ORIGINAL ARTICLE

Touchless gesture user interface for interactive image visualization in urological surgery

**Guilherme Cesar Soares Ruppert · Leonardo Oliveira Reis ·
Paulo Henrique Junqueira Amorim · Thiago Franco de Moraes ·
Jorge Vicente Lopes da Silva**

Abstract

Purpose In the operating room (OR) a touchless interface is an ideal solution since it does not demand any physical contact and still can provide the necessary control features in a cleansed and sterilized environment.

Methods Using open-source software libraries and image processing techniques, we implemented a hand tracking and gesture recognition system based on the Kinect device that enables surgeon to successfully touchlessly navigate through the image in the intraoperative setting through a personal computer. We used the InVesalius software, which provides high-quality 3D reconstruction of medical images.

Results Computed tomography data were intraoperatively fruitfully accessed through a simple and cheap solution in 4 tumor enucleations in 3 male patients in whom elective nephron-sparing surgeries were performed for small non-exophitic tumors. Mean tumor length was 2.7 cm (2.1, 2.7, 2.9 and 3.1 cm), and real-time ultrasound was not necessary for intraoperative identification in 3 of 4 endorenal tumors. All pathological reports revealed renal cell carcinoma, Fuhrman grade I, and negative inked surgical margins. No intra- or postoperative complication was reported.

Conclusions For the first time in the literature, a touchless user interface solution applying the Kinect device showed to be very efficient and enabled a low-cost and accurate control of the software InVesalius intraoperative, just using hand gestures. It can be used with any mouse-controlled software, opening an avenue for potential applications in many other areas, such as data visualization, augmented reality, accessibility, and robotics. The further validation and advancement of this technology are underway.

Keywords Accessibility · Robotics · Operating room · Natural · Control · Surgeon · Joystick-free environment

G. C. S. Ruppert · P. H. J. Amorim · T. F. de Moraes · J. V. L. da Silva
Division of Tridimensional Technologies, Renato Archer Center for Information Technology, Campinas, Brazil

L. O. Reis
Department of Surgery (Urology), School of Medical Sciences, University of Campinas, Unicamp, Campinas, Brazil

L. O. Reis (✉)
Faculty of Medical Sciences (Urology), Center for Life Sciences, Pontifical Catholic University of Campinas PUC-Campinas, Tessália Vieira de Camargo, 126, Cidade Universitária "Zeferino Vaz", Campinas, SP CEP: 13083-887, Brazil
e-mail: reisleonardo@zipmail.com.br; reisleo@unicamp.br; reisleo.l@gmail.com

Introduction

Computer interfaces based on gestures have been intensively researched [1, 2], but limitations like high cost, bad accuracy, and setup complexity have contributed to keep this technology impractical for real-world applications. This situation changed drastically recently with the release of an accurate, low-cost, and easy-to-setup device by Microsoft, called Kinect.

Gesture user interfaces have applications in many areas, and the present work is primarily focused on one particular application that is the visualization of 3D medical images during a urological surgical procedure. This scenario is particularly important because the operating room (OR) is a cleansed and sterilized environment to adhere to the principles of asepsis, and the contact of the surgeon with traditional computer interfaces (like mouse and keyboard) could lead to contamination, increasing the risk of patient infection.

Integrating recently developed virtual reality technologies into the OR to facilitate performance of nephron-sparing surgery could be applied to traditional open and laparoscopic surgery. Such systems are usually controlled using mouse and keyboard, requiring an undesired physical contact during the surgery [3].

We present a touchless gesture user interface that allows the surgeon to control medical imaging software just by performing hand gestures in midair. The system is also open-source, low-cost, and simple to deploy which allows a widely adopted solution.

Methods

The imaging software used in our system is the InVesalius [4], open-source software for visualization of 3D medical images developed at Renato Archer Information Technology Center (Campinas, SP, Brazil) that offers visualization in 2D slices (multi-planar reconstruction) and in 3D using a high-quality volume and surface rendering technology. Also, measurement tools (linear, angular and volumetric) are provided. It is freely available for Windows, Mac OS and Linux platforms in seven languages and is compatible with the most common file formats such as DICOM, Analyze and STL.

Two prototypes of gesture interface were developed using two different technologies to communicate with Kinect. The first one uses a simpler but completely open-source library called Libfreenect and is based on depth threshold and post-processing filtering to detect the hand. The second prototype uses a non-open-source solution (NITE/OpenNI) to detect the position of the hand. In both prototypes, the hand of the user is tracked and the corresponding position is used to move the mouse pointer and button clicks events are also virtually generated. The next sections describe in details the Kinect device and the two prototypes.

