Touch-less Interaction with Medical Images Using Hand & Foot Gestures

Shahram Jalaliniya
IT University of Copenhagen
2300 København S, Denmark
jsha@itu.dk

Lars Büthe
ETH Zürich
Gloriastrasse 35, 8092 Zürich
lars.buethe@ife.ee.ethz.ch

Jeremiah Smith
Imperial College of London
London, SW7 2RH, UK
jeremiah.smith@imperial.ac.uk

Thomas Pederson
IT University of Copenhagen
2300 København S, Denmark
tped@itu.dk

Miguel Sousa
Future-Shape GmbH
Höhenkirchen-Siegertsbrunn
miguel.sousa@future-shape.com

UbiComp '13, September 08 - 12 2013, Zurich, Switzerland

Abstract
Sterility restrictions in surgical settings make touch-less interaction an interesting solution for surgeons to interact directly with digital images. The HCI community has already explored several methods for touch-less interaction including those based on camera-based gesture tracking and voice control. In this paper, we present a system for gesture-based interaction with medical images based on a single wristband sensor and capacitive floor sensors, allowing for hand and foot gesture input. The first limited evaluation of the system showed an acceptable level of accuracy for 12 different hand & foot gestures; also users found that our combined hand and foot based gestures are intuitive for providing input.

Author Keywords
Gesture-based interaction, Touch-less interaction in Hospital, Wearable sensor, Floor sensor

ACM Classification Keywords
H.5.2 User Interfaces: Interaction Styles; J.3 Life and Medical Sciences: Health

Introduction
Surgeons need to interact frequently with an increasing number of computerized medical systems before and during surgeries in order to review medical images and

records. However, computers and their peripherals are difficult to sterilize, so usually during a surgery, an assistant or nurse operates the mouse and keyboard for such interactions. This indirect interaction suffers from communication problems and misunderstandings [9]. That is one of the main reasons why, in recent years, touch-less interaction has been considered for use in operation theatres. In general, touch-less interaction has been implemented using vocal commands and body gestures. A limitation of voice-based methods is that they cannot usually distinguish between different people speaking in the same room, in addition to being sensitive to environmental noise [7]. As for gesture-based approaches, these systems can detect body gestures using different kinds of cameras or body-worn sensors. The main challenge of vision-based systems is low accuracy in difficult lighting conditions, difficulty in coping with cluttered backgrounds, and occlusions by staff or equipment [4].

In this paper, we designed and implemented a hand gesture-based system based on inertial body-worn sensors to interact with medical images in the operation room. We believe that such sensors could eventually be integrated into the surgeon's garment. We also implemented a foot gesture-based system using a capacitive floor sensor, as a clutch mechanism, to control interaction with different systems. Our approach is motivated by the fact that hand-based gesture interaction has been recognized as a highly natural way of interaction [16], and clinicians already use their foot to control certain medical devices in hospitals; therefore, foot-based gesture could potentially be a suitable input mechanism for interaction in the operation theater. We report on a first limited evaluation of the system both with respect to the accuracy of the hand & foot based gesture recognition system and on usability aspects of the system.

Related Work

Interaction, or more precisely "user input", is touch-less if it can happen without mechanical contact between any part of the system and human body [3]. In past touch-less interaction research, voice and gesture have been investigated by many researchers. Beyond these approaches, HCI researchers have also explored for example gaze tracking and brain-computer interfaces for touch-less interaction.

Voice input can be realized by recording the voice with a microphone and processing it through dedicated algorithms. To ensure a good functionality, training of the system usually is needed and special voice commands have to be remembered. As mentioned before, the main problem of this approach is the fact that the system will react to all people speaking in the room, especially if they have similar voices. This problem has been tackled by installing a microphone array instead of just a single microphone. The microphone array is able to detect where the speaker is standing and through this decide if the speech is an intended command [7]. The advantage of interacting through voice lies in the fact that it is independent of any occlusions, and the user does not need to wear any additional devices.

In general, body gestures could be detected in different ways from using wearable sensors to environmental sensors. The most common approaches for touch-less interaction in operation rooms fall into two main categories: detecting gestures either with a vision-

based approach or with the help of body-worn sensors. In the vision-based approach, as with vocal interaction, the user does not need to wear any additional devices. However, a direct line of sight is needed for the interaction, where the users typically have to hold their hand in an unnatural position in order for the system to detect the gestures. The detection of the gestures can for instance be done with regular webcams [17], a stereo camera [6], a time of flight camera [15], or the Microsoft Kinect [5], [14]. The latter is becoming increasingly popular thanks to its low cost and easy implementation.

