

Real-Time Hand Gesture Interface for Browsing Medical Images

Juan Wachs¹, Helman Stern^{1*}, Yael Edan¹, Michael Gillam², Craig Feied², Mark Smith², Jon Handler²

¹Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, Be'er-Sheva, Israel, 84105, {helman, yael,juan}@bgu.ac.il.

²Institute for Medical Informatics, Washington Hospital Center, 110 Irving Street, NW, Washington, DC, 20010, {feied,smith,handler}@medstar.net, michael.gillam@microsoft.com

Received 1 January 2007; revised 2 February 2007, accepted 3 March 2007

Abstract

A gesture interface is developed 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 objects in an image visualization environment. A color distribution model of the gamut of colors of the users hand or glove is built at the start of each session resulting in an independent system. The gesture system relies on real-time robust tracking of the user's hand based on a color-motion fusion model, in which the relative weight applied to the motion and color cues are adaptively determined according to the state of the system. Dynamic navigation gestures are translated to commands based on their relative positions on the screen. A state machine switches between other gestures such as zoom and rotate, as well as a sleep state. Performance evaluation included gesture recognition accuracy, task learning, and rotation accuracy. Fast task learning rates were found with convergence after ten trials. A beta test of a system prototype was conducted during a live brain biopsy operation, where neurosurgeons were able to browse through MRI images of the patient's brain using the sterile hand gesture interface. The surgeons indicated the system was easy to use and fast with high overall satisfaction.

Keywords

Hand gesture recognition, medical databases, browsing, sterile interface, image visualization

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 into the corresponding location in the image space of a high-performance computer monitor. Also, touch-screens are a popular means of interaction. In a setting like an operating room (OR), touch screen displays must be sealed to prevent the buildup of contaminants, and require smooth surfaces for easy cleaning with common cleaning solutions.

Many of these deficiencies may be overcome by introducing more natural human computer interaction (HCI) 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 these concepts were exploited in systems for improving medical procedures and systems. In FACE MOUSE [4], a surgeon can control the motion

*Helman Stern; Department of Industrial Engineering and Management; Ben-Gurion University of the Negev; Be'er-Sheva, Israel, 84105; helman@bgu.ac.il.

of the laparoscope by simply making the appropriate face gesture, without hand or foot switches or voice input. Graetzel, et al. [5] developed a computer vision system that enables surgeons to perform standard mouse functions (pointer movement and button presses) with hand gestures. Other systems [6] suggest 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 capable of recognizing the user's commands indicated by hand gestures [7]. The Gesture Pendant [8] 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 [9] 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 [10] 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 [11] 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 [12]. These patients receive therapy while being immersed in a virtual environment.

In this paper we explore only the use of hand gestures, which can in the future be further enhanced by other modalities. 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. We develop a vision based (unencumbered) hand gesture capture system that can be used to manipulate windows and objects, especially images, within a graphical user interface (GUI).

Performance of the gesture interface was tested in real-time in a hospital environment. In this domain the non-contact aspect of the gesture interface avoids the problem of possible transfer of contagious diseases through traditional keyboard/mouse user interfaces.

The system specifications and description appear in Section 2. In Section 3 image processing operations using adaptive color-motion fusion for hand segmentation and tracking is described. Section 4 provides details of the gesture operational mode, with state machine switching between directional navigation, rotate, zoom and sleep modes. Section 5 provides performance evaluation results. Section 6 reports on the test of the system during a live brain biopsy operation. Final conclusions are provided in Section 7.

2. SYSTEM OVERVIEW

System Description

A web-camera placed above a screen 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 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 a return to its original position. The direction of the 'flick' is used to navigate through a visual image data browser. Other actions/commands such as: zoom, rotate and system sleep 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.

