Multi-modal Controller for Image Manipulation in the Operating Room

Alexandre Sierro Pierre-André Mudry
University of Applied Sciences Western Switzerland
HES-SO Valais
Rte du Rawyl 47
1950 Sion, CH - Switzerland
{pierre-andre.mudry, alexandre.sierro}@hevs.ch

Abstract

In the domain of orthopedics, surgeons often rely on radiology images during operations. In this context, manipulating the images displayed on a computer screen is an issue as their hands have to remain sterile. In this article, we present a multi-modal controller (foot and voice) coupled with an existing state-of-the-art radiology display and analysis software used in operating rooms. The controller itself consists of a battery-operated wireless embedded system integrated into a shoe that provides multiple foot pressure-points as well as an absolute orientation sensors. In addition, a wireless microphone is used to acquire voice commands.

To demonstrate the validity of our approach, we present a randomized user study conducted on ten subjects that had to perform image manipulation tasks using the controller.

Keywords

Foot-based controller, inertial measurement unit, voice control, orthopedics, surgery, radiology imagery.

1 INTRODUCTION

During orthopedic operations, surgeons often rely on existing radiology images (X-ray, MRI, …). Displayed on computer screens, those images are often manipulated with a mouse by operating room (OR) assistants as the surgeon's hand are often busy manipulating the patient. Another reason behind those assisted manipulations are sterility issues related to hand-based controllers (keyboards or mice for instance).

In this paper, we present a multi-modal controller based on voice and foot input for radiology image manipulation during surgery. The advantages of this approach are two-fold: first, the advantage of sterility and hand-free operation and, second, the independence in the positioning of the surgeon towards the input device.

1.1 Paper Organization

We proceed as follows: in the next section, a brief overview of existing human-machine interaction methods in the operating room is presented. After that, focus is put on the hardware and software implementation of the controller and the means of interfacing it with a standard, PC-based, radiology image manipulation program. We then present a randomized experimental setup to demonstrate the strengths and weaknesses of the applied approach before concluding.

2 RELATED WORK

High-sterility and non-encumbered interaction are paramount in the OR. For this reason, camera-based approaches tracking surgeon gesture have been successfully applied in the past ([1, 2]). However, one major difficulty with this technique is the proper detection of gestures which still remains a challenge today ([3]). To improve the situation, researchers have demonstrated that integrating the third dimension can be useful (for instance by using a *Kinect* device [4], or *Leap Motion device* [5]). However, using an imaging device requires the surgeon to be positioned at a precise location in the OR.

To partially circumvent this limitation, voice commands can be added to the setup in order to perform some control (see for instance [6, 7]) when not in the field-of-view of the imaging device.

In the last decade, developments in the field of micro electronic mechanical systems (MEMS) enabled the production of cheap and reliable orientation sensors. Of particular interest is the appearance of devices integrating a fusion of accelerometers, gyroscope and geomagnetic sensors which can be used to extract hand or foot movements of a user. This enabled the creation of position-capturing devices which can be used in gaming or control (for instance as described in [8, 9]) and that we will be using in our multi-modal controller to capture foot-orientation information.

3 IMPLEMENTATION

Our multi-modal controller is based on three different sources of information: foot pressure points distribution, foot gestures as well as voice commands. Foot-based information is captured via a dedicated embedded system which has been integrated into the sole of a shoe.

3.1 Architecture

As depicted in Fig. 1, foot sensor information is wirelessly transmitted to a control software which integrates this information thanks to a very efficent open-source voice recognition package called *Sphinx*[1].

Once the proper pointing method (see 3.3) has been selected, the appropriate commands are then generated and sent via telnet to *Weasis DICOM Viewer*[2], an open-source radiology image manipulation program..