Body-worn sensors pose a good alternative to vision-based systems, as they do not require a direct line of sight. Furthermore, this type of sensor only allows a designated person to interact with the system, avoiding the potential confusion associated with having multiple people in the room the system is deployed in. A combination of inertial orientation sensors [4], gyroscope [19], and accelerometer are some of the most common body-worn sensors, which have been used to detect hand gestures for touch-less interaction in the hospital setting. For detecting foot gestures, several approaches can be used mainly divided into on-body sensors [1] and environmental sensors (i.e. sensitive floors). The latter category features different technologies including pressure [13], light refraction [2] or capacitive [18]. While capacitive approaches usually have less resolution, when compared to the others, these do not require for a soft or transparent surface, and easily allow for swipe-like foot gestures.

With body-worn sensors it is important to differentiate between gestures performed as intentional input the interactive system and gestures that take place as part of other activities (and should be ignored by the interactive system). One approach is to detect dedicated, easily distinguishable, gestures out of a continuous data stream [10] or enabling the system with another modality like user voice commands [4].

The main difference of our approach compared to previous work is that we are using a combination of hand and foot gestures to interact with several medical systems using both wearable and environment embedded sensors. Hand gesture commands have been designed for interacting with medical images, while foot gestures can enable, disable, and switch interaction between different systems.

Hand Gesture recognition method

Our hand-gesture detection is based on a wrist-worn IMU (Inertial Measurement Unit), which features a three-axis accelerometer, gyroscope and magnetometer, and transmits the data wirelessly, in real-time.

Modeling gestures

For interaction with the system a set of six well-distinguishable gestures have been identified with the aim of being intuitive for the user. During interaction, the initial state of the right arm is in a 90-degree angle. All hand gestures are depicted in Figure. 1. Apart from four regular movements "up" (1), "down" (2), "left" (3) and "right" (4) that are used for any menu-like navigation, an additional two gestures are implemented, i.e. tilting the hand to the left (5) and the right (6). These are used for special actions, such as a zoom-in or zoom out.

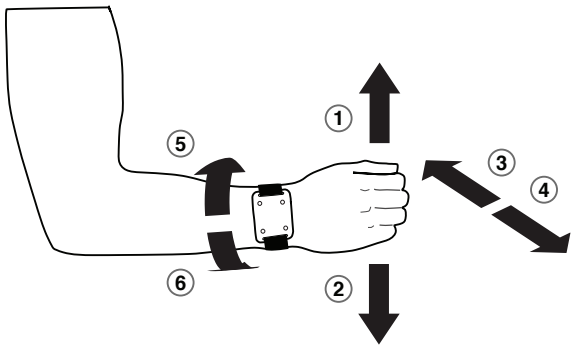

Figure 1. Hand gestures

The inertial sensor is attached, as shown in Figure 1, to the outer side of the right wrist in order to ensure that all movements are detected with maximum amplitude by the device while at the same time the hand and fingers are not obstructed.

Learning gestures

We had two different options for training our system to detect hand gestures: a person independent learner or training system for each individual. The latter strategy increases the accuracy of the system while the system needs to be calibrated for each user separately, which is time consuming. Since our hand gestures are simple and easy to learn, we followed the person independent strategy. For each gesture a training set of around 20 gestures is recorded and labeled. These are used as the input for a learner component, which is trained on the specific gestures. The trained learner is then used for the online gesture recognition.

Gesture recognition

The six gestures we chose for our system were best recognized using a two-stage strategy with a set of single-output artificial neural networks. The first stage used a single neural network to discriminate between the *idle* class and any of one the *gesture* classes. If a gesture class was detected in the first stage, a further set of 6 neural networks was used to predict the exact gesture class. This was done using 6 neural networks each trained to recognize one of the target gestures. The overall predicted gesture class was chosen to be the one for which the network output score was closest to the value assigned during training to its target class. All neural networks used default Matlab settings with a single hidden layer of size 4.

