Touchfree Medical Interfaces

Nathaniel Rossol[1], Irene Cheng[1], Rui Shen[1], and Anup Basu[1]

***Abstract*— Real-time control of visual display systems via mid-air hand gestures offers many advantages over traditional interaction modalities. In medicine, for example, it allows a practitioner to adjust display values, e.g. contrast or zoom, on a medical visualization interface without the need to re-sterilize the interface. However, when users are holding a small tool (such as a pen, surgical needle, or computer stylus) the need to constantly put the tool down in order to make hand gesture interactions is not ideal. This work presents a novel interface that automatically adjusts for gesturing with hands and hand-held tools to precisely control medical displays. The novelty of our interface is that it uses a single set of gestures designed to be equally effective for fingers and hand-held tools without using markers. This type of interface was previously not feasible with low-resolution depth sensors such as Kinect, but is now achieved by using the recently released Leap Motion controller. Our interface is validated through a user study on a group of people given the task of adjusting parameters on a medical image.**

I. INTRODUCTION

Markerless, mid-air hand gestures tracked by computer vision techniques have become increasingly popular for human-computer interaction. Prior work has shown that their non-contact nature makes them especially well-suited for controlling visual displays at a distance [9]. They are also useful for controlling medical visualization interfaces where the lack of physical touch improves safety by avoiding possible contamination [1].

Some of the challenges when users are performing tasks that require them to hold a tool in their hand(s) such as a pen, computer stylus, laser pointer, or a medical instrument, are that it can interfere with finger tracking or make it awkward for the user to perform gestures in a way that can be recognized. For example, a clinician might be holding an ultrasound probe in one hand, and a biopsy needle in the other during a procedure, but would still like to control their real-time medical display. As another example, an artist might be using a computer drawing tablet with a stylus but would still prefer to make display interactions without switching to their non-dominant hand. Being forced to constantly put a tool down causes an undesirable interruption in workflow.

To address the challenges above, we propose a novel interface design for controlling displays that enables mid-air gestures from hands and also hand-held tools. In our approach, hand-held tools can automatically become gesture devices that the user may use to help control the display as shown in Figure 1. Previously, our proposed interface design would not be practical with low resolution commercial depth sensors, such as the Kinect [7], but is achievable in this work using the Leap Motion sensor's [8] higher resolution and sampling rate.

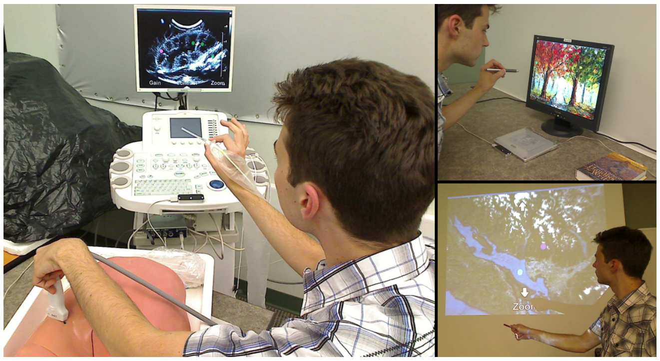


Fig. 1. Several applications of our approach. Left: The user controls an ultrasound display without touching it, using a needle to gesture. Top Right: A computer stylus is lifted off its drawing table and used to adjust the display without needing to switch modes. Bottom Right: During a presentation, an ordinary laser pointer is used to interact with the display.

The contributions of our work are:

- An interface design that uses gestures specifically designed to be equally efficient with bare hands or hand-held tools.

- Unique filtering rules introduced to help prevent unintentional gestures caused by hand movements while holding tools.

- Developing an application for touch-free control of an interface coupled with an Ultrasound display.

- The ability to operate in both bimanual, and unimanual modes, without the need for context switching, or the use of markers on hands or tools.

Finally, to evaluate the effectiveness of our approach, a prototype medical display interface was implemented that uses our interaction framework. A 3-stage user study conducted on our implementation revealed that users were capable of using either their hands or pointed tools to perform the gestures with no major loss in task performance times being observed. Applications of our work include touch-free interfaces for Ultrasound devices, where a technician does not need to worry about having one hand free of gel for interacting with the device or a keyboard, or surgical environments where a dedicated opertor (in addition to surgeons) is not needed for visualizing surgery related images.

The remainder of this paper is organized as follows: Section 2 summarizes related work on interacting with displays

---

[1] The authors are with the Department of Computer Science, University of Alberta, Edmonton, Alberta T6G 2E8, Canada. Emails: nrossol@ualberta.ca, locheng@ualberta.ca, rshen@ualberta.ca, basu@ualberta.ca

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

using gestures. The proposed approach and implementation details are described in Section 3. The evaluation of results is presented in Section 4. Finally, concluding remarks are given in Section 5.

II. RELATED WORK

In 2003, the VisionWand [2] was proposed as a simple, low-cost tool for interacting with displays. The wand itself was simply a colored plastic rod that contained no electronics and was tracked in 3D by a pair of calibrated color cameras as the user performed 3D mid-air gestures. In 2007, Guo et al. [6] proposed a wand with similar features but with more degrees of freedom. However, in both of these approaches, only pre-calibrated marked tools were considered for interaction, and the tool gestures were considered in a completely separate context from hand gestures.

In 2005, Vogel et al. [12] investigated unimanual hand gestures for highly precise manipulations on a large high-resolution display. Their work identified that small quick finger gestures, such as tapping in mid-air (called an "air tap"), were highly intuitive for users and could be used for efficient and precise context switching. Similar findings were also confirmed by Fikkert el al. in 2010 [3] in their interactive map display. These small, precise finger gestures provide a good solution to the common issue of determining when to begin and end tracking of a user's hand for manipulation tasks, which is known as "gesture spotting" [13]. However, Vogel's system requires a costly, pre-calibrated motion capture rig and needs infrared reflectors placed on users' hands in order to achieve its precise tracking. Therefore, the system is designed to track only one hand at a time and does not consider the presence of unmarked hand-held tools. Techniques using data gloves or other sensors placed on the hand are similarly limited [4].

To overcome the limitations of markers, several recent approaches have made use of low-cost commercial depth sensors, such as the Microsoft Kinect, to interact with displays via hand gestures. Gallo et al. [5] proposed a Kinect-based interface in 2011 for visualizing medical images in a sterile surgery room requiring the user to be standing, and using both hands for interaction. Song et al. [11] proposed a novel 2-handed 3D gesture interface in 2012 for making precise manipulations using a handle-bar metaphor, and similarly, Schwaller et al. [10] used a 2-handed approach for panning and zooming displays. These techniques achieve their goals by using different, asymmetrical static hand poses on each hand for gesture spotting and context switching. This overcomes the inability of modern commercial depth sensors to precisely track dynamic finger gestures in 3D, because of noise and low resolution depth images. This approach also takes advantage of the fact that 2-handed gestures are often preferred by users for interacting with displays, and can allow for tasks to be completed quicker, as was demonstrated by Nancel et al. [9]. Users may not always have both hands free for interaction depending on the application domain, but it is ideal for the option to be available if possible. The weakness of this class of methods is the low accuracy which makes dynamic finger tracking impossible.

In contrast to the above approaches, our method: (i) is completely markerless while accommodating the use of tools, (ii) can track dynamic fingers and tools, thereby gestures like tapping, by exploiting the higher precision of the Leap Motion Sensor and novel filtering rules.

III. PROPOSED APPROACH AND IMPLEMENTATION

To test the efficacy of our design, we chose to apply it to the challenge of controlling an ultrasound machine's display parameters during procedures. A design was made in collaboration with our industrial partner who manufactured and provided the test machine. The parameters chosen were the ones identified as most frequently adjusted during procedures: Gain, Zoom, and Contrast. The gesture interpretation system was implemented on a dedicated low cost commodity laptop for mobility purposes (Intel x86 dual core processor at 2.2 GHz, 2 GB RAM), while display parameter adjustments were computed asynchronously and then sent to the display of an Sonix RP ultrasound machine. The sensor used was a low cost commercial LEAP Motion controller which uses small IR (Infrared Red) LEDs and a pair of IR cameras to track unmarked objects with millimeter level precision at over 60 frames per second in an interaction volume of about 1 cubic meter. The LEAP Motion SDK (the only such API available) was used to track the the positions of finger tips, hands, and tool tips.

The API restricts the tools that can be tracked to have dimensions similar to that of a pen with a thickness between approximately 30mm and 3mm, which still covers a large range of hand-held tools (pens, markers, styluses, needles, probes, laser pointers, etc.) without any pre-calibration required. For larger tools, our interface allows users to still hold it in their palm and use free fingers to gesture, as the hand tracking is usually robust enough to remain stable in this state. Visual feedback lets users know what is being tracked as shown in Figure 2.

