Gestix: A Doctor-Computer Sterile Gesture Interface for Dynamic Environments

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

[1]Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, Be'er-Sheva, Israel, 84105,
{juan, helman, yael}@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 design a sterile gesture interface for users, such as doctors/surgeons, to browse medical images in a dynamic medical environment. A vision-based gesture capture system interprets user's gestures in real-time to navigate through and manipulate 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 tracking of the user's hand based on color-motion cues. A state machine switches from navigation gestures to others such as zoom and rotate. A prototype of the gesture interface was tested in a operating room by nuerosurgeons conducting a live operation. Surgeons feedback was very positive.

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 (HCI). Unfortunately, it has been found that a common method of spreading infection 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 face gestures. Hand gestures for mouse functions for doctor-computer interfaces appeared in Graetzel et al. [6]. Zeng et al. [7] use finger position to obtain data on breast palpations. Other systems [8] suggest a teleoperated robotic arm using hand gestures for multipurpose tasks. Wheelchairs are guided by hand gestures in [9]. In [10] a Gesture Pendant is used to control home devices. 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).

In this paper we explore only the use of hand gestures. In particular we propose a doctor-computer interface system based on the recognition of 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 this 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, as well as, description of the Gibson 3D data browser used as our domain of application.. 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. Example results including implementation in a neurosurgeon operating environment is described in section 5. Final conclusions are provided in section 6.

2 System Overview

2.1 System Specifications

Some structural characteristics of a gesture interaction model for a medical environment are presented in [11], 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: (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. A method to include Psycho-Physiological factors in the design of a hand gesture interface is given in [12].

2.2 Architecture and Methodology

A web-camera placed above a screen (Fig. 1) 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. A two layer architecture is shown in Fig 2. The lower level, Gestix, provides tracking and recognition functions, while the higher level, Gibson, manages the user interface.

