

A Usability Study of a Gesture Recognition System Applied During the Surgical Procedures

Antonio Opromolla^{1,2}, Valentina Volpi^{1,2(✉)}, Andrea Ingresso¹,
Stefano Fabri³, Claudia Rapuano³, Delia Passalacqua³,
and Carlo Maria Medaglia¹

¹ Link Campus University, Via Nomentana 335, 00162 Rome, Italy
{a.opromolla, v.volpi, a.ingrosso,
c.medaglia}@unilink.it

² ISIA Roma Design, Piazza della Maddalena 53, 00196 Rome, Italy

³ Consorzio Roma Ricerche, Via Giacomo Peroni 130, 00131 Rome, Italy
{fabri, rapuano, passalacqua}@romaricerche.it

Abstract. Within an operating room, surgeons need to interact with a large amount of patient's medical information and data. In order to avoid misunderstandings among the staff and protecting the patient safety, the medical staff may use a touchless interaction system that allows the surgeons to directly interact with digital devices that visualize digital images. The RISO project aims to create a gesture recognition system for the visualization and manipulation of medical images, useful for the surgeons even during the surgical procedures. In this paper we show the main findings from a usability study carried out with the aim to evaluate, among others, the learnability of the system and the memorability of the gestures employed for the interaction.

Keywords: Touchless interaction · Gesture recognition · Usability · Surgery · Operating room

1 Introduction

Within an operating room, surgeons need to interact with a large amount of patient's medical information and data (e.g.: images, records, etc.). So, the number of computerized devices employed for accessing them is growing.

Since the electronic devices (PC primarily) are difficult to sterilize, currently a middle-person, usually a nurse, assists the surgeon by retrieving data and information through mouse and keyboard at a PC station located inside the operating room [1]. The indirect nature of this type of interaction increases the risk of causing several misunderstandings among the staff and generating an overall slowness in surgery procedures. A touchless interaction with the digital devices can represent an effective solution for surgeons, since it allows them to directly interact with digital images.

The most part of touchless systems have been implemented using voice commands and gestures. However, the speech recognition systems are limited because of their

difficulties to discriminate between different people talking in the same room, as well as because of their high noise sensitivity. The gesture recognition systems can be implemented through different typologies of video cameras and sensors, but they generally lack in accuracy under low light conditions. However, in comparing speech and gesture recognition systems it emerges that the latter allow a more “natural interaction”; first of all, gesture recognition systems free the surgeon from interacting with additional accessories which might cause him/her physical hindrances.

In the last years, the use of hardware and software systems in medical field has increased. These systems are used for storage, transmission, display, and printing of digital diagnostic images, such as PACS (Picture Archiving and Communication Systems). Indeed, thanks to DICOM (Digital Imaging and Communications in Medicine), a standard used in PACS, the surgeons can manipulate the diagnostic images both in 2D and in 3D [2, 3].

The RISO project (in Italian the acronym stands for “Rilevazioni Immagini in Sala Operatoria; in English, “Image Recognition in Operating Room”) presented in this paper aims to create a gesture recognition system for the visualization and manipulation of medical images, useful for the surgeons even during the surgical procedures.

This paper is organized as follows. In the second section, we will discuss about the main projects that use natural interaction systems into the operating room, during a surgery. In the third section, we will describe the RISO system, focusing both on the system architecture and on the user interface and the interaction modes. In the fourth section, we will focus on the usability evaluation study conducted on the RISO system, by describing the methodology and the main findings of this analysis. Finally, in the fifth section, the conclusions.

2 Related Work

In the touchless interaction, the user inputs data without any contact between the computer components and the human body. So, gestures are suitable means to be used in touchless systems. In literature, several interaction techniques using gestures have been analysed, in order to study the different ways to recognize them, such as wearable or environmental sensors. The more common touchless interactions technique used in the operating room can be divided into two main categories: systems detecting gestures with a visual approach or with wearable sensors. The first does not require the user to wear any additional device, but it needs a direct line of sight between the user and the video capture device to let the system detect gestures. Webcam, stereo camera, ToF (Time of Flight) camera or Microsoft Kinect are some of the devices that can be used to recognize gestures. The wearable sensors instead do not require the direct line of sight. Moreover, since they allow only the person wearing the sensors to interact with the system, they avoid the possible confusion deriving from the presence of more people in the room where the system is located.