Kinect

Released in November 2010, Kinect is a relatively low-cost (about $149) device developed by Microsoft that provides joystick-free gaming just using body gestures recognition [5, 6]. The device provides a color camera, microphones, accelerometers, and a motorized tilt; however, the most important and novel component is a depth camera.

The depth sensing is provided by an infrared (IR) projector that projects a fixed structured light (consisting of a pattern with millions of small dots), and an infrared camera captures the reflections of this pattern and provides an 11-bit single-band image (i.e., from 0 to 2,048). The microprocessor inside Kinect performs stereo triangulation

[7] between the image captured by the infrared camera and the original pattern, obtaining the depth information for each point on the field of view. The output of this triangulation is the depth image, which is a gray-scale image, where the intensity of each pixel is proportional to the depth, which is the distance from the device to the object on the scene in the corresponding pixel position.

All the cognitive features like players and gesture recognition are not really performed by Kinect. The device itself just provides the cameras, leaving the recognition itself to be executed by software that implements state-of-art computer vision methods based on depth images.

Libfreenect

Kinect was designed only to be connected to Xbox, and there was no official support for any other platform. Libfreenect was the first successful attempt to communicate with Kinect from a personal computer. Basically, this library offers basic functions to fetch the RGB (red, green, and blue light) and depth images, as well as to read the other sensors.

OpenNI/NITE

OpenNI (Open Natural Interaction) and NITE are software components provided by Prime Sense, which is the company behind the Kinect depth camera technology licensed by Microsoft. OpenNI is an open-source framework for programming natural interfaces, providing a common programming interface for developing skeleton and hand tracking algorithms (plugin components) with a predefined list of body joints that the algorithms can track.

However, OpenNI does not come with any embedded algorithm; thus, it is necessary to install these components so OpenNI can actually be useful. That is where NITE takes place, being an OpenNI-compliant module that implements real-time skeleton tracking methods.

These algorithms use machine learning and probabilistic templates of the human body shape (based on thousands of modeled body samples). It is capable of detecting and segmenting multiple users on the scene and tracking 15 body joints (hands, elbows, shoulders, neck, head, torso, hips, knees and feet) in real-time with an impressive robustness and accuracy. Both are multi-platform and free of charge, but although OpenNI is open-source, NITE is proprietary and distributed only as a binary executable.

Results

A first prototype was made by using Libfreenect together with image processing operations in order to simulate mouse

events according to hand gestures. Only the depth image was used, and it was applied an image threshold according to a predefined distance where the user is expected to be standing in a way to get just a 50-cm range in front of the person's body. So when the user raises the hand in front of the body, just the hand pixels are selected. However, the result of this threshold is still noisy, unstable and present outliers pixels so post-processing steps are required.

We applied a minimum component filter to remove outliers regions smaller than 10 pixels and a mean filter [8] between frames to get better stability on the hand detection. Finally, the center of gravity of the hand is calculated. This position is used to generate a virtual mouse event moving the mouse cursor to that relative position in the screen. Figure 1 shows the processed image.

The mouse press event also is virtually generated when the user holds still the hand in the same position for one second, enabling rotations and button clicks. As it is almost impossible to hold the hand perfectly still, a displacement of up to 20 pixels was considered as no movement. The mouse press event is generated when the user removes the hand back outside the threshold depth area.

A second prototype was developed with OpenNI/NITE also generating virtual mouse event (move and clicks) based on the position of the hand joints. In this solution, the right hand of the user moves the mouse cursor in the screen. The cursor position was calculated as the relative position of the user's hand to his torso joint position, so the user can be standing anywhere in the camera angle of view and he can even move around the room during the use.

Also, as we have all the joints coordinates, we can calculate the length of the arms and calibrate the range of the hand movement to fit each user. The left hand is used to activate virtual mouse clicks. By raising the left hand up to

the level of his elbow, a right button press event is generated. When the user returns the left hand down to relaxed position, a left-button release is performed.

This feature allows the user to perform mouse-drag functions like 3D rotations, 2D slices change, as well as regular clicks for buttons and menus (Fig. 2). However, when the user raises the left hand up to the head level, a right button event is generated, enabling the user to perform a right-click drag, which in the case of InVesalius controls the zoom. In order to detect which gesture is being performed, the coordinates of left-hand, left-elbow and head joints are compared, and relative spatial thresholds are applied.

We have validated the proposed solution in 4 tumor enucleations in 3 male patients in whom elective nephron-sparing surgeries were performed for small non-exophitic tumors. Mean tumor length was 2.7 cm (2.1, 2.7, 2.9, and 3.1 cm), and real-time ultrasound was not necessary for intraoperative identification in 3 of 4 endorenal tumors. All surgeries were guided by image review through gesture interface during the procedure, allowing anatomical support based on the 3D imaging previously obtained by computed tomography (CT) scan. All pathological reports revealed renal cell carcinoma, Fuhrman grade I, and negative inked surgical margins. No intra- or postoperative complication was reported.