System Specifications

Some structural characteristics of a gesture interaction model for a medical environment are presented in [13], and extended for the OR domain in [5]. For the correct design of a hand gesture interaction system for doctors/surgeons, the following specifications should be considered [14]: (1) Real time feedback and operation - during surgery the system should be fast and enable the surgeon to obtain visual feedback of the evoked gestures, (2) Fatigue - gestural commands must be concise and rapid to minimize effort, (3) Intuitiveness - gestures should be cognitively related to the commands or actions they represent, (4) Unintentionally - most systems capture every motion of the user's hand, and as a consequence an 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 backgrounds containing object motion and variable illumination, (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.

3. HAND SEGMENTATION AND TRACKING

Color Cue

An initial hand or glove color 2D histogram is generated in real-time during the ‘calibration’ stage of the system. An interface preview window shows an outline of a hand palm drawn on the screen (Figure 1). The calibration process is initiated when the user places his hand slowly into the region of the template. The user then positions the hand image within the screen template which is scanned to construct a hand color distribution model. The calibration allows individuals with differing skin colors, as well as different colored gloves, to operate the system resulting in an independent user gesture system. At each frame k during operation of the system, using a 2D histogram lookup table, the color of a pixel at location (x, y) is converted to the probability that the pixel is classified as a hand (or gloved hand). It is assumed that the operator is wearing a long sleeved-shirt to avoid confusion due to the position of the arm. An exception is when single colored gloves are worn with short sleeves, as is often the case for doctor/surgeons in an operating room (Figure 1).



Figure 1. User hand skin color calibration

Background Maintenance

It is possible that the calibration phase can be initiated through the detection of motion in the background which results in an erroneous sample of

the hand or glove colors. In order to avoid false motion clues originated by non-hand motion, a background maintenance operation is maintained. An initial image of the background is stored immediately after the application is launched. 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 image contains only a background, otherwise, an upper threshold level t_2 is checked to test whether the present object is a hand. Between t_1 and t_2 the background is not updated as it is assumed that something other than small pixel noise or a fast placement of the hand into the template is present. The background stored image is updated using a smoothed average of each pixel (i, j) according to (1) below.

$$B_k(i, j) = \lambda \times f(i, j) + (1 - \lambda) \times B_{k-1}(i, j) \quad (1)$$

Here, B_k is the updated stored background image at frame k , B_{k-1} is the stored background image at frame $k-1$, λ is the smoothing coefficient (regulates update speed), and $f(i, j)$ is the current background image at frame k .

Motion Cue

To detect the motion of the hand, a grayscale image is obtained from the RGB image, and smoothed using a Gaussian filter to obtain $I_k(i, j)$, the grayscale value of pixel (i, j) at frame k . Denote $\Delta_k(i, j) = |I_k(i, j) - I_{k-1}(i, j)|$ as the absolute value of the frame difference between frames k and $k-1$ at pixel (i, j) . Define a binary motion indicator $\phi_k(i, j) = 1$, if $\Delta_k(i, j) \geq \epsilon$ (where ϵ is a small value), and 0 otherwise. Morphological image processing operations are then employed to clean holes and small noise in the image.

Color and Motion Fusion

As a result of the color cue we have an intensity image $p_k(i, j)$ representing the hand skin or glove color probability at frame k , and a second black and white image determined by a motion indicator ϕ_k . At each frame k , a fused intensity image I_k is computed according to (2) based on the work found in Stern and Efron [15].

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

For the scene in frame k , the total normalized motion is computed and represented as a scalar $\alpha_k \in [0, 1]$. This provides the relative weights between the motion and color cues in (2). This motion weight is adaptive, increasing or decreasing in response to the total amount of motion in the scene. In the limit at $\alpha_k = 0$ only the color cue is used. The value d is an amplifying factor ($d = 1.3$ was found

empirically to provide the best performance). The min function ensures that the probability does not exceed a value of one. Motion indication reinforcement is introduced to overcome the weak ability of color only to 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 motion but also that of the entire body.

Hand Tracking

The CAMSHIFT algorithm is used to track the hand. CAMSHIFT, as described by Bradski [16], uses a probability distribution image comprised of pixels representing hand colors. This hand image is created from a 2D hue-saturation skin color histogram [17]. A histogram is used as a look-up-table to convert the acquired camera images of the hand into corresponding pixel intensities, a process known as back projection. In the original CAMSHIFT algorithm the probability of a pixel belonging to the hand is determined by the grayscale value of the pixel only. In lieu of using color probability alone, we modify it with motion information according to (2) to represent a hand pixel probability. The relative weights between color and motion are shifted according to the amount of motion in the scene resulting in an adaptive fusion system. Using the centroid and size of the hand blob of pixels an iterative procedure based on a generalization of the mean shift algorithm [18] is used to update the tracking window at each frame. Thresholding to black and white followed by morphological operations is used to obtain a single component for further processing to classify the gestures.

4. OPERATION MODES AND GESTURE RECOGNITION

The gesture interface was implemented in a medical database named Gibson developed by IMI [19] for the purpose of interacting with medical images such as X-rays and MRIs. A finite state machine (Figure 2) is used to define the operational architecture of the 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 users image is a rectangular frame (Figure 3). The area inside the frame is called the "neutral area".

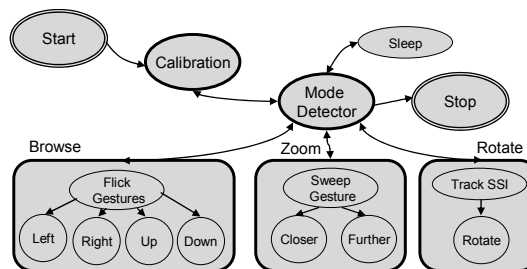


Figure 2. State machine for the gesture-based medical browser

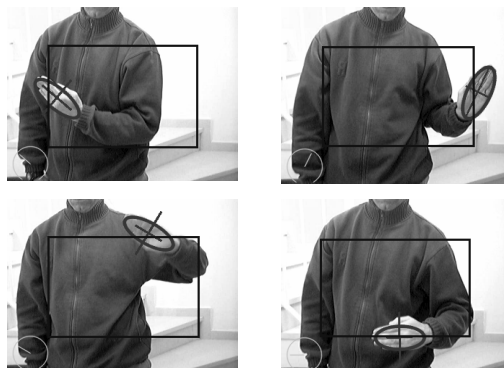


Figure 3. Four quadrants mapped to cursor movements

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 a posture are sent to a "mode detector" to determine whether a zoom or rotation gesture is presented. The hand must always be in the neutral area to conduct a mode change gesture.

Directional Navigation

Navigation gestures were designed to browse through a Gibson data browser. Gibson represents medical imaging information as 3D objects where each image is a flat rectangular plate formed around a cylinder and arranged in numerous levels. The cylinder can be rotated CW and CCW, and moved up and down in the vertical direction to exhibit various levels on the screen. Because of the wrap around feature both in vertical and horizontal directions, one can consider the images to be mounted on a torus. 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. 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. For clarification, if the user wishes to look at the image to the left of the image appearing in the center of the screen, the hand moves initially to the left across the boundary of the rectangle, and “flicks” back into the neutral area again (see Figure. 3). This moves the torus to the right as viewed by the user, and is similar to how one flicks through the pages of a book.

Zoom Mode

The main purpose of the zoom is to change the size of an image. Once the zoom-mode is activated the size of image is adjusted in proportion to the area of the hand as its distance to the screen is changed. 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 90° to 180°. To test if the hand is rotated 90° 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.
2. Four reference frames are marked. The first frame is the current frame at $f_4=0$ frames, $f_3=-w/3$ 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: $f_4= 0$, $f_3=-13$, $f_2= -26$, and $f_1=-39$.

Let, ψ - The difference of the angles of two consecutive frames; $\bar{\psi}_{(i,j)}$ - The average angle between frames f_i and f_j ; $std(\psi_{(i,j)})$ -The standard deviation of the angles between frames f_i and f_j ; ϵ - The error margin allowed in degrees (5° was used).

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

$$90 - 2 * \epsilon \leq \bar{\psi}_{(3,4)} - \bar{\psi}_{(1,2)} \leq 90 + 2 * \epsilon \quad (3)$$

$$std(\bar{\psi}_{(3,4)}) \leq 1.5 * \epsilon \quad \vee \quad std(\bar{\psi}_{(1,2)}) \leq 1.5 * \epsilon \quad (4)$$

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. For most situations rotation will be in increments of 90°. The instrument can be found in an image by applying the Probabilistic Hough Transform, referred to as pHT [20]. 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, the wrist, etc. 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.

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 1°, 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 end points 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 the line segment. To select among all line segments, that segment 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 line length is determined and the following tests are made:

- 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) Trace a line from the centroid of the tracking window to the far endpoint of the instrument.

Find the shortest distance from the closest endpoint of the instrument to the traced line. Is this distance small?

Every positive response adds one vote to the total votes of a candidate line segment. The line with the highest number of votes is selected as the line segment representing the instrument. An example of successful tracking of an instrument is shown in Figure 4.

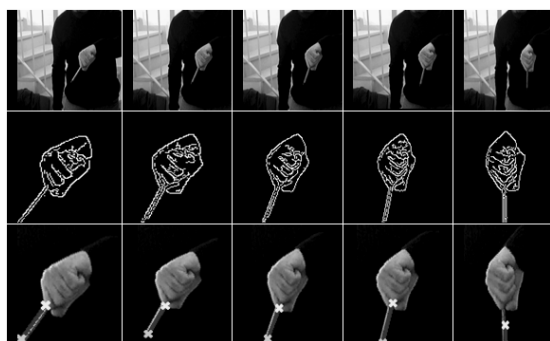


Figure 4. Detection of instrument vertices

Sleep Mode

There are occasions when the doctor wishes to discuss the current image with his/her colleagues or to attend other matters. He/she might gesticulate while talking, or pointing to the image, which may trigger the recognition of an unintentional gesture. This is known as the "Immersion Syndrome" [13]. To avoid these situations the system may be placed "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" a wake up gesture is used whereby the user waves the hand over the small rectangle in the upper left corner of the screen.

5. PERFORMANCE EVALUATION

Description of the experiments

Four types of performance evaluations were conducted; (a) gesture recognition accuracy, (b) task learning, (c) excess gestures used, and (d) rotation accuracy. In addition, the system was tested by surgeons during a neural operation in a real hospital setting. Gesture recognition accuracy, task completion time, and number of excess gestures used were measured for 10 non experienced users in the experiments defined below. The exception was for the rotation gesture accuracy which was tested by an experienced user. Users were selected from the 24-37 age range with medium to high levels of dexterity and medium to high experience in computer use. The subjects were

queried on the ergonomic aspects such as comfort and intuitiveness after completing the experiments.

Gesture Accuracy

At the start of the experiment for each subject, a tutor demonstrates each of the following eight gestures: left, right, up, down, zoom-in, zoom-out, sleep, and wake-up. To ensure that each subject understands how to perform the gestures, each subject is allowed to practice each gesture four times. The user then performs each gesture an additional four times and a log file is created showing the gestures recognized by the system. This information is used to obtain a confusion matrix from which the recognition accuracy of each gesture and the overall recognition accuracy are obtained (Table 1).

Table 1. Recognition accuracy of each gesture and overall accuracy.

Total	LEFT	RIGHT	UP	DOWN	ZOOM IN	ZOOM OUT	SLEEP	WAKE	Acc
LEFT	38	0	0	2	0	0	0	0	0.95
RIGHT	0	40	0	0	0	0	0	0	1.00
UP	0	0	37	3	0	0	0	0	0.93
DOWN	1	0	0	39	0	0	0	0	0.98
ZOOM IN	0	0	0	4	35	1	0	0	0.88
ZOOM OUT	0	0	0	2	0	38	0	0	0.95
SLEEP	0	0	0	0	0	0	40	0	1.00
WAKE	0	0	0	0	0	0	0	40	1.00
Acc									0.96