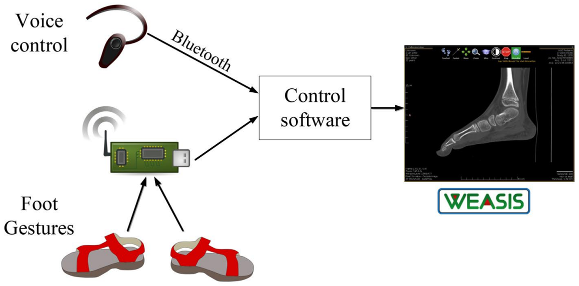

Figure 1: System architecture

3.2 Hardware Implementation of the Foot-based Controller

The foot-based controller embedded system (Fig. 2) contains four main components articulated around an ARM Cortex M0 micro-controller:

- **Pressure sensors** – Foot pressure is measured at three different locations using resistive load cells from *Alpha Electronics*. The resultant resistance is converted into a voltage and then digitized using the micro-controller's analog to digital converter.

- An **inertial measurement unit** (IMU) – The exact model is `BNO055`, which is a module already containing the required sensor fusion algorithms to provide fast and accurate readings of absolute orientations extracted from 9 axes : 3 axes accelerometer, 3 axes gyroscope and 3 axes magnetometer.

- A low-power **wireless communication** chip – Active in the 2.4 GHz range, the `NRF24L01+` chip is connected with an antenna directly printed on the printed-circuit board.

- An autonomous **power supply** which consists in a 3.7 V, 850 mAh Lithium-Polymer battery charged using either with a micro-USB connector or an inductive charger.

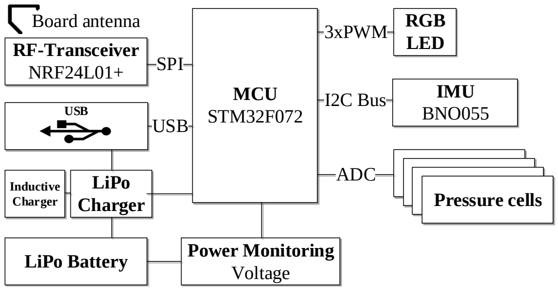

Figure 2: Hardware architecture

The embedded system is integrated into a standard sandal, which can be seen in Fig. 3. The system can be used using one or two shoes, depending on the selected interaction mode, as we will discuss in the next section.

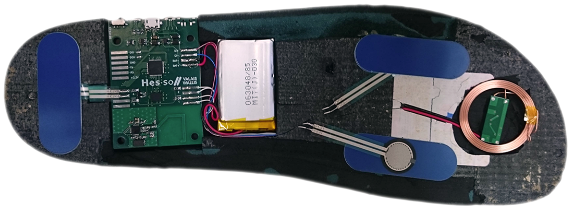

Figure 3: Shoe integration (from left to right) : rear pressure sensor, system board, left and right pressure sensors and wireless charging receiver.

3.3 Software implementation

The visualization software we used is tailored to be used with a standard mouse input. Early tests showed that a direct translation from foot gestures to mouse commands is not feasible. In fact, a clicking gesture with the foot can be very tiresome and therefore a different selection mechanism based on voice commands was chosen.

The valid actions implemented in the context of this project are : *move*, *zoom*, *contrast* and *slice*. For this last point, it is worth noting that radiology data might be three-dimensional and therefore it is possible to navigate into the "depth"of the radiology image by changing the actual slice of the data.

To select between those different actions and interact with the software, three interaction strategies have been implemented:

---

[1] http://cmusphinx.sourceforge.net/

[2] https://github.com/nroduit/Weasis

1. **Voice** method, which lets the user choose between the different actions using voice commands. In this mode, inclination of the main foot, measured by IMU, only acts on the selected action.

2. **Fusion** method, which combines pressure and inclination of the main foot to act simultaneously on movement and magnifying. In this mode, pressure applied on the tip of the foot will zoom-in and pressure on the back will zoom out.

3. **Two feet** method, which uses the main foot inclination to move the picture and the second foot to control magnifying.

In every strategy, voice commands can be used to cancel the current action or reset the visualization to a known state. As depicted in Fig. 4, a control panel using an icon-based UI appears as an overlay in front of *Weasis*, displays valid commands and provides a feedback of the currently selected mode.

Figure 4: The method selector UI which is displayed atop *Weasis*.

3.3.1 Acquisition process

Raw data output from the sensors is converted to valid user inputs with a relatively simple software on the PC. The conversion algorithms starts by applying specific thresholds and gains to each sensor and then their values are routed to a specific *Weasis* command according to the currently selected pointing method.

In order to improve the user experience of the system and increase its controllability, a profile containing threshold values and main foot selection is generated for each end-user. This profile enables the integration of taste-specific values into the controller and act as calibration for the system.

4 EXPERIMENTAL SETUP

To demonstrate the validity of the approach and to provide a user-based feedback on the multi-modal controller, we implemented an experimental setup reproducing a typical OR scenario.

During the experiment, the three aforementioned pointing methods were evaluated as well as the standard mouse control which serves as a reference.

To test the setup, ten persons were presented with the experimental setup and the detection thresholds of the foot-controller were adjusted to their taste.

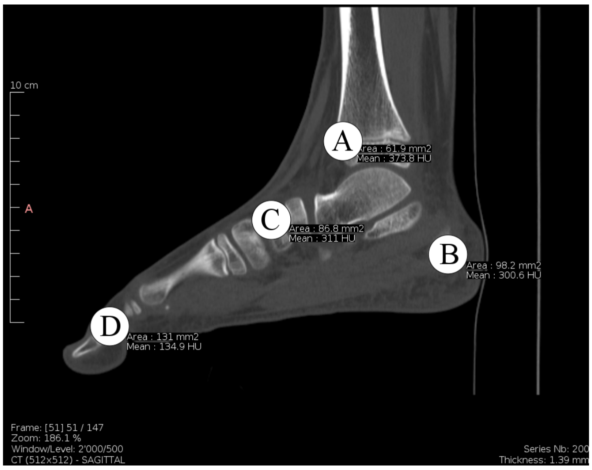

Figure 5: Points of interest that had to be zoomed to.

4.1 User Objectives

For the experiment, the users had to zoom on four points in an X-ray image following a specific order (highlighted as ABCD on Fig. 5).

Measurements are performed twice with different complexities: the first time the user has to zoom-in then zoom-out for every point whereas the second time only the first point has to be zoomed-in before moving over the other points. These complexity levels are labeled respectively *Z+M* and *M*.

In both cases, the time to reach the first point and the subsequent transitions times are measured. Before each measure, users had time for practice. At the end of experiment, users also had to rate their satisfaction level for each of the pointing method in terms of accuracy, speed and usability. The marks given could vary between 1 (not satisfied) and 4 (very satisfied).

4.2 Results

Fig. 6 depicts the pointing duration for the various strategies and task complexities.

When considering interaction speed, a first result that can be extracted is that the reference mouse method is on average 2.5 times faster than any other method. The voice method is the slowest strategy for aggregated zoom and movement ; this can be explained by the fact that changing from one mode to the other requires voice commands. However, as zoom and move commands are clearly separated, this reduces interferences in the movement and allows more accurate movements for the M complexity.

Overall, the fusion method seems to be the most appropriate to achieve a reasonable speed for most users using this setup. Unfortunately, direct speed comparisons with other input techniques are difficult as use-case scenarios differ too much.

From a user evaluation standpoint, Fig. 7 shows how the various methods were evaluated in terms of usability,

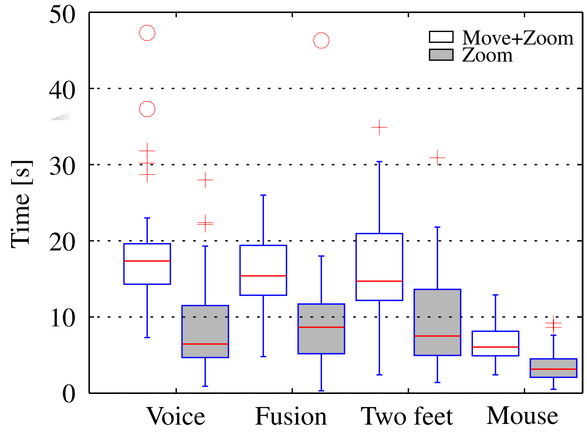

Figure 6: Duration analysis for the different methods and task complexities.