Although a preliminary experience, while surgeon keeps his arms over the surgical field during the procedure, inadvertent movements are not able to initiate unwanted commands. Otherwise, when surgeon turns to Kinect device, doing the gesturing in order to get the computer to respond properly, movements are always straightforwardly picked up. To learn the routine of the device was very easy, taking less than a few minutes, allowing more than one user to approve the device.
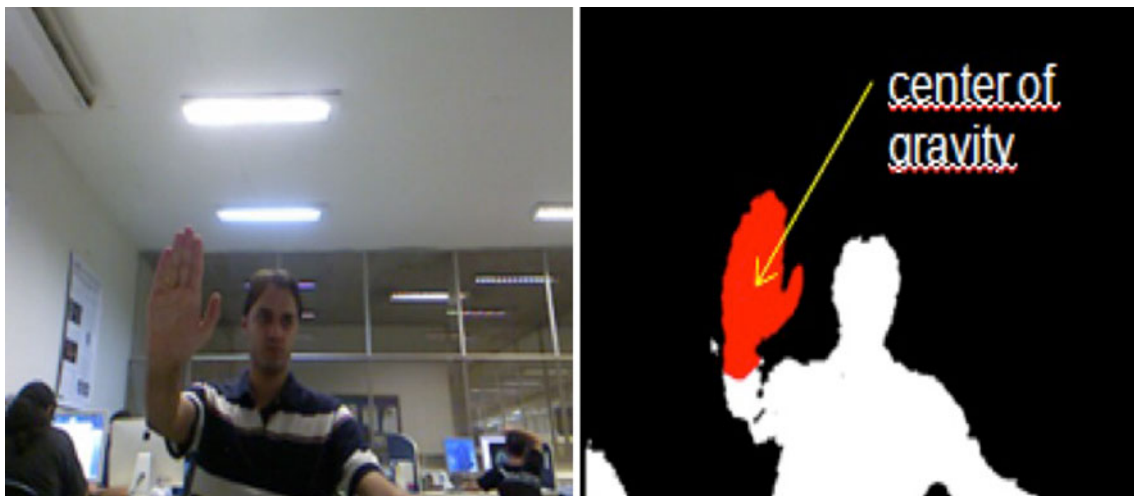

Fig. 1 The color image (*left*) and the result for the threshold in the depth image (*right*), followed by post-processing operations (in *red*) and center of gravity computation

Fig. 2 Touchless gesture interactive image visualization during surgical procedure: InVesalius screenshots showing 2D and 3D images

Discussion

The prototype with Libfreenect revealed to be effective to do tasks like controlling 3D rotation, zoom, slice changes, and contrast adjustments. However, there were two major drawbacks of this solution. The first is that the user has to be located in a specific spot so the threshold can separate the hand from the rest of the scene. The second problem is that the mouse-clicking scheme shares the same hand of mouse moving and that can lead to undesired mouse-moves when trying to perform clicks. Also, this solution is dependent on the arms length of the user, requiring different calibrations.

On the other hand, the prototype using OpenNI/NITE showed to overcome all these problems. Using this approach, the user was able to accurately access the visualization controls of InVesalius. The user experience in this setup was considerably better than the solution presented by Libfreenect because the responsiveness is more stable, does not require the user to be fixed in the same spot, and is much more robust. Also, the ability to track both hands and generate right and left-button clicks with the left-hand revealed to be efficient in the usability and precision of the interface.

Due to the widespread use of imaging technologies and the consequent increasing diagnosis of small renal lesions in young and healthy patients, minimally invasive techniques to perform nephron-sparing surgery have been developed and have been increasingly implemented in an attempt to better outcomes, including tumor enucleation. However, it may result in a greater incidence of positive margins and an increase in intraoperative complications; therefore, additional technological refinements are required to overcome such challenges, incorporating intraoperative real-time image and anatomical support [3].

Although the real-time ultrasonography can be of significant intraoperative assistance in this context, it can be added to and potentiated by the proposed system, especially to distinguish small, hypo-echoic renal tumors from renal scarring. In this setting, viewing intraoperative renal and tumor CT and magnetic resonance imaging (MRI) would provide better and more accurate resection of the renal mass.

Information regarding the position of the kidney, location, and depth of the tumor extension, renal arteries, and veins, and the relationship of the tumor with the collecting system can be intraoperatively viewed and reviewed according to the underlying surgical step.

The presented solution also allows the surgeon to view and manipulate the previously proposed 3D virtual reality model [3] during the surgical procedure while maintaining surgical sterility in an efficient, natural, and non-intrusive human–machine interface, avoiding surgeon distraction from their focus on the procedure. Additionally, no special gloves or sensors placed on surgeon hands underneath the surgical gloves, neither special nor virtual reality environment, are needed as previously suggested [3] and Kinect works in all room lighting conditions, simultaneously tracking multiple users [9].

Furthermore, although having a sterile mouse or tracking device for controlling the computer would be effective in the proposed scenario, it would add unnecessary pieces limiting the surgeon space and movements during the procedure. In this regard, the proposed invention represents the evolution of such devices, avoiding special gloves and/or kinds of trackers/mouses, even if sterile.

The wide impact of such system is not limited to a precise tumor identification, but also helps to prevent wrong site surgeries which are usually the result of a cascade of small errors that in addition to ineffective communication and distractions, deficiencies related to the preoperation and scheduling processes, relies on the limited access to the images in the OR, reducing the available information to those of the surgeon memory, mainly after intraoperative antiseptic measures, when the team is prepared and is not allowed to touch anything but the patient and the surgical instruments, adding to the wellbeing of the patient and the surgical team.

Many developers and research groups around the world are exploring possible applications of Kinect and similar devices that go beyond the original intended purpose of these sensors as home entertainment and video game controllers. In this scenario, this is the first low-cost gesture interface solution applied in the OR evolving Kinect for the field of touchless visualization and analysis of three-dimensional anatomical images (e.g., CT or MRI scans) and opens an avenue for improvements and expansion.

To the best of our knowledge, there is only one Kinect application described in the medical literature; it is regarding physical rehabilitation, just as a motivation for movements but in its original concept for Xbox [10].

Given the rising interest in expanding the application of joystick-free environment, a Microsoft official support for Kinect/PC was released recently (on June/2011), when this work was finished, but it still has important limitations, such as Windows-only compatibility and restriction only for non-commercial use.

Integrating the OpenNI/NITE prototype into InVesalius and exploring different combinations of gestures will provide even easier experiences. Other technologies like fingers tracking, face detections, and speech recognition have been incorporated and are under evaluation. Paralleling the optimization of the presented endeavor, the next steps are heading to its expansion to different fields including a number of aspects of urology such as endourology and minimally invasive urology, as well as a number of other medical/surgical specialties.

Although a step forward in increasing the use of technology toward more efficient utilization of resources in the OR, the further validation of this technique is underway, expanding the number of procedures and patients, accounting for the benefits of this technology (fewer times the surgeon had to scrub out to view images, shorter OR time, lower infection rate, and validated survey results).

Conclusions

For the first time in the literature, we presented a touchless user interface solution to enable a surgeon to control the software InVesalius in an OR just using hand gestures. Kinect device showed to be very efficient and enabled a low-cost and accurate system. Two prototypes were developed using different software approaches, and in our experiments, both succeeded in providing the functionality.

The solution using OpenNI/NITE to track the skeleton and the hands was more accurate and easier to use. Further validation of this approach as a tangible benefit to operative outcomes paralleling its advancement is currently in progress.

References

1. Kadobayashi R, Nishimoto K, Mase K (1998) Design and evaluation of gesture interface of an immersive walk-through application for exploring cyberspace. 3rd IEEE international conference on digital object identifier
2. Keskin C, Aran O, Akarun L (2005) Real time gestural interface for generic applications. European signal processing conference
3. Knudsen BE, Campbell G, Kennedy A et al (2005) Design of functional simulation of renal cancer in virtual reality environments. Urology 66(4):732–735
4. In Vesalius Medical Image Software. http://svn.softwarepublico. gov.br/trac/invesalius
5. Kinect Official Web Site. http://www.xbox.com/kinect/
6. Kinect (Wikipedia, 21 June 2011 22:26). http://en.wikipedia. org/w/index.php?title=Kinect&oldid=435540952
7. Agarwal A, Triggs B (2006) Recovering 3D human pose from monocular images. IEEE Trans Pattern Anal Mach Intell 28(1):44–58. doi:10.1109/TPAMI.2006.21. http://ieeexplore.ieee. org/stamp/stamp.jsp?tp=&arnumber=1542030&isnumber=32932
8. Gonzalez R, Woods R (2007) Digital image processing, 3rd edn. Prentice-hall, New York
9. Product Technology: The PrimeSensor™ Technology. http:// www.primesense.com/?p=487
10. Chang YJ, Chen SF, Huang JD (2011) A kinect-based system for physical rehabilitation: a pilot study for young adults with motor disabilities. Res Dev Disabil 32(6):2566–2570