Voice commands and sensors are often used together with visual recognition systems to enable or disable particular states or mode of the system. Ebert et al. [2] developed a touchless PACS system using the Microsoft Kinect (to capture the video streaming) and a wireless microphone (to capture voice commands). The system has

three main modes to control the patient's data that can be switched by using voice commands: stack navigation mode, move mode, and the window mode. All three use the movement of one or two hands.

However, in a further system implementation proposed by Ebert et al. [4], in addition to enable gestures for viewing and manipulating the medical images, the voice commands were deactivated, since the results from the previous study showed that the voice recognition were too subject to background noise and it struggled with accents (it worked poorly if the user had a non-American accent). Moreover, many users considered the required headset ungainly and distracting. The main problems encountered during the previous test were overcome by the use of finger gesture detection, by which it was possible to control the basic functionality of the medical image viewer. The finger gesture detection allows gestures such as panning, scrolling, and zooming.

Even the system designed by Ruppert et al. [5] uses the Kinect to capture the video stream, but in this case the tracked position of the user hand is used to move the mouse pointer. This allows the user to perform mouse-drag functions like 3D rotations and 2D slices change by using one or both hands. Button clicks events are also virtually generated.

Soutschek et al. [6] include the "mouse" movable cursor and click (used to measure the size of anatomical structures or specify a Volume Of Interest for further analysis) among the functional requirements considered as basic in the visualization of the medical data set (the others are rotation and translation, in order to explore and navigate 3-D data sets, and reset). These actions correspond to 5 gestures captured by a ToF camera system.

Even a simple camera positioned on the screen and a tracking module can be used to acquire and to interpret the movements, as in Gestix, a system developed by Wachs et al. [7], which recognizes both static and dynamic poses.

On the contrary, the WagO system developed by Kipshagen et al. [8] combines an image processing component, which automatically determines the user's hand and its contour position, and a gesture recognition component. Some stereo-cameras positioned below the screen and pointing to the ceiling of the operation theatre triangulate the hand positions in 3D and map them to the 2D environment of the OsiriX (an open source DICOM viewer) application. The system recognizes a set of 4 static gestures to control the file DICOM viewer OsiriX.

Hands are not the only human body parts tracked by video capture devices in order to interact with touchless medical imaging systems. Gallo et al. [9] developed an open-source system for the exploration of medical imaging data that capture both static and dynamic hand and arm gestures through a Microsoft Kinect.

Similarly, Jacob et al. [10] uses the Microsoft Kinect to capture the user's skeleton, in order to provide the positions of various landmarks placed on the human body. The system includes an intention recognition module that is able to decide whether a performed gesture is intentional or not on the basis of anthropometric and kinematic features of the human body, such as the torso and the head orientation and the hands position.

3 The RISO System

The RISO project aims to create a gesture recognition system for the visualization and manipulation of medical images.

In order to appropriately design and develop the RISO system, an accurate analysis has been conducted. It concerned both the technical solutions (mainly about the gesture recognition devices and the toolkits for the application development) and the gestures to use in a touchless interaction. The characteristics of the context of use, i.e. the operating room, was the main criterion in the selection of the technological solutions and consequently of the gestures for the interaction with the RISO system.

In details, referring to the RISO system, the interaction with the medical images occurs through the Leap Motion, a hardware sensor device that supports hand and finger motions as input. Thanks to its sensor accuracy, the Leap Motion recognizes the hands and fingers small motions, allowing a touchless interaction with the devices. This is a perfectly suitable feature for an operating room, where usually the medical team, the equipment and the machinery could hamper the surgeon movements. Furthermore the operating room is a perfect environment for the use of Leap Motion, because their surgical lights do not have IR components that interfere with the device.