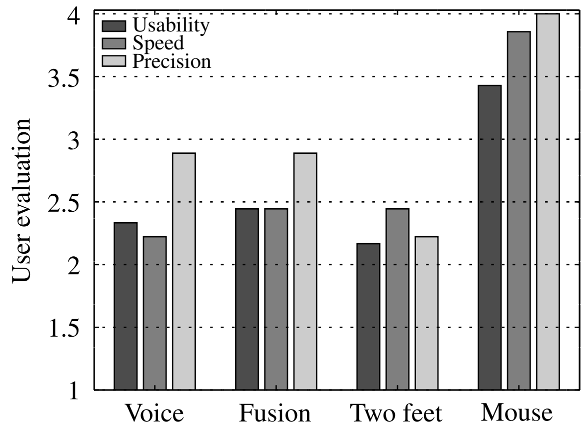

Figure 7: User evaluation (average) of the different methods.

speed and precision. Analyzing this data reveals that test impressions correspond to time measurements, i.e. voice method has a good precision but is slower, two feet is the fastest method but is less precise. For those experiments the reference mouse method is still preferred.

5 CONCLUSION

We showed in this article how a multi-modal controller can be successfully used to provide a robust HMI in the context of an OR. Even if users seem to favor a mouse as input device for image manipulation, we showed that mixing voice commands with foot gestures provides both accuracy and speed whilst preserving sterility and position independence for the surgeon.

Further work will include testing the multi-modal controller in a real OR scenario to adjust the system to real-world constraints and integrate feedback from surgeons.

6 ACKNOWLEDGEMENTS

This research, part of the *Lunamed* research projet, was funded by an UAS Western Switzerand (HES-SO) internal grant.

7 REFERENCES

[1] M. Ma, P. Fallavollita, S. Habert, S. Weidert, and N. Navab, "Device- and system-independent personal touchless user interface for operating rooms," *International Journal of Computer Assisted Radiology and Surgery*, pp. 1–9, 2016.

[2] G. C. S. Ruppert, L. O. Reis, P. H. J. Amorim, T. F. de Moraes, and J. V. L. da Silva, "Touchless gesture user interface for interactive image visualization in urological surgery," *World journal of urology*, vol. 30, no. 5, pp. 687–691, 2012.

[3] T. Kopinski and U. Handmann, "Touchless interaction for future mobile applications," in *International Conference on Computing, Networking and Communications (ICNC)*, pp. 1–6, February 2016.

[4] M. Strickland, J. Tremaine, G. Brigley, and C. Law, "Using a depth-sensing infrared camera system to access and manipulate medical imaging from within the sterile operating field," *Canadian journal of surgery. Journal canadien de chirurgie*, vol. 56, pp. E1–6, June 2013.

[5] A. Zocco, M. D. Zocco, A. Greco, S. Livatino, and L. T. De Paolis in *Proc. of the 2nd Int. Conference on Augmented and Virtual Reality (AVR2015)*, pp. 432–445, Springer, 2015.

[6] A. M. Hötker, M. B. Pitton, P. Mildenberger, and C. Düber, "Speech and motion control for interventional radiology: requirements and feasibility," *International Journal of Computer Assisted Radiology and Surgery*, vol. 8, no. 6, pp. 997–1002, 2013.

[7] Y. Kim, S. Leonard, A. Shademan, A. Krieger, and P. C. W. Kim, "Kinect technology for hand tracking control of surgical robots: technical and surgical skill comparison to current robotic masters," *Surgical Endoscopy*, vol. 28, no. 6, pp. 1993–2000, 2014.

[8] A. Gams and P.-A. Mudry, "Gaming controllers for research robots: controlling a humanoid robot using a WIIMOTE," in *Proc. of the 17th Int. Electrotechnical and Computer Science Conference (ERK08)*, pp. 191–194, 2008.

[9] K.-B. Cho and B.-H. Lee, "Intelligent lead: a novel HRI sensor for guide robots," *Sensors*, vol. 12, no. 6, p. 8301, 2012.