesture representations. Our system is user independent without the need of a large multi-user training set. Operation of the gesture interface was tested in a hospital environment in real-time. In this domain the non-contact aspect of the gesture interface avoids the problem of possible transfer of contagious diseases through traditional keyboard/mice user interfaces.

System specifications, architecture and methodology are presented in Section 2. In Section 3 image processing operations using color-motion fusion for segmentation of the hand from the background are described. Section 4 provides details of the tracking module, its mapping into navigational gestures, and state machine switching between other gestures such as zoom and rotates. Final conclusions are provided in section 5. The appendix provides a description of the Gibson 3D data browser used as our domain of application.

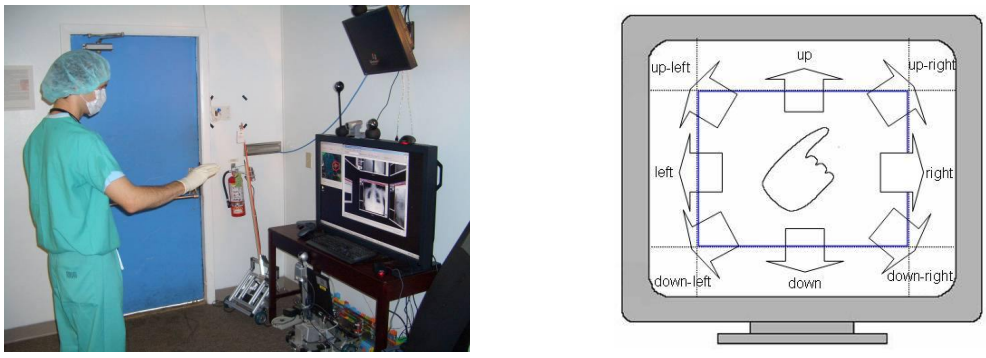
2 System Overview

2.1 System Specifications

Some structural characteristics of a gesture interaction model for a medical environment are presented in [15], and extended for the OR domain in [6]. For the correct design of a hand gesture interaction system for doctors/surgeons, the following specifications should be considered [16]: (1) Real time interaction - during surgery the surgeon can watch a computer monitor to see the position of hand gesture command. (2) Fatigue - gestural commands must be concise and rapid to minimize effort. (3) Intuitiveness – gestures should be cognitively related to the command or action it represents, (4) Unintentionality - most systems capture every motion of the user's hand, and as a consequence unintentional gesture may be interpreted by the system. The system must have well-defined means to detect the correct intention of the gesture. (5) Robustness - the system should be capable to segment hand gestures from complex backgrounds containing; object motion, variable lighting and reflected color, (6) Easy to learn – doctors/ surgeons are time pressed individuals, so long training times should be avoided. (7) Unencumbered –doctors/ surgeons may wear gloves and frequently hold instruments, so additional devices attached to the hand, such as data gloves, colored or infrared markers must be avoided. The above considerations should improve computer usability for doctors/surgeons in medical environments.

2.2 Architecture and Methodology

A web-camera placed above a screen (Figure 1(a)) captures a sequence of images of the hand. The hand is tracked by a tracking module which segments the hand from the background using color and motion cues. This is followed by black/white (BW) thresholding and various morphological image processing operations. The location of the hand in each image is represented by the 2D coordinates of its centroid. The temporal information is represented by the current frame number. This spatio-temporal data is mapped into a 'flick gesture'. A flick gesture is the rapid movement of the hand from a neutral position to a specific direction, and returned to the original state. The direction of the 'flick' is in one of eight possible navigation directions of the screen (see Figure 1(b)) and is used to navigate through a visual image data browser. The doctors/surgeons intended actions/commands are recognized by extracting features from the spatio-temporal data of the gestures. With these actions/commands doctors can bring up X-rays images, select a patient record from the database or move objects and windows in the screen. A two layer architecture is used. The lower level provides tracking and recognition functions, while the higher level manages the user interface.



(a) (b)

Figure 1 (a) Gesture capture, (b) Screen Navigation Map

3 Segmentation

3.1 Color Cue

The CAMSHIFT [17] algorithm is used to track and recognize gestures. Within the CAMSHIFT module, a probability distribution image comprised of pixels representing hand colors is created from a 2D hue-saturation skin color histogram [18]. This histogram is used as a look-up-table to convert the acquired camera images of the hand into corresponding hand pixels, a process known as back propagation. Thresholding to black and white followed by morphological operations is used to obtain a single component for further processing to classify the gestures.