Features were computed from the ETHOS IMU [8] using a moving sample window of size 25, displaced by 9 instances between each collection (FIFO with 64% overlap between samples). Samples are drawn at ~21hz with about 2% error due to missing packets while transmitting the data via the Ant protocol.

The features extracted from each sample window were the following:

(Var(acc.x), Var(acc.y), Var(acc.z), Var(gyr.x), Var(gyr.y), Var(gyr.z), dGyr.x, dGyr.y, dGyr.z)

Where *Var()* stands for the variance measure taken over *acc* (accelerometer) and *gyr* (gyroscope) signals in the sample window. *dGyr* signals whether a gyroscope signal initially increases or decreases with respect to the signal baseline of the sample window. This was marked by a +1 or -1 respectively in the corresponding feature. The axes are specified after the '.' of each

signal source. We note that the neural network responsible for discriminating between the idle and gesture class only used the first six features while the 6 gesture-specific neural networks used all 9. The features were obtained by analyzing raw IMU signals, observing that each gesture had a characteristic gyroscope signal signature.

Foot gesture recognition method

Foot gestures are used for two main purposes in the system. In the first case, these serve as an enabler for user interaction with the displays. That is, different foot gestures are used to activate/deactivate the hand gesture detection sub-system. This allows the user to *freeze* the displays with a desired view on the display and start performing a given real-world task that requires hand movements, which otherwise would trigger false (unintentional) gesture commands.

The second case concerns switching between the multiple screens. This task could be accomplished by a more *exotic* hand gesture (such as circling) or a combination of the simple gestures (e.g., three times *gesture_right*). However, the first case requires movements that may not feel natural to the user, whilst in the second one, the user needs to memorize certain gesture combinations, in addition to taking more time to perform the switching task.

Six foot gestures are implemented in total. These are divided into two types: *tapping* and *swiping*. The detection of the latter type is eased by the use of a proximity sensitive floor (vs. pressure sensitive). Furthermore, both left and right feet can be used for interaction. The complete gesture set consists of: *swipe_left* [Figure 2, (1)], *swipe_right* [Figure 2, (2)], *double_tap_left*, *double_tap_right* [Figure 2, (3)], *swipe_down_left*, *swipe_down_right* [Figure 2, (4)]. Detection of the defined gestures on the floor is straightforward. Each of the triangular-shaped sensor cells is considered active or inactive (*ON/OFF*) depending on the proximity of the foot. As such, a *double-tap* is simply translated into a sequence of *ON-OFF-ON-OFF-ON* on a given cell (within a specific timing). The *(x,y)*-location of the cell encodes which foot was used for the tapping. As for the swipe movements, a set of *OFF-ON* sequences of neighboring cells is used, with the swipe direction depending on which cells are activated in a row.

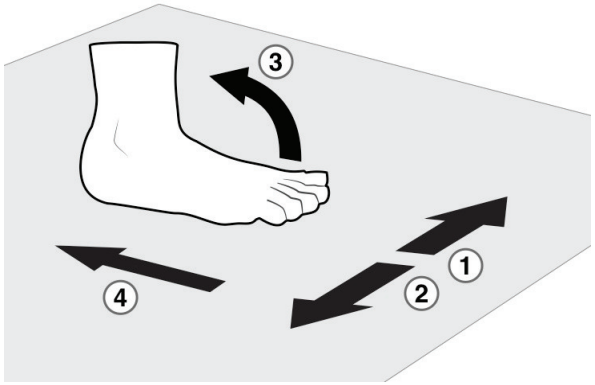

Figure 2. Foot gestures

System architecture

A schematic representation of the main components of the implemented system and the interconnection between different components is illustrated in Figure 3. The system consists of a wristband-style inertial sensor, a capacitive floor sensor, and the gesture recognition system. The inertial sensor and the floor

sensor transmit data to the gesture recognition system wirelessly.