Overall recognition accuracy of the eight gestures is 96 percent. Individual gesture recognition accuracies are in the range of 93 to 100 percent, except for the zoom in gesture. As the user moves the hand closer to the screen, the tendency is for the palm to tilt down. This may cause the hand's centroid to move below the neutral rectangle, causing a switch between the zoom mode and the browse mode, where a down gesture is subsequently recognized. Thus, the lower recognition accuracy of the zoom gesture is not due to the recognition algorithm, but due to failure of the subject to hold the palm parallel to the screen throughout the performance of the gesture. Alerting the subject to the proper execution of this gesture should improve its recognition accuracy.

Browsing Task

Here subjects view a monitor screen displaying a set of medical image from the Gibson browser. Each subject is requested to start at image A, and arrive at image B (Figure 5) by using navigation gestures only. Once the image was found, the subject is asked to return to the original image A. To complete the task, as fast as possible, 12 browsing gestures are necessary; six to reach image B, and another six to return to image A. Ten trials are repeated with the completion times and gesture sequences recorded in the log-file.

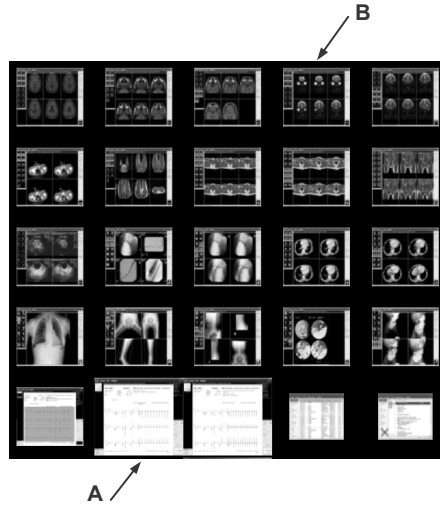
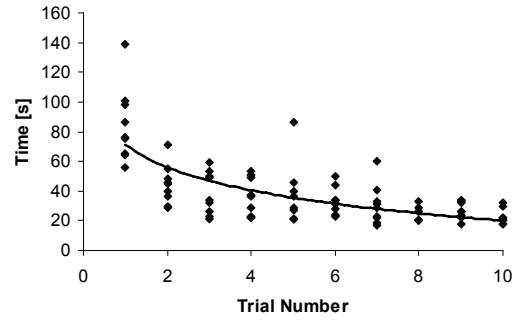


Figure 5. Medical image search task

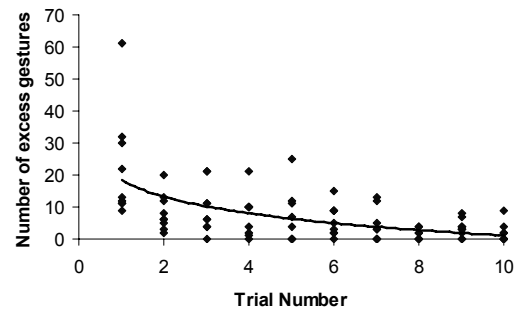
Only data for 9 of the 10 subjects were used. The results for one of the subjects was deemed unacceptable as he exhibited a low level of dexterity, and did not follow the guidelines (sitting instead of standing, switching hands, and taking a break in the middle of the experiment). Also one outlier at trial 8 was removed for one of the subjects. The bad performance was traced to the fact that the pictures "wrap" around when reaching any edge (instead of getting stuck). The subject felt very experienced, and was gesticulating very fast. Because she felt confident, she was not waiting enough time to check whether she was close to the edge of the cylinder of images. This caused a wrap around, and therefore she had to browse through a lot of images to return to the same place that she wanted. This feature is a fault of the design of the image browser, and not that of the recognition capability of the system. As such, when this fault is corrected the performance of the system should improve.

As the user gains experience, the task completion time is expected to decrease. A learning curve was fit through the data of ten trials to establish how fast the browsing task could be learned. The data points and learning curve appear in Figure 6(a).

The model for the learning curve is $Y_n = Y_1 n^{-b}$, where n is the trial number, Y_n is the estimated value of the completion time in seconds on the n^{th} trial, Y_1 is the time of the first trial, b is $\log r / \log 2$, and r is the learning rate. The best fit learning curve equation for task time is $Y_n = 69.4 n^{-0.49}$. This results in a learning rate of $r = .72$, which means that for every doubling of trials the completion time



(a)



(b)

Figure 6. Learning curves for (a) task completion time, (b) excess gestures (9 subjects, 10 trials each)

decreased by 28%. As can be seen from the learning curve, learning leveled off at the 10th trial with a value of 22.5 seconds.

It was also decided to examine the number of excess gestures used beyond the minimum of 12 required to complete the task. This was determined by examining the gesture sequence for each trial. The number of excess gestures in each trial for each user is plotted in Figure 6(b), and used to obtain the learning curve $Y_n = 18.4 n^{-0.90}$, which represents a learning rate of $r = .54$. This implies very fast learning (reducing the number of extraneous gestures used by 46% for every doubling of the number of trials. At the end of 10 trials an estimate of .88 excess gestures were used.

Rotation Accuracy

Three tasks were designed to test the accuracy of the rotation gesture. As this gesture will mostly be used for rotation of an image CW or CCW by 90° or 180°, the following tasks were defined: (a) Rotate from 90° towards 180°, (b) Rotate from 180° towards 90°, and (c) Rotate from 90° to 0°. Each task was repeated three times using slightly different background clothing. For each trial upon

arrival at the target angle, and holding the pose, a sample of 5 frames was taken and the angle output of the rotation detector was recorded. The result of this test was an overall mean absolute error of 3°. This accuracy is reasonable considering the difficulty to obtain fine positional accuracy of the hand especially because of its jitter in free space. As this gesture will mostly be used for a 90° or 180° rotation of an image in a CW or CCW direction, any average within ±5° can be used as a sufficient signal to rotate and lock the image into the intended position.

Ergonomic Aspects

To test the ergonomic aspects of the gesture interface system, every user was asked two questions: Q1 – What was the strength of the comfort level?, and Q2 – How intuitive were the gestures? Table 2 summarizes the answers. Numbers are based on Borg Scale [21] with verbal anchors from weak to strong association according to: 0-Not at all, 1- Very Weak, 2-Weak, 3-5 Moderate, 5-6 Strong, 7-9 Very Strong, and 10-Extremely Strong.

Table 2. Comfort (Q1), and intuitive (Q2) measures of the gesture interface system

Subject	1	2	3	4	6	7	8	9	10	mean
Q1	4	7	6	7	4	10	5	4	5	5.778
Q2	8	8	7	7	7	10	8	8	8	7.889

6. REAL - TIME OPERATIONAL TEST

A beta test of the system was conducted in July, 2006 in an OR at Washington Hospital Center, Washington, DC, during a live biopsy operation. Neurosurgeons were able to browse through MRI images of the patient's brain using the Gestix hand gesture interface with the Gibson visual data browser (Figure 7). Surgeons that used the system had no prior experience with gesture interfaces. The setup time for the whole Gestix system was approximately 20 minutes. The system installed in the OR consisted of a commercially available Canon VC-C4 color camera. This camera was placed just over a large flat screen monitor. Additionally, an Intel Pentium IV, (2,4 GHz, OS: Windows XP) with a Matrox Standard II video-capturing device allowed a frame-rate of over 60 frame/sec.

From the two images at the top of Figure 7 one can see during the OR session the system adapting to changes in illumination (when the overhead lights

are turned off). The complete video can be seen at the Institute of Medical Informatics' (IMI) web site [19]. The third image from the top shows the OR setting in which the gesture interface was installed. The fourth image is a view of the interface screen showing a brain MRI image and a view of the user's gesture feedback.

Some comments made by the neurosurgeons subsequent to the operation were: "It was very easy to use the system, after all we are good with our hands", and "We were very satisfied with its speed and ease of use". At the end of the entire operation procedure, the main surgeon who conducted the task of browsing MRI images, filled in a questionnaire with questions on task experience, ease of task, time of task completion, and overall task satisfaction. The surgeon's response indicated that the Gestix system was easy to use, fast, with high over all satisfaction.

7. 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. The system is user independent due to the fact that the gamut of colors of the users hand or glove is built at the start of each session. Hand segmentation and tracking uses a new adaptive color-motion fusion function. Dynamic navigation gestures along with zoom, rotate, and system sleep gestures are recognized.

Three types of system performance evaluations were conducted; (a) gesture recognition accuracy, (b) task learning, and (c) rotation accuracy. Subjects were also queried on the ergonomic aspects of the system such as comfort and intuitiveness. In addition, the system was tested by surgeons during a neural operation in a real hospital setting. A test of the system was conducted in an OR at Washington Hospital Center, Washington, DC, during a live biopsy operation where neurosurgeons browsed through MRI images of the patient's brain using the Gestix hand gesture interface. Surgeons were given a post operation satisfaction questionnaire which revealed high scores for ease of use, task completion time and overall satisfaction.

The following are some of the major advantages of the hand gesture interface for use by surgeons and doctors: (i) Easy to use: - the system allows the use of hands, which is the natural work tool for the surgeons, (ii) Rapid reaction: - nonverbal instructions by hand gesture commands are intuitive and fast. In practice, the Gestix system can

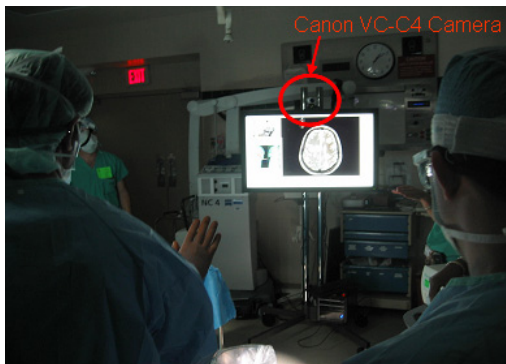


Figure 7. Top pair - surgeons operating gesture interface under bright and dark illumination in the operating room. Bottom pair – Views of gesture interface system installation in operation room

process images and track hands in real-time, (iii) Unencumbered: - the proposed system does not require the surgeon to attach a microphone, use head-mounted (body-contact) sensing devices or use foot pedals, (iv) Sterile interface: - a sterile non contact interface is a major advantage for use in operating rooms, and (v) Controlled from afar: - the hand gestures can be performed up to 5 meters from the camera and still be recognized accurately.

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. It is also planned to assess the use of stereo and/or infrared cameras.

ACKNOWLEDGEMENTS

This work was supported by the Institute for Medical Informatics, Washington Hospital Center, Washington, D.C. Partial support was given by the Paul Ivanier Center for Robotics Research and Production Management, and by the Rabbi W. Gunther Plaut Chair in Manufacturing Engineering, Ben-Gurion University of the Negev.

REFERENCES

- [1] M. Schultz, J. Gill, S. Zubairi, R. Huber, and 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, and P. Vickery, "Health information: an evaluation of the use of touch screen kiosks in two hospitals," *Health Information Librarian Journal*, vol. 18, no. 4, pp. 213-9, 2001.
- [3] K. R. Smith, K. J. Frank, and 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] 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.
- [5] 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, pp. 245-257, 2004.
- [6] J. Wachs, H. Stern, Y. Edan, and U. Kartoun, "Real-time hand gestures using the fuzzy-C-means algorithm," in Proc. of WAC 2002, Florida, 2002.
- [7] 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.
- [8] 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.
- [9] 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, 1996.
- [10] M. Gutierrez, P. Lemoine, D. Thalmann, and 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, 2004.
- [11] 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.
- [12] R. Boian, R. Sharma, C. Han, A. Merians, G. Burdea, S. Adamovich, M. Recce, M. Tremaine, and H. Poizner, "Virtual reality-based post-stroke hand rehabilitation," Studies in Health and Technology Information, no. 85, pp. 64-70, 2002.
- [13] T. Bade 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.
- [14] H. Stern, J. Wachs, and Y. Edan, "Optimal hand gesture vocabulary design using psycho-physiological and technical factors," in Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, FG2006, Southampton, UK, pp. 257-262, 2006.
- [15] H. Stern and B. Efron, "Adaptive color space switching for tracking under varying illumination", Journal of Image and Vision Computing, vol. 23, no. 3, pp. 353-364, 2005.
- [16] G. R. Bradski, "Computer vision face tracking for use in a perceptual user interface," Intel Technical Journal, pp. 1-15, 1998.
- [17] D. Comaniciu and P. Meer, "Robust analysis of feature spaces: color image segmentation," CVPR'97, pp. 750-755, 1997.
- [18] Y. Cheng, "Mean shift mode seeking, and clustering," IEEE Transactions on pattern analysis and machine Intelligence, Vol. 17, pp. 790-799, 1995.
- [19] National Institute for Medical informatics. Online Website: <http://www.imedi.org/>
- [20] Kiryati, Y Eldar, and AM Bruckstein. "A probabilistic Hough Transform," Pattern Recognition, vol. 24, no. 4, pp. 303-316, 1991.
- [21] G. Borg, "Psycho-physical bases of perceived exertion," Medicine and Science in Sports, vol. 14, no 5, pp. 377-381, 1982.

AUTHOR INFORMATION



Juan Wachs received his B.Ed.Tech degree in Electrical Education from the ORT Academic College in Jerusalem, Israel, in 1995. He received M.Sc. in Information Systems, and is currently working toward the PhD degree in intelligent systems at Ben-Gurion University. Since 2000, he has become a critical member of a team of researchers working on Virtual Reality Telerobotic Operations through the use of a hand gesture-computer interface at the Ben-Gurion University of the Negev. His current research interests include machine-vision, telerobotics and virtual reality. Mr. Wachs is a Member of the Operation Research Society of Israel.



Helman Stern received a Ph.D. in Operations Research, University of California, Berkeley, a Masters in Engineering Administration, George Washington University, and a BS in Electrical Engineering, Drexel University. He has taught at the University of

California, at Berkeley, Rensselaer Polytechnic Institute, SUNY at Albany and San Francisco State University. He is a Professor at Ben Gurion University of the Negev where he founded the Telerobotics and Intelligent Systems Labs, and teaches Intelligent Systems, and Machine Vision. His research interests are in human robotic cooperative learning, optimal gesture vocabulary design, color face and multiobject tracking, and intelligent multi-modal man-machine interfaces.



Yael Edan is a Professor in the Dept. of Industrial Engineering and Management, Ben Gurion University of the Negev. She holds a B.Sc. in Computer Engineering and M.Sc. in Agricultural Engineering, Technion-Israel Institute of

Technology, and Ph.D in Engineering, Purdue University. Her main research includes robotic and

sensor performance analysis; robotic gripper analysis and design; systems engineering of robotic systems; robotic control of dynamic tasks; sensor selection procedures; sensor fusion; multi-robot control methodologies; telerobotics control; and human-robot collaboration methods. She has made major contributions in the introduction of intelligent automation and robotic systems to the field of agriculture.



Michael Gillam, MD, is Director of Research and Partnerships, Microsoft Azyxxi Health Solutions Group. He is a board certified emergency physician and member of the Department for Emergency Medicine, Evanston Northwestern Healthcare,

Northwestern University Medical School. He has been active in medical informatics for over 18 years. He lectures nationally and has served as Chair of Medical Informatics Society for Academic Emergency Medicine and American College of Emergency Physicians. He is also the Director of the Medical Media Lab at the National Institute for Medical Informatics in Washington D.C. where he manages projects ranging from advanced data visualization, augmented and virtual reality, to medical robotics.