Shown below the adopted solutions, concerning both the system architecture (Sect. 3.1) and the interaction modes (Sect. 3.2).

3.1 System Architecture

The hardware architecture of the RISO system is composed of a depth sensor used as input device, a computer that processes input data and manage the GUI (Graphical User Interface), and a display that shows the results. The software architecture is composed of two components: the module for gesture recognition and the module for data handling and visualization.

Input Device. The input device is the component that required more attention. In the gesture recognition field, ToF cameras (Time of Flight), also called “depth cameras”, are the most used devices for image acquisition. They allow estimating the distance between the cameras and the objects in real time, by measuring the time-of-flight of a light signal between the camera and the subject for each point of the image. Using these devices allows making faster the segmentation of a hand, because they can detect skin of different colours even in conditions in which there are noisy backgrounds.

Generally, all ToF devices exploit the same technology. In details, they use in combination one or more cameras VGA and a depth sensor. The main differences that we note between the various devices are the price, the sensor accuracy, and the camera resolution.

Among all the analysed devices, Leap Motion was the most suitable for RISO. This is an innovative USB device with three IR sensors and two near IR cameras. It is designed to be placed on a flat surface under the area where hand movements are performed. It can track the movements of the fingers, palms or objects used as pointers (like pen or pencils). Movements are detected in a range between about 0.1 m and

1.0 m. Leap Motion has good performance, high precision, minimal amount of space and its price is much lower than the cost of other available devices (Fig. 1).



Fig. 1. The Leap Motion device

Libraries. Leap Motion provides a development kit that provides high-level methods for identifying hands and fingers. So, in order to develop the gesture recognition component of RISO system we used Leap Motion SDK. By using it, it is possible to perform the segmentation of the hands, feature extraction, such as fingers, palms or pointers, and finally tracking them. RISO uses Leap Motion libraries to recognize gesture and poses made by the users. These gestures and poses are mapped with the possible inputs transmitted to the module that manages clinical data.

To develop the module for clinical data handling and visualization we used the Medical Imaging Interaction Toolkit, an open-source software system for development of interactive medical image processing software.

3.2 Interaction Modes

Once defined the technological solutions to be adopted, as described in the Sect. 3.1, the 9 gestures through which the user has to interact with the RISO system were identified mainly on the basis of intuitiveness and memorability. Each gesture is named as the action that it enables: “cursor positioning” (moving along x-axis and y-axis), “moving in depth” (along z-axis), “zooming” (reshaping the image), “windowing” (changing the values on the Hounsfield scale, that is the quantitative scale describing radio-density), “image shifting” (moving the image when it is not totally visualized in the window), “resetting” (taking the cursor and the image to the starting point), “selecting” (set the focus on a specific image), “releasing” (a specific image area), “asking help”. Figure 2 shows the enabled gestures.

Referring to the interaction flow, the user has to: (1) select the folder collecting the health examinations of a specific patient; (2) see the list of the examinations carried out by the patient; (3) select a specific health examination and explore it (Fig. 3).

4 The Usability Evaluation

A usability evaluation study was conducted on a prototype of the RISO system. Since a prototype version of the Leap Motion device was used during the test, in some cases the overall system lacked in stability. However, this critical condition had not invalidated the aim of the usability evaluation study.