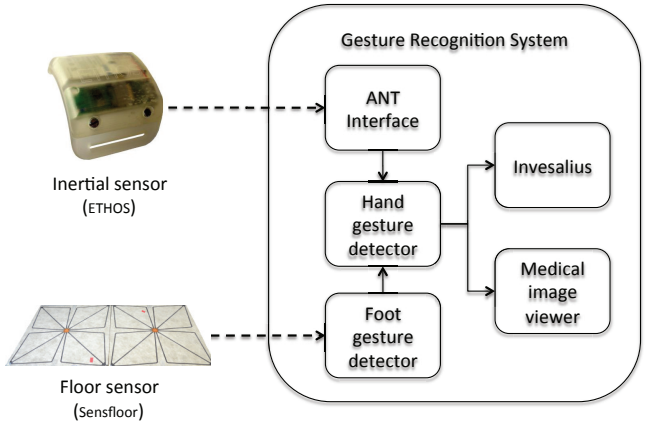

Figure 3. System Architecture

Inertial orientation sensor
In order to detect hand movements, we used ETHOS, which is a miniaturized IMU sensor developed at ETH Zurich for unobtrusive measurement of body movement. The ETHOS measures 3-D acceleration, 3-D rate of turn, 3-D magnetic field. The ANT+ module of the ETHOS provides wireless connection to transmit data to the server.

Capacitive floor sensor
SensFloor [11] is a textile-based underlay with embedded electronic modules equipped with radio transceivers. The floor uses a capacitive measurement approach to detect conductive objects, such as a *human foot*, placed upon itself (or hovering up to a few centimeters). A typical SensFloor unit (0.5x0.5m) has one electronic module with eight surrounding triangular sensor pads (Figure 4).

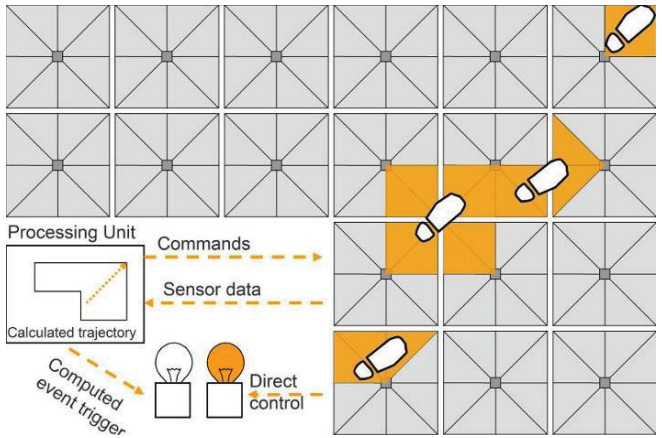

Figure 4. Floor sensor

Each time a significant change in capacitance is detected in one of the sensor pads, the module sends a wireless message to a central receiver, designated as *Smart Adapter*. Such message contains the (x,y) position of the module and the measurement data.

Gesture Recognition System
The entire system is running on a laptop with a dual core 1.4 GHz CPU and 4 GB RAM, running Windows 7. Matlab is used for processing the acquired data and interfacing with visualization applications. The laptop is connected to two screens that are used for displaying the image viewer as well as the InVesalius. Both applications are stand-alone and are controlled with commands received through TCP-IP. All parts of the system are described in more detail in the following subsections.

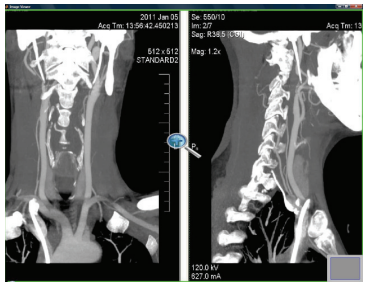

Figure 5. Screenshot of the Medical image viewer system including the magnifier symbol and colored border as feedback mechanisms of the system

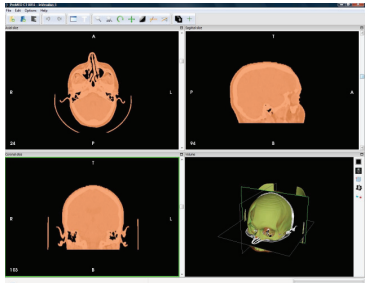

Figure 6. Screenshot of the Invesalius system

ANT INTERFACE MODULE

In order to connect the ETHOS to the gesture recognition system wirelessly, we developed an interface application, which receives the IMU data through the ANT protocol. This application corrects for any lost data and writes the received data on a virtual COM port, which is accessible by the Matlab application.

FOOT GESTURE RECOGNIZER

The sensor data transmitted from the floor modules is received by a Smart Adapter, which streams it to a COM port on the PC. A Matlab application was implemented to read and process such data, in order to detect the relevant gestures (see Figure 2). Each time a foot gesture is detected, the state of the system is changed. Specifically, the application encodes whether a screen is active, and which screen is currently in focus. The foot gesture recognizer gives a vocal feedback to the user after detecting a gesture, which is helpful for users to repeat a non-detected gesture immediately. Without this feedback, the user needs to wait for a longer time to see a visual feedback from one of the two displays.