The initial 2D histogram is generated in real-time by the user in the 'calibration' stage of the system. The interface preview window shows an outline of the palm of the hand gesture drawn on the screen. The user places his hand within the template while the color model histogram is built (Figure 2), after which the tracking module (Camshift) is triggered to follow the hand. This allows individuals with differing skin colors to operate the system.

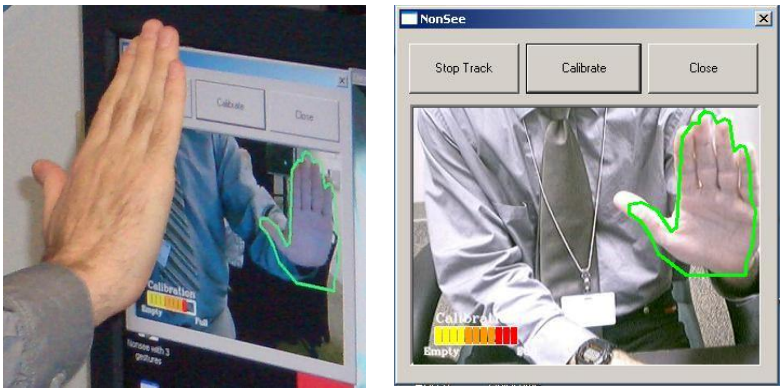


Figure 2. User hand skin color calibration

The calibration process is initiated by detecting the motion of a hand wave within the region of the template. To avoid false motion clues originated by non-hand motion a background maintenance operation is maintained. A first image of the background is stored immediately after the application is launched, and then background differencing is used to isolate the moving object (hand) from the background. Since background pixels have small variations due to changes in illumination over an extended period of time, the background image is dynamically updated. Background variations are identified by thresholding the absolute difference between two consecutive frames. If the difference is under some threshold t_1, then the current images contain only a background, otherwise, an upper threshold level t_2 is checked to test whether the present object is a hand. In case that the current image is a background, the background stored image is updated using a running smoothed average (see (1)).

$$Bcc_k(i,j) = (1-\alpha) * Bcc_{k-1}(i,j) + \alpha * f(i,j) \tag{1}$$

Here Bcc_k is the updated stored background image at frame k, Bcc_{k-1} is the stored background image at frame k-1, α is the smoothing coefficient (regulates update speed), f(i,j) is the current background image at frame k. Small changes in illumination will only update the background while huge changes in intensity will trigger the tracking module. It is assumed that the hand is the only skin colored object moving on the area of the template.

3.2 Motion Cue

The grayscale image obtained from the RGB channels is smoothed using a Gaussian filter. The absolute difference between two consecutives images is computed, and thresholded to convert the intensity image to BW. Morphological operations assist us to clean holes and small noise in the image (see Fig 3)

Figure 3. Segmentation using 'motion cue'

3.3 Color and Motion Fusion

As a result of the color cue we have an intensity image p_k, representing the skin color probability at frame k, and a second BW image used as a motion indicator ϕ_k, obtained from the motion cue at frame k. At frame k, I_k is the fused intensity image [19] according to (2).

$$I_k(i,j) = \alpha_k \min\{1, p_k(i,j).d\} * \phi_k(i,j) + (1-\alpha_k) * p_k(i,j) \tag{2}$$

Here d is an amplifying factor (d=1.3 for best performance), and α is a motion assessment variable which increases and decreases for large small amounts of motion. Motion indication reinforcement is introduced to overcome the weak ability of color only handle extreme color changes and noise from light variations. Motion only cannot be fully trusted because of the resultant halo effect, reflections and cast shadows. Also, color fusion avoids the defect of motion only, which detects not only the hand but the entire body movement

4 Hand Tracking and Operation Modes

The gesture interface was implemented in a medical database named Gibson (see Appendix), developed by IMI [20] for the purpose of interacting with medical images such as X-rays and MRIs. The finite state machine (Fig. 4) is used to illustrate the operational architecture of gesture system. Gesture operations are initiated by a calibration procedure in which a skin color model of the users hand is constructed. Control between dynamic gestures used for browsing through images and pose gestures (used for rotation and zoom) are affected by mode switch gestures. Superimposed over the image is a rectangular frame (Fig. 5). The area inside the frame is called the "neutral area". Movements of the hand across the boundary of the rectangle constitute directional browser commands. When a doctor decides to perform a specific operation on a medical image, he/she places the hand in the 'neutral area' momentarily, and an attention window event is called. The spatio-temporal information and other attributes of the posture are sent to a "mode detector" to determine whether a zoom or rotation pose gesture is presented.