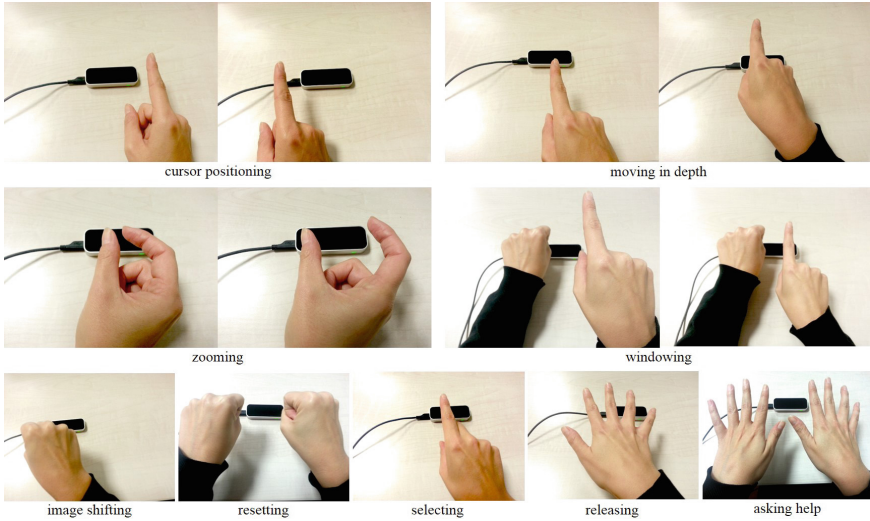


Fig. 2. The gestures enabled for interacting with the RISO system



Fig. 3. The interaction flow of the RISO system. In order: (1) the user selects the folder collecting the health examinations of a specific patient; (2) he/she sees the list of the health examinations carried out by the selected patient; (3) he/she selects a specific health examination and explores it.

In details, the study aimed to evaluate the ease of learning of the system, the ease of execution of the gestures, the suitability of the employed technologies, and the memorability of the gestures.

The Fig. 4 shows some frames collected during the usability test.

4.1 Methodology

The usability tests involved (one by one) 10 users and were organized as follows.

First of all, the user filled an “entry questionnaire” (consisting of 4 closed questions and 3 open-ended questions) designed to investigate his/her confidence with the gesture recognition systems and technologies. From this questionnaire emerged that the respondents had an average knowledge of the gesture recognition systems and technologies. They mostly used the “Nintendo Wii” (a home video game console) and the

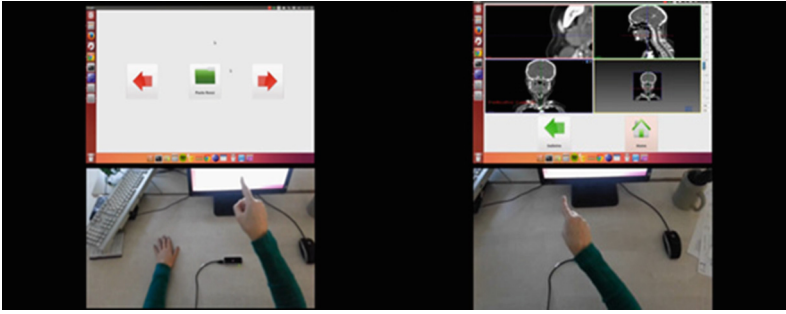


Fig. 4. Frames collected during the usability test of the RISO system

“Microsoft Kinect” (a motion sensing input device), mainly with playful purposes, while none of the respondents had used the Leap Motion device before.

Once filled the entry questionnaire, a short explanatory video focused on the interaction modes for the RISO system was shown to the test participant, in order to get him/her familiar with the RISO application and with the 9 available gestures.

Then, the user started the usability test. In particular, the participant individually performed a sequence of 11 tasks repeated three times. The repetition of this sequence had been necessary to measure the memorability of the gestures. A camera pointing on the monitor screen of the RISO system recorded each sequence.

Each task of the sequence required the participant to use one (or more) gesture. In detail: task 1 – assessing to the X-ray examination (“cursor positioning” and “selecting”); task 2 – “zooming” a selected image; task 3 – “shifting” the image; task 4 – “releasing” and “resetting” the image; task 5 – “windowing”; task 6 – choosing the CT Scan examination (“cursor positioning” and “selecting”); task 7 – “selecting” an image and “zooming” it; task 8 – “shifting” the selected image; task 9 – “windowing”; task 10 – “moving in depth” the image; task 11 – “asking help”.

The usability tests were carried out in a controlled test environment (not an operating room): this condition has been a weak point of our study of usability, but it has not been an impediment to the satisfaction of its overall objective, since we mainly aimed to evaluate the learnability of the first prototype of the RISO system.