HAND GESTURE RECOGNIZER

The data stream from the IMU input into the classifier, which processes the data and returns the predicted gesture. In order to avoid that one real gesture is detected as two subsequent gestures by the classifier, a gesture is only output when the previous three windows did not identify any gesture at all.

MEDICAL IMAGE VIEWER

Two different applications are used to demonstrate the user's interaction with the system. One consists of an image viewer, where the user can flip through medical pictures, as well as zoom in, zoom out, and pan in all directions (Figure 5). We used the color of the window around the image as a feedback to show if the image viewer application is activated or not. The green color illustrates the activated status while the gray window means it is inactive. When the user sends a command by performing a specific gesture the image viewer displays appropriate symbols on the screen for each action in addition to performing the action. (i.e. displaying a magnifier symbol for zooming in Figure 5). The other application is a public and open-source software for visualization of 3D medical images, *InVesalius* [12]. Here, the user can navigate through 2D slices of a CAT scan, in different views (Figure 6). When the InVesalius application is activated, the color of the activated window changes from black to green.

Evaluation

The main goal of our evaluation is answering to the following questions: To what degree does our combined hand and foot gesture recognition system correctly identify the commands from users of the system? Do users find the proposed combination of hand and foot gestures a viable alternative to the traditional indirect control of visualizations in operation theaters?

Evaluation of the classifier

The overall gesture recognition system was evaluated in two different ways. We first used a 3-fold cross-validation experiment over a recorded dataset from one of the authors and then confirmed the results in our user study where we recorded the live performance of the system over 5 other users. In both cases, the learner component was oblivious as to whether the data was coming from our recorded dataset or from a live sensor stream.

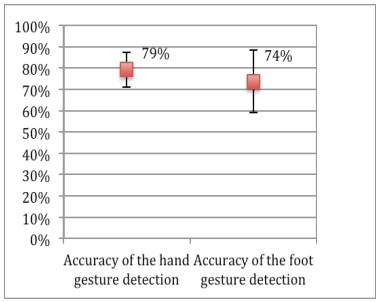

Figure 8. Accuracy of gesture detection in the user study

RESULTS FOR THE 3-FOLD CROSS-VALIDATION EXPERIMENT

In this experiment, we have a dataset of approximately 350 gestures, equally distributed among the 6 classes. Because we are streaming the data to the learner, a change in the sample window size results in a change in the number of points (and their features) received by the learner. We chose to segment the data into 3 roughly equal sized sub-datasets, i.e. 3 data collection runs taken over 3 different days, which determined the parameter for our cross-validation experiment. We report the confusion matrix in Figure 7 and precision and recall in Table1.

Gestures	Up	Down	Left	Right	Tilt_L	Tilt_R	Idle
Precision	0.91	0.93	0.96	1	0.95	0.98	0.99
Recall	1	0.98	1	1	1	1	0.99

Table 1. Precision and recall of the classifier

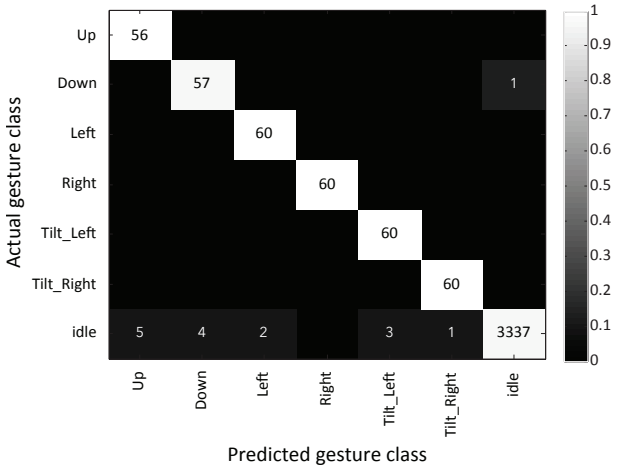

Figure 7. Gesture classification confusion matrix of the cross-validation experiment

The performance on the recorded data shows that the recall rate through classes is very high (low number of deletions), with a single missed 'down-gesture' throughout the experiment. The precision is slightly lower as some idle classes have been falsely predicted as gestures (insertions).

Evaluation of usability

In order to evaluate the usability and intuitiveness of the implemented system we conducted a qualitative user study in our (non-clinical) research lab with five participants (two females and three males), mainly computer science researchers (Figure 10). After introducing the whole setting and a short training on how to do the gestures, we asked users to try the system for about one minute, and then they were asked to go through a predefined scenario including ten steps which took in average three and half a minutes for each participant. Finally, we had a short interview with participants and they were asked to answer a list of seven questions with a five-point Likert scale that ranged from strong disagreement (1) to strong agreement (5). The result of the questionnaire (Figure 9) shows that in general users found that the combination of the foot and hand gestures is a valuable alternative for the indirect interaction. However during the evaluation, the average accuracy of the foot gesture detection was about 74% and this value was 79% for the hand gesture detection (Figure 8). In order to obtain the ground truth, we recorded the video of the participants during the evaluation, and the accuracy of the gesture detection has been calculated based on reviewing the recorded videos.

Figure 10. A user interacting with the system

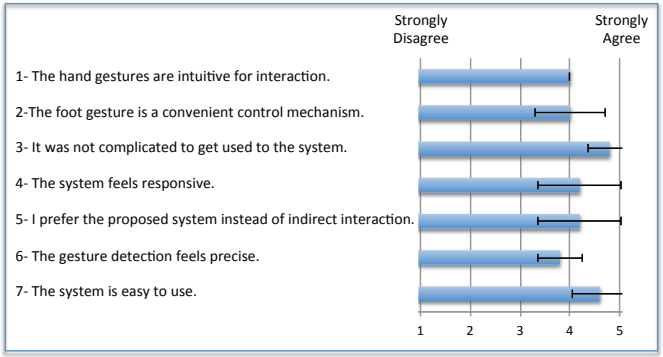

Figure 9. Result of the usability evaluation

Discussion & Conclusions

As the previous studies [4] also showed, using inertial body-worn sensors for interaction in the operation room gives us the possibility of interacting with several displays and distributing interaction between different users while it is independent of the user location. Moreover, since we are using a finger-free method surgeons can interact with the system when they have surgical instruments in their hands.

The combination of hand and foot gestures helped us implement several gestures (6 hand gestures with just one wristband inertial sensor and 6 foot gestures) with a high accuracy of detection. In addition from our limited user study, we can argue that using our foot gesture detection mechanism could be an intuitive and accurate alternative for other mechanisms of controlling interaction such as user's voice. However, our users found some of the foot gestures more intuitive. For example, the double tap gesture was preferable to the swipe gestures because of the user's balance problem during the swipe gesture.

Our gesture detection method showed a high accuracy in theory; however, we observed some gesture confusion during the user study. Because our hand gesture recognition system was calibration-free, which means we used a fixed training dataset sampled by one person for all participants. But each participant did the gestures slightly different that decreased the accuracy of classification. In the further steps, we expect that by adding more training data from different people or calibrating the system for each user, the accuracy of the system would be improved. Also for the foot gesture recognition, we observed some none-detected gestures since we used a capacitive floor sensor, and our users wore dissimilar shoes with different thicknesses. In order to avoid wrong positive gestures, we set the threshold of the capacitive sensor on the minimum value; however, for a real application, this value could be adjusted based on the standard shoes in the operation theatre. As it is clear in the Figure 10, in our current implementation we used a small piece of floor sensor contacting two cells that limited users to stand on a specific position to perform foot gestures. But as a future step, it is possible to detect gestures independent of the user's location on a bigger SensFloor. In the next step, we will ask real surgeons to evaluate our system for a specific surgery.

ACKNOWLEDGEMENT

This work was supported by the EU Marie Curie Network iCareNet under grant number 264738.

References

[1] Alexander, J., Han, T., Judd, W., Irani, P., Subramanian, S. Putting Your Best Foot Forward: Investigating Real-World Mappings for Foot-based Gestures. In Proc. *CHI'12*, 1229-1238.

[2] Augsten, T., Kaefer, K., Meusel, R., Fetzer, C., Kanitz, D., Stoff., T., Becker, T., Holz, C. and Baudisch, P. Multitoe: High-Precision Interaction with Back-Projected Floors Based on High-Resolution Multi-Touch Input. *UIST '10*, 209-218.

[3] Barré, R., Chojecki, P., Leiner, U., Mühlbach, L., & Ruschin, D. Touchless Interaction- Novel Chances and Challenges. In Proc. of the *13th International Conf. on Human-Computer Interaction, Part II*, *San Diego, CA, USA, 2009*, 161-169.

[4] Bigdelou, A., Schwarz, L., Navab, N. An adaptive solution for intra-operative gesture-based human-machine interaction. In Proc. Of the *ACM international IUI '12 Conf.*, *2012, p. 75.*

[5] Ebert, L.C., Hatch, G., Ampanozi, G., Thali, M. J., Ross, S. "You Can"t Touch This: Touch-free Navigation Through Radiological Images", *SAGE*, 2011.

[6] Graetzel, C., Grange, S., Fong, T., and Baur, C., A Non-Contact Mouse for Surgeon-Computer Interaction, NCCR-COME Research Networking Workshop, Brauwald, Switzerland, August 2003.

[7] Hands-free interaction in the hospital. Retrieved April 29,2013 from: http://www.healthcare.philips.com/pwc_hc/main/about/assets/Docs/medicamundi/mm_vol49_no3/mm_493_Technology%20News.pdf

[8] Harms, H., Amft, O., Winkler, R., Schumm, J., Kusserow, M., Troester, G., ETHOS: Miniature orientation sensor for wearable human motion analysis, Proceedings of *IEEE Sensors Conf.*, 2010, 1037-1042.

[9] Johnson, R., O'Hara, K., Sellen, A., Cousins, C., and Criminisi, A. Exploring the potential for touchless interaction in image-guided interventional radiology. In Proc. *SIGCHI Conf. CHI,* 2011, 3323-3332.

[10] Junker, H., Amft, O., Lukowicz, P., Troster, G. Gesture spotting with body-worn inertial sensors to detect user activities. *Pattern Recogn.* 41, 6, June 2008, 2010-2024.

[11] Lauterbach, C., Steinhage, A. and Techmer, A., Large-area wireless sensor system based on smart textiles, *9th International Multi-Conf. on Systems*, Signals and Devices, 2012, 1-2.

[12] Martins, T.A.C.P., Santa Ba´rbara, A., Silva, G.B., Faria, T.V., Cassaro, B.,Silva, J.V.L., InVesalius: Three-dimensional medical reconstruction software. In Proc. of the *3rd International Conf. on Advanced Research in Virtual and Rapid Prototyping*, 2008, 135-141.

[13] Paradiso, J., Abler, C., Hsiao, K., Reynolds, M. The Magic Carpet - Physical Sensing for Immersive Environments. In Proc. *CHI' 97*, 277-278.

[14] Ruppert, G. C. S., Reis, L. O., Amorim, P. H. J., de Moraes, T. F., & da Silva, J. V. L. Touchless gesture user interface for interactive image visualization in urological surgery. *World Journal of Urology*, 30(5), 2012, 687-691.

[15] Soutschek, S., Penne, J., Hornegger, J., Kornhuber, J. 3-D gesture-based scene navigation in medical imaging applications using time-of-flight cameras, in: Workshop on Time-of-Flight based Computer Vision, Anchorage, Alaska, June, 2008.

[16] Varona, J., Jaume-i-Capó, A., Gonzàlez, J., Perales, F.J. Toward natural interaction through visual recognition of body gestures in real-time. *Interacting with Computers* 21(1), 2008, 3–10.

[17] Wachs, J., Stern, H., Edan, Y., Gillam, M., Feied, C., Smith, M., Handler, J. Gestix: A Doctor-Computer Sterile Gesture Interface for Dynamic Environments, *Soft Computing in Industrial Applications* vol. 39, Springer Berlin / Heidelberg, 2007,30-39.

[18] Zimmerman, T. G., Smith, J. R., Paradiso, J. A., Allport, D., Gershenfeld, N. Applying electric field sensing to human-computer interfaces, In Proc. *SIGCHI Conf. CHI*, Denver, CO, 1995, 280–287.

[19] Zinnen, A., Schiele, B., Ziegert, T. Browsing patient records during ward rounds with a body worn gyroscope. In Proc. of the *ISWC '07*, 1-2.