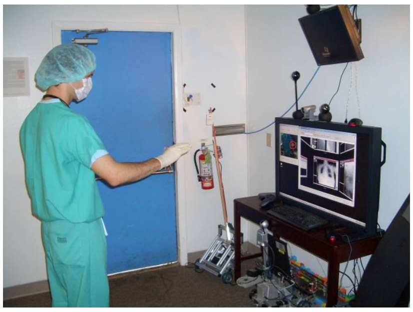


Fig. 1 Gesture capture

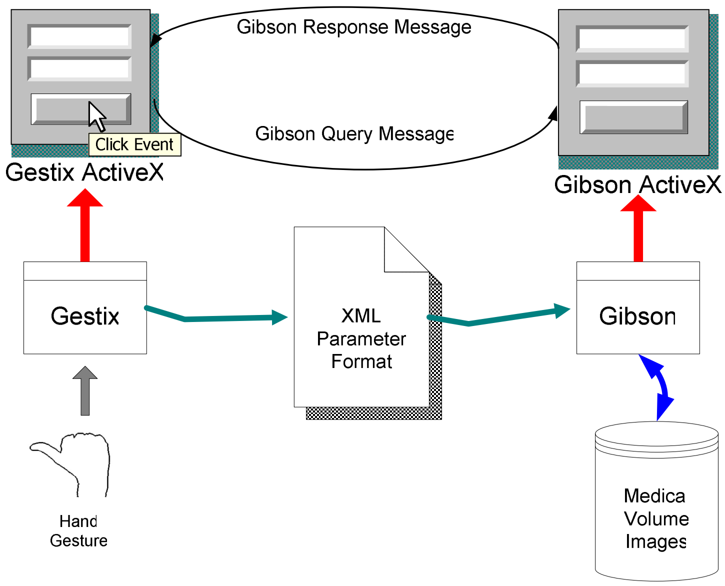


Fig. 2. Architecture of Gestix-Gibson Interface

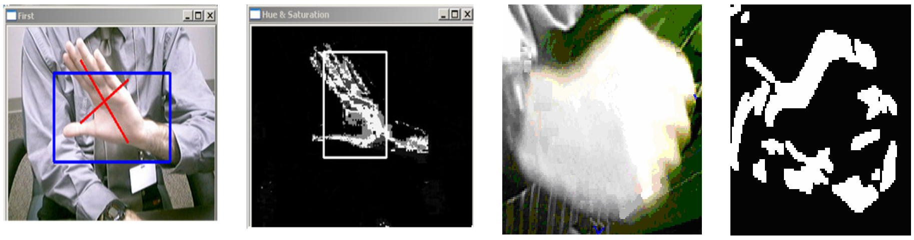
2.3 The Gibson Data Browser

The Gibson image browser is 3D visualization medical tool that enables examination of images, such as CT scans and X-rays. To interface the gesture recognition routines with the Gibson system, information such as, the centroid of the hand, its size representing zoom, and orientation for rotation angle, is used to enable screen operations in the Gibson GUI. Both the gesture interface and the Gibson image browser are embedded in ActiveX controls which are communicated using messages and windows events. The messages between Gestix and Gibson are based on customized windows events (GibsonQuery and the GibsonResponse) and are intended to browse and manipulate images in the database. The current status of the hand detection module (Gestix) is sent to the Gibson control in XML format as metadata.

3 Segmentation

The CAMSHIFT [13] 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 [14]. 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 projection, see Fig 3(a-b). The initial 2D histogram is generated in real-time by the user in the 'calibration' stage of the system, for more details see [15].

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 to a BW motion image Morphological operations are used clean the image, removing holes and small noise. (see Fig 3(c-d))



(a) (b) (c) (d)

Fig. 3. .(a) User hand image. (b) Back-projected image (color cue). (c) Motion gesture. (d) Image differencing (motion cue)

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 [16] according to (1).

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

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 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. 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

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.

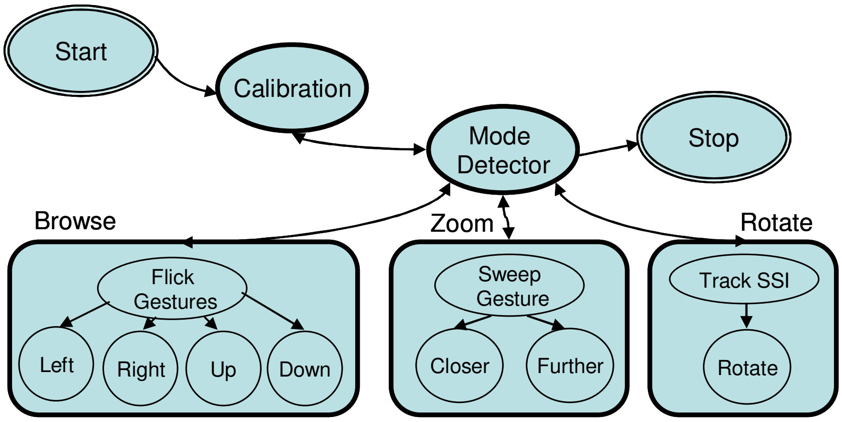


Fig. 4. State machine for the gesture-based medical browser

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 quadrants (left, right, up, down), is the image cylinder moved in the direction of the flick.

4.2 Zoom Mode

Once the zoom-mode is activated the size of image is changed 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 counter clock wise, from 90º to 180º.

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 (SSI) in the fist of the hand, and holds it in the range of .25 m to 2 m 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 between the SSI output and the horizontal axis. The architecture to detect the SSI segment in an image is shown in Fig. 5.

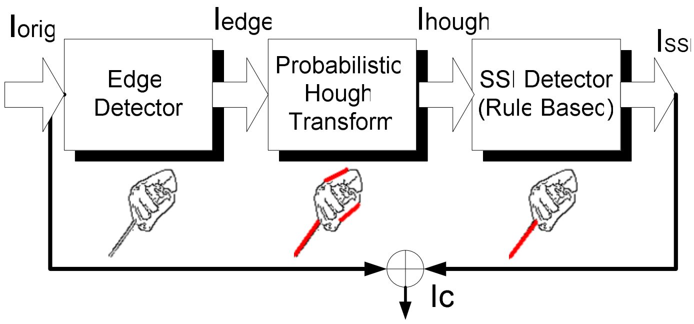
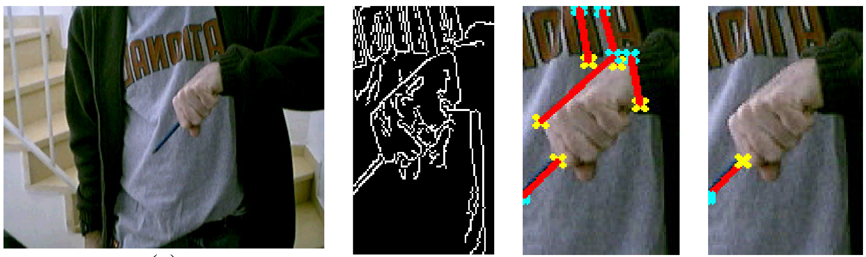


Fig. 5. Architecture of the SSI Algorithm

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, I_{org}. This image is input to a canny edge detector using two thresholds t_1=50 and t_2=210, and a 3x3 mask, to obtain an edge image, I_{edge}. This is followed by the probabilistic Hough Transform, referred to as pHT [17], with distance resolution=1, angle resolution of 1°, threshold of 30, minimum line length of 30, and the maximum allowable gap between collinear line segments of 2. The pHT algorithm finds the set of straight line segments in the image longer than some minimum.. The entire set of line segments is represented in the image $I_{hough..}$ These segments are good candidates to be the sides of the SSI; however, they may also represent other straight lines such as: doors, windows, tables, wrist, etc. The best segment to represent the SSI is selected by a SSI detector which uses prior knowledge to design a set of crisp rules which is applied to the set of line segments. Using the end points of each line i segment returned from pHT. The closest end point of a line segment i to the hand is determined by finding the end point closest to the centroid of the hand tracking window. This end point is designated as the start end point $S_f(i)$, and the other as the end point, $E_f(i)$. With both endpoints identified it is possible to find the rotation angle of each line segment $\Phi_f(i)$, as measured CCW from the horizontal of the current frame f. To select among all line segments that one most likely to represent the SSI, a rule based evidence test is conducted. This rule based algorithm is described below.

SSI Detector Algorithm (Rule Based)

At frame f, for each candidate line segment i, represented by (S_f (i), E_f(i), Φ_f(i)) found in the image the length of all line segments and perform the following tests: (1) Is i the longest of the lines?, (2) Is the difference between the angle of the major axis of the tracking ellipse of the hand and $\Phi_f(i)$ is small (within a given ε value), (3) Is the change of the angle of the i^{th} line segment with respect to its angular position in the preceding frame small, i.e.; Is $(\Phi_f(i)-\Phi_{f-1}(i))<\varepsilon$)?, (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? The test is comprised of a set of queries for each line segment. If enough responses are positive, then enough evidence is accumulated to infer that the line segment represents the SSI. The test looks for a line segment that is long, aligned with the major axis of the detected hand tracking ellipse and has not changed significantly from the last SSI detected, and is pointing to the center of the hand. This last test (test 4) is designed to eliminate all line segments except those radiating out of the center of the hand. Of the set of segments passing test 4 only those aligned with the major axis of the hand will be selected by 3. 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 SSI. Considering I_{SSI} as the image with the single detected line segment, it is overlaid on I_{org} obtain the composite image I_c. An example of I_c is shown in Fig. 6(d). An alternative method for detecting the SSI is to use a Fuzzy Inference System by replacing the above test questions by fuzzy rules. For example, for the question "is i the longest of the lines?"; one may define a linguistic variable line length and membership functions for the terms; short, medium and long, and a linguistic variable, degree of line i being a SSI with terms; not likely, likely, very likely, most likely. Then the following fuzzy rule may be defined: If **<line i>** is< **long>** then the **<degree of line i being a sterile instrument>** is< **very likely>**. A composite premise may be defined to generate a SSI fuzzy acceptance test rule. This method, however, is left for future work.



(a) (b) (c) (d)

Fig. 6. The SSI algorithm applied to a single frame. (a) Original image. (b) After canny edge detector. (c) Output of the pHT. (d) Output of the SSI detector

5 System Implementation

An example of a rotation gesture with a complex background is shown in Fig, 6(a). The pHT output, Fig. 6.(c) consists of five candidate line segments. Three of them violate rules 3 and 4, because they are short, their angles are significantly different from the previously tracked line segment, and the angles are far different from the camshift tracking window. As for the remaining two candidates, those with a positive slopes, only the bottom segment is directed toward the centroid of the hand and is this selected. An example of detection and tracking is shown in Fig. 7, where SSIs were found for 80% of the frames of a 1min, 13 sec video. The system was also tested in a hospital environment during a live neurosurgeon operation where surgeons browsed MRI images using the gesture system.

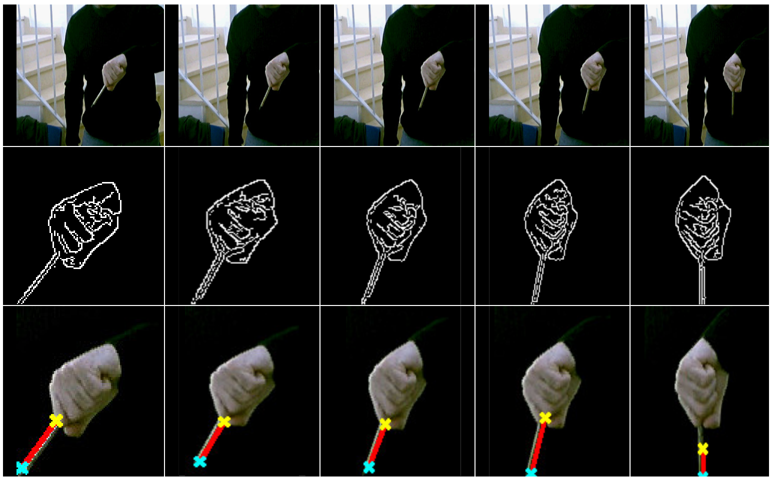


Fig. 7. Sequence of successful detection and tracking of an SSI

6 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). The system was tested in a hospital environment during a live neurosurgeon operation where surgeons browsed MRI images using the gesture system. 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. In addition, the development of a fuzzy inference system will be conducted for detecting the hand held sterile instrument..

7 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 Proc. of 15th Intl Conf. 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 Intl Symp. on Wearable Computers, pp. 87-94, 2000.
11. 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.
12. 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
13. G. R. Bradski, "Computer vision face tracking for use in a perceptual user interface," Intel Technical Journal, pp. 1-15, 1998.
14. D. Comaniciu and P. Meer, “Robust Analysis of Feature Spaces: Color Image Segmentation,” CVPR’97, pp. 750-755.
15. J. Wachs, H. Stern, Y. Edan, M. Gillam, C. Feied, M. Smith, J. Handler, “A Real-Time Hand Gesture Interface for Medical Visualization Applications,”, in the 10th Online World Conference on Soft Computing in Industrial Applications, September 19th - October 7th, 2005.
16. 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.
17. N Kiryati, Y Eldar, and AM Bruckstein. “A probabilistic Hough Transform,” Pattern Recognition, vol. 24, no. 4, pp. 303-316, 1991.