During each test session three people were taking part: a test supervisor, who taught participants the basic instructions to perform the usability test (e.g.: follow the script and run through tasks); an observer, who took notes regarding events and activities (e.g.: what the participant did, his/her facial gestures, body language, and verbatim comments, etc.); a camera operator, who recorded the test session.

Finally, the user filled an “exit questionnaire” addressed to evaluate his/her own experience in using the RISO system. The questionnaire was composed of 1 closed question, 3 open-ended questions, and a 5-point Likert scale with 21 items. Each of the 21 items was represented by a sentence describing RISO, in relation to which the respondents had to express their level of agreement or disagreement (1: strongly disagree; 5: strongly agree). Some of these sentences described RISO positively (positive position), the other sentences negatively (negative position). The evaluation of the exit

questionnaire concerned: the user experience with the employed technologies and the gestures enabled for interacting with the system, the adequacy of the system features, the suitability of the RISO system within the context of use, and the user perception of the main advantages and disadvantages of the application.

4.2 Findings

Following, the main findings from the usability test analysis illustrated on the basis of the specific investigated elements.

Employed Technologies. Although the users declared they had some difficulties in understanding the operational mode of the controller (Leap Motion), they said that the activities to be performed had not been hindered by the use of this device.

Context of Application. Although the usability tests were carried out in a controlled test environment, the test participants declared that in their opinion the RISO system could be easily integrated within the real environment (the operating room). Most of the users thought that the RISO system would be able to radically innovate the context of application and to improve the work of a surgeon. However, some of the respondents pointed out the possible hindrances that could occur into the real environment conditions (e.g. low light, the distance from the Leap Motion, etc.).

Gestures. The users declared that repeating three times the same task sequence was useful in order to memorize the gestures and to perceive more natural the interaction with the system. The gestures were considered easy to perform, although some of them turned out to be not totally intuitive (especially the “windowing” gesture) or not well performing (especially the “zooming” gesture).

System Features. The users considered the system features useful for surgeon purposes. However, the test participants recommended additional features that could be implemented. Among them: sending via Internet the patient information, updating the patient information, adding comments to a selected image.

Perceived Advantages. The respondents declared that the main advantages of the RISO system were: the possibility to access the patient information without touching other instruments or compromising the hygiene of the operating room, and bypassing the misunderstandings with assistants; the higher speed in consulting patient information, with respect to the traditional tools; the large freedom of movement of the surgeon having no physical obstacles; the natural interaction, that allowed the surgeons to employ the cognitive resources only on the surgical procedures (rather than in understanding how to interact with the devices).

Perceived Disadvantages. The respondents declared that the main disadvantages of the RISO system were: the necessary training for interacting with the system; the possibility to soil the devices (primarily the Leap Motion), compromising the operation of the system; the greater lack of precision in interacting through the Leap Motion, with respect to the traditional input devices (e.g. the mouse); the time (considered too long)

spent by the system in recognizing the single gestures; the general lack of “flexibility” of the system.

Improvement. As mentioned above, each of the involved users performed the tasks sequence three times, in order to measure the memorability of the gesture and, as a consequence, the improvement in interacting with the RISO system. The time spent for each task (and for each repetition) by the participants was considered as an index useful to measure these abilities. The Fig. 5 shows the average of the time (in seconds) spent by users for each task during the three repetitions.

According to these data, for 5 tasks out of 11 there was a decrease in the time spent to perform each task, from the first to the third repetition; on the contrary, 6 tasks out of 11 showed a more irregular trend.

However, if we consider the total time taken to carry out the single repetition, we note that the average of the time spent for the repetition 1 amounts to 249,3 s, 195 s for the repetition 2, and 186,8 s for the repetition 3. So, we can affirm that the average of the time spent for each repetition decreases from the first to the third one.

Moreover, we note that each task, taken one by one, shows a different average of the time spent to accomplish it (the “longer” is the task 10).

The Fig. 6 represents the average of the time (in seconds) that each user spent during the three repetitions of the 11 tasks.