4.1 Directional Navigation

Navigation gestures were designed to browse through a Gibson data browser. Gibson represents an image as a 3D object where each image is a flat rectangular plate formed around a cylinder and arranged in numerous levels. The cylinder can be rotate CW and CCW,

and moved vertically to exhibit various levels on the screen. Thus any image on the screen can be accessed directly by four navigation commands.

When a doctor/surgeon wishes to browse the image database, he/she moves the hand rapidly out from a 'neutral area' toward any of four directions, and then back to the neutral area. This movement is referred to as a 'flick' gesture.

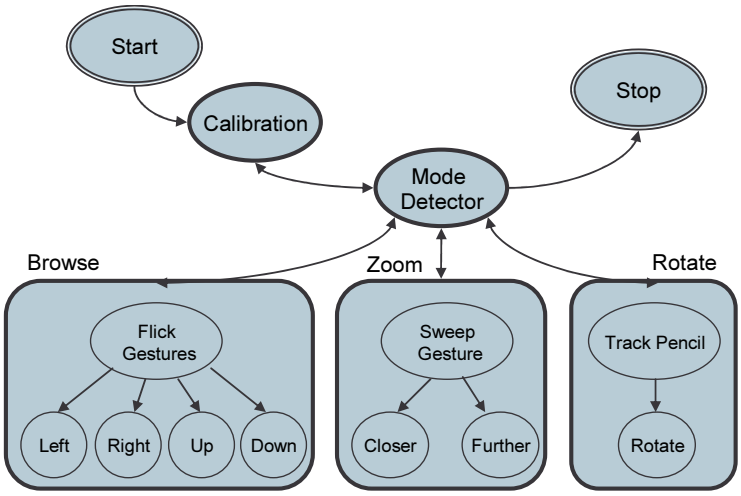


Figure 4. State machine for the gesture-based medical browser

Interaction is designed in this way because the doctor will often have his hands in the 'neutral area' without intending to control the Gibson data browser. Only when a flick gesture is moved towards one of the four four quadrants (left, right, up, down), is the image cylinder moved in the direction of the flick. Flick directions are detected when the hand moves from the neutral area, across the boundary of the rectangle, and repositioned back into the neutral area again (see Fig. 5).

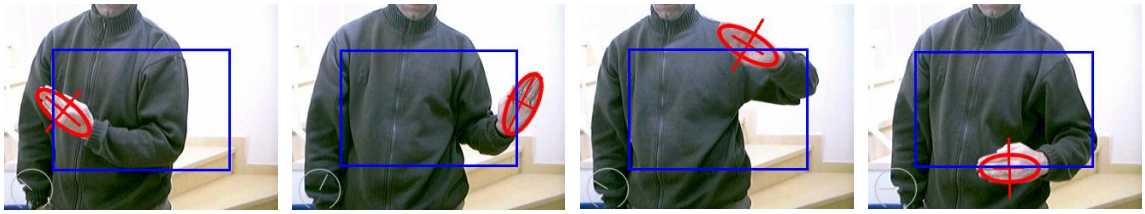


Figure 5. Four quadrants mapped to cursor movements

4.2 Zoom Mode

The main purpose of the zoom is to increase or decrease the size of an image. Once the zoom-mode is activated the size of image is increased according to the detected area of the hand as it moves toward and away from the screen. To go back to the normal mode, the hand is moved out from the neutral area to any of the 4 directions. The "zoom mode" is activated, when the hand is in the neutral area, by an abrupt rotation (sweep gesture) of the wrist. The rotation (sweep) must be counter clock wise, from 90º to 180º. To test if the hand is rotated 90º CCW from the vertical position, in a specified amount of time, the following procedure is conducted:

1. A temporal window of the last w frames is created. Here w=39 is used.
2. Four reference frames are marked. The first frame is the current frame at f_4=0 frames (current), f_3=-w/3 frames, f_2=-2/3w frames and f_1=-w frames. The

value of w used was 39 frames, so that the four reference frames are [f_4=0, f_3=-13, f_2=-26, f_1=-39]

Let,

α - The difference of the angles of consecutive frames.

$\overline{\alpha}_{(f_i,f_j)}$ The average angle between frames f_i and f_j.

$std(\alpha_{(f_i,f_j)}))$ The angle standard deviation between frames f_i and f_j.

ε The error margin allowed is in degrees, the epsilon used was 5º.

To enter into the zoom-mode, eq (3) and eq (4) must be true.

$$90 - 2 * \varepsilon \leq \overline{\alpha}_{(f_3,f_4)} - \overline{\alpha}_{(f_1,f_2)} \leq 90 + 2 * \varepsilon \quad (3)$$

$$std(\alpha_{(f_3,f_4)}) \leq 1.5 * \varepsilon \ \vee \ std(\alpha_{(f_1,f_2)}) \leq 1.5 * \varepsilon \quad (4)$$

4.3 Rotation

The rotation operation is helpful when the doctor wants to rotate the image to a desired angle. To pursue this goal, the physician/surgeon places a sterilized straight instrument in the fist of the hand, and holds it at least three meters from the camera. When the area of the tracking window becomes smaller than some threshold, the rotation mode is activated. When in rotation mode, the angle to which the medical image is rotated is determined by the angle made by the instrument and the horizontal axis. The instrument can be found in an image by applying the Probabilistic Hough Transform, referred to as pHT [21]. The pHT algorithm finds all the straight segments in an image longer than some minimum. These are good candidates to be the sides of the instrument; however they also may represent other straight lines such as: doors, windows, tables, wrist, etc. Using prior knowledge a set of crisp rules is applied to the set of lines to find the best line that represents the instrument. To quickly eliminate unlikely candidate lines far from the hand an expanded (2.5 times) window around the tracking window is cropped out of the image (see Figure 6).

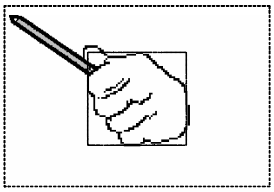


Figure 6. Tracking windows: original and resized

A canny edge detector is applied to the cropped image, with two thresholds t_1=50 and t_2=210, and a mask of 3x3. This is followed by the probabilistic Hough Transform with distance resolution=1, angle resolution of 1º, threshold of 30, minimum line length of 30, and the maximum gap allowed between line segments lying on the same line as two pixels. The end points of each line segment are returned from pHT. The closest end point to the hand is determined by finding the minimum distance of all end points to the centroid of the hand tracking window. With both endpoints identified it is possible to find the rotation angle represented by the hand. This angle, β_f, is measured CW from the horizontal of the current frame f. To select among all segments that most likely to be the instrument an evidence test is conducted. The test is comprised of a set of queries. If enough responses are positive, it is assumed that enough evidence has been accumulated to infer that the instrument has been identified.

For each candidate line segment i, found in the image, the following tests are made:

Initially find the length of all line segments.

1) Is i the longest of the lines?
2) Is the difference between the major axis of the tracking ellipse of the hand and β_f, small?
3) Is the change between $\beta_f - \beta_{f-1}$, small? (i.e.; within some small angle change).This is the change in the angle of the instrument found in the previous frame, and the angle in the current frame, f.
4) Draw a line from the centroid of the tracking window to the far endpoint of the instrument. Find the shortest distance from the close endpoint of the instrument to this line traced. Is this distance small?

Every positive response adds one vote to the total votes of a candidate line. The line with the highest number of votes is selected as the segment representing the instrument. An example of successful tracking of an instrument is shown in Fig. 7.

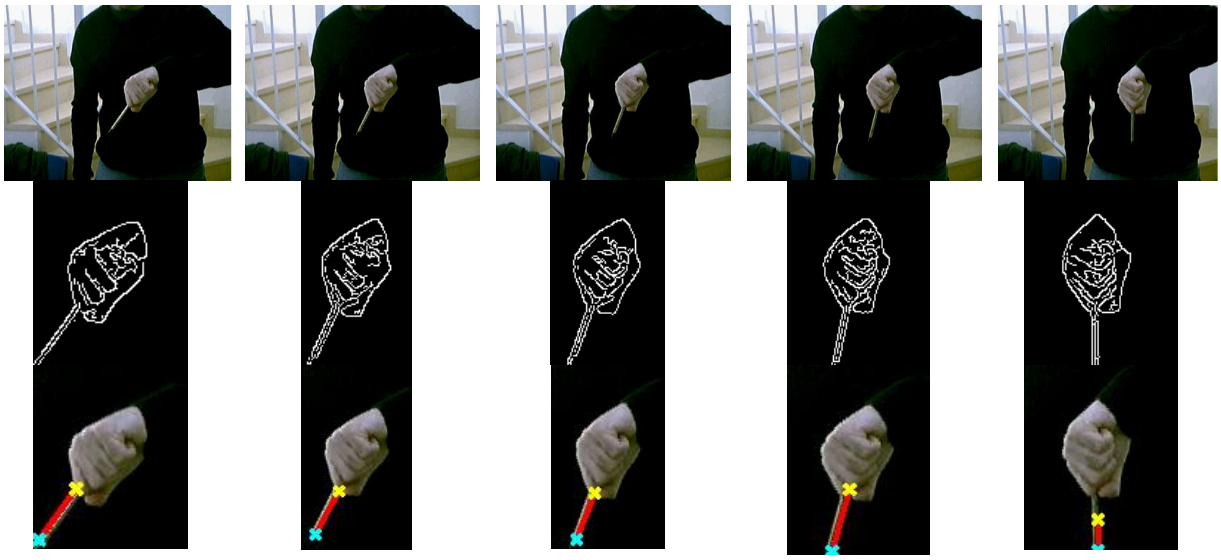


Figure 7. Instrument vertices detection and tracking

4.3 Sleep Mode

There are occasions when the doctor wishes to discuss the current image/or record with his/her colleagues or to attend to other matters. He/she might gesticulate while talking, or pointing to the image, which may trigger the recognition of unintentional gestures. To avoid this doctor may place the system “on-hold” or “sleep”. While the system is in sleep mode, the recognition processes are disabled. To switch to “sleep-mode” the user moves the hand to the lowest part of the screen, keeping it within the screen boundaries. To return to the “normal mode” the user waves the hand over the small rectangle in the lower left corner of the screen.

5 Conclusions

A vision-based system that can interpret user’s gestures in real-time to manipulate windows and objects within a medical data visualization environment is presented. A hand segmentation procedure using color-motion fusion extracts binary hand blobs from each frame of an acquired image sequence. Dynamic navigation gestures are translated to commands based on their relative positions on the screen. Static gesture poses are identified to execute non-directional commands, such as zoom and rotate. The gesture recognition system was implemented in a sterile medical data-browser environment (named Gibson).

Future work includes replacement of the rotation gesture to operate with the hand palm only, and the development of two handed gestures to achieve increased accuracy for the zoom and rotation gestures.

6 Acknowledgement

This work was partially supported by the Paul Ivanier Center for Robotics Research and Production Management, Ben-Gurion University of the Negev

References

[1] M. Schultz, J. Gill, S. Zubairi, R. Huber, F. Gordin, “Bacterial contamination of computer keyboards in a teaching hospital,” Infect Control Hosp. Epidemiol., vol. 4, no. 24, pp. 302-303, 2003.

[2] D. Nicholas, P. Huntington, P. Williams, P. Vickery, “Health information: an evaluation of the use of touch screen kiosks in two hospitals”. Health Information Librarian Journal. 2001 Dec, vol. 18, no. 4, pp. 213-9.

[3] K. R. Smith, K. J. Frank, R.D. Bucholz, “The NeuroStation- a highly accurate, minimally invasive solution to frameless stereotatic neurosurgery. Comput Med Imaging Graph, no. 18, pp. 247-256, 1994.

[4] H. Colle, K. Hiszem, “Standing at a kiosk: Effects of key size and spacing on touch screen numeric keypad performance and user preference,” Ergonomics, no. 18, pp. 1406-1423, 2004.

[5] A. Nishikawa, T. Hosoi, K. Koara, D. Negoro, A. Hikita, S. Asano, H. Kakutani, F. Miyazaki, M. Sekimoto, M. Yasui, Y. Miyake, S. Takiguchi, and M. Monden. "FAce MOUSe: A Novel Human-Machine Interface for Controlling the Position of a Laparoscope," IEEE Trans. on Robotics and Automation, vol. 19, no. 5, pp. 825-841, 2003.

[6] C. Graetzel, T.W. Fong, S. Grange, and C. Baur, "A non-contact mouse for surgeon-computer interaction," Technology and Health Care, vol. 12, no. 3, 2004, pp. 245-257.

[7] T J. Zeng, Y. Wang, M.T. Freedman and S.K. Mun, "Finger tracking for breast palpation quantification using color image features", SPIE Optical Engineering, vol. 36, no. 12, pp. 3455-3461, Dec. 1997.

[8] J. Wachs, H. Stern, Y. Edan, U. Kartoun, “Real-Time Hand Gesture Using the Fuzzy-C Means Algorithm”, In Proc. of WAC 2002, Florida, June 2002.

[9] Y. Kuno, T. Murashima, N. Shimada, and Y. Shirai, “Intelligent Wheelchair Remotely Controlled by Interactive Gestures,” In Proceedings of 15th International Conference on Pattern Recognition, vol.4, pp.672-675, 2000

[10] T. Starner, J. Auxier, D. Ashbrook, and M. Gandy. “The Gesture Pendant: A Self-illuminating, Wearable, Infrared Computer Vision System for Home Automation Control and Medical Monitoring”. In Fourth International Symposium on Wearable Computers, pp. 87-94, 2000.

[11] D. A. Becker and A. Pentland. “Staying Alive: A Virtual Reality Visualization Tool for Cancer Patients,” Proceedings of the AAAI'96 Workshop on Entertainment and Alife/AI, Portland, Oregon, August 1996

[12] M. Gutierrez, P. Lemoine, D. Thalmann, F. Vexo. “Telerehabilitation: Controlling Haptic Virtual Environments through Handheld Interfaces”.In Proceedings of ACM Symposium on Virtual Reality Software and Technology (VRST 2004), Hong Kong, November 2004

[13] R. Patel and D. Roy. “Teachable interfaces for individuals with dysarthric speech and severe physical disabilities”. In Proceedings of the AAAI Workshop on Integrating Artificial Intelligence and Assistive Technology, pp. 40-47, 1998.

[14] R. Boian, R. Sharma, C. Han, A. Merians, G Burdea, S. Adamovich, M. Recce, M. Tremaine, H. Poizner, “Virtual reality-based post-stroke hand rehabilitation,” Studies in Health and Technology Information, , no. 85, pp. 64-70, 2002.

[15] T. Baudel, and M. Beaudouin-Lafon. “CHARADE: Remote Control of Objects using Free-Hand Gestures,” Communications of the ACM. vol. 36, no. 7, pp. 28-35, 1993.

[16] H. Stern, J. Wachs, Y. Edan, “Optimal Hand Gesture Vocabulary Design Using Psycho-Physiological and Technical Factors,” 7th International Conference Automatic Face and Gesture Recognition, FG2006. Southampton, UK, April 10-12 2006

[17] G. R. Bradski, "Computer vision face tracking for use in a perceptual user interface," Intel Technical Journal, pp. 1-15, 1998.

[18] D. Comaniciu and P. Meer, “Robust Analysis of Feature Spaces: Color Image Segmentation,” CVPR’97, pp. 750-755.

[19] H. Stern, B. Efros, "Adaptive Color Space Switching for Tracking under Varying Illumination”, Journal of Image and Vision Computing, vol. 23, no. 3, 2005, pp. 353-364.

[20] National Institute for Medical informatics. Online Website: http://www.imedi.org/

[21] N Kiryati, Y Eldar, and AM Bruckstein. “A probabilistic Hough Transform,” Pattern Recognition, vol. 24, no. 4, pp. 303-316, 1991.

Appendix: Gibson Data Browser

The Gibson image browser is 3D visualization originally designed to help scientists explore multi-dimensional graphical representation of their data at various levels of detail. It has been adapted as a medical analysis tool that enables medical experts to easily and efficiently examine images, such as CT scans of the brain and X-ray radiographs. Rendering these images requires high processing power, which would not allow smooth rendering of the volume when rotating the collection or zooming into it, while using a standard processing-power machine. To avoid these performance problems, Gibson represents the surfaces (images) as hyper-squares manifolds, as seen in Figure A.

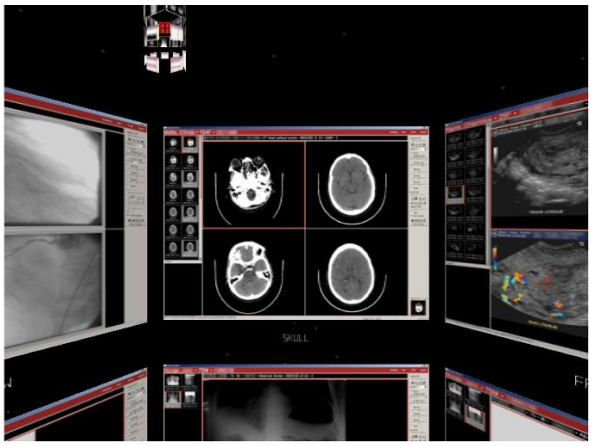


Figure A. Hyper-square representations of medical images