According to these data, for 3 users out of 10 there was a decrease in the time used to perform each sequence of tasks, from the repetition 1 to the repetition 3; on the contrary, 6 users out of 10, show a more irregular trend. This is mainly attributed to the low stability of the system, due to the use of a prototype of the Leap Motion in this release of the RISO system.

Moreover, the participants carried out the single repetition (and the single tasks) at different times. So, we can identify three typologies of users: “Fast users” (users 2, 4,

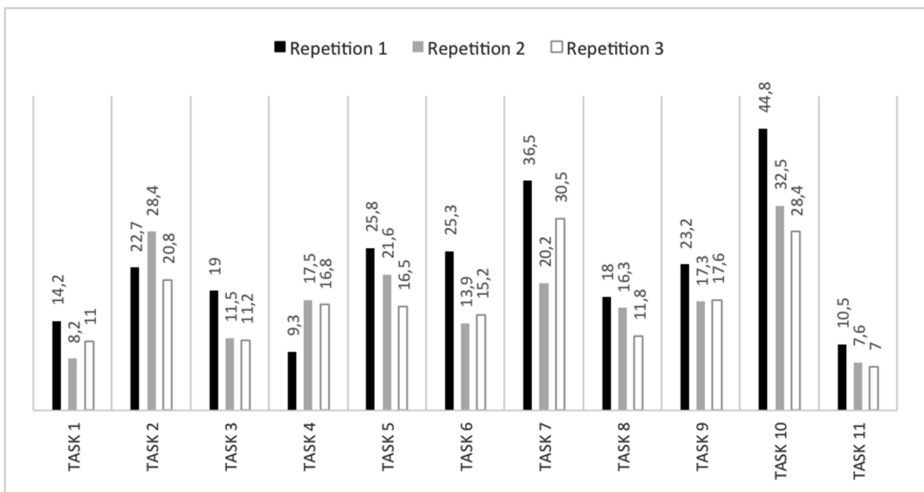


Fig. 5. The average of the time (in seconds) spent by users for each task during the three repetitions.

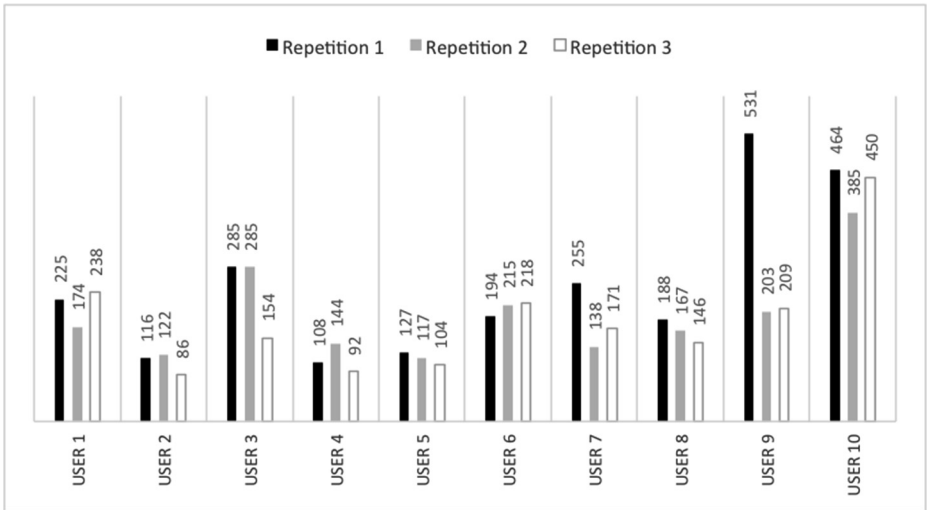


Fig. 6. The average of the time (in seconds) spent by each user during the three repetitions of the 11 tasks.

and 5), who carried out the test in a very short time; “Average users” (users 1, 3, 6, 7 and 8), who carried out the test in a longer time; “Slow users” (users 9 and 10), who carried out the test in a long time.

Success Rate. Although some tasks (i.e. the gestures employed during the interaction with the RISO system) were more problematic than others and the system was not so much stable, all the involved participants completed all the repetitions and all the assigned tasks. So, we recorded a 100 % success rate. For this reason, the effectiveness of the RISO system, represented by “the accuracy and the completeness with which specified users can achieve specified goals in particular environments” [11] can be considered as high.

5 Conclusion

In this paper, we showed the main findings of the usability study of the first prototype of the RISO system, a gesture recognition system for the visualization and manipulation of medical images during surgical procedures.

On the whole, we can affirm that the results of this study are positive, considering the learnability of the gestures enabled for the interaction and the usefulness of the system perceived by users.

The test setting had two main weak points. The first was the overall system stability, due to: the use of a Leap Motion prototype, the noise caused by IR components of the environment lighting, and the prototypal stage of the software. The second was the test environment, that was a “controlled” environment, rather than a “real” environment.

Considering this, it would be interesting to carry out a usability study on a more stable RISO prototype, within an environment very similar to an operating room. Our purpose will be to evaluate the precision, the flexibility, and the time spent in recognizing a gesture enabled by the RISO system in the real context of use.

References

1. Hartmann, B., Benson, M., Junger, A., Quinzio, L., Röhrig, R., Fengler, B., Färber, U.W., Wille, B., Hempelmann, G.: Computer keyboard and mouse as a reservoir of pathogens in an intensive care unit. *J. Clin. Monit. Comput.* **18**(1), 7–12 (2004)
2. Ebert, L.C., Hatch, G., Ampanozi, G., Thali, M.J., Ross, S.: You can't touch this: touch-free navigation through radiological images. *J. Surg. Innov.* **19**(3), 301–307 (2012)
3. Rosset, A., Spadola, L., Ratib, O.: OsiriX: an open-source software for navigating in multidimensional DICOM images. *J. Digit. Imaging* **17**(3), 205–216 (2004)
4. Ebert, L.C., Hatch, G., Thali, M.J., Ross, S.: Invisible touch—control of a DICOM viewer with finger gestures using the Kinect depth camera. *J. Foren. Rad. Imaging.* **1**(1), 10–14 (2013)
5. Ruppert, G.C., Reis, L.O., Amorim, P.H., de Moraes, T.F., da Silva, J.V.: Touchless gesture user interface for interactive image visualization in urological surgery. *World J. Urol.* **30**(5), 687–691 (2012)
6. Soutschek, S., Penne, J., Hornegger, J., Kornhuber, J.: 3-D gesture-based scene navigation in medical imaging applications using Time-of-Flight cameras. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1–6 (2008)
7. Wachs, J., Stern, H., Edan, Y., Gillam, M., Feied, C., Smith, M., Handler, J.: A real-time hand gesture interface for medical visualization applications. In: Tiwari, A., Roy, R., Knowles, J., Avineri, E., Dahal, K. (eds.) *Applications of Soft Computing*. AISC, vol. 36, pp. 153–162. Springer, Heidelberg (2006)
8. Kipshagen, T., Graw, M., Tronnier, V., Bonsanto, M., Hofmann, U.G.: Touch-and marker-free interaction with medical software. In: Dössel, O., Schlegel, W.C. (eds.) *World Congress on Medical Physics and Biomedical Engineering*, Munich, Germany, 7–12 September 2009. *IFMBE Proceedings*, vol. 25(6), pp. 75–78. Springer, Heidelberg (2009)
9. Gallo, L., Placitelli, A.P., Ciampi, M.: Controller-free exploration of medical image data: experiencing the Kinect. In: *24th International Symposium on Computer-Based Medical Systems*, pp. 1–6 (2011)
10. Jacob, M., Cange, C., Packer, R., Wachs, J.P.: Intention, context and gesture recognition for sterile MRI navigation in the operating room. In: Alvarez, L., Mejail, M., Gomez, L., Jacobo, J. (eds.) *CIARP 2012*. LNCS, vol. 7441, pp. 220–227. Springer, Heidelberg (2012)
11. ISO 9241-11:1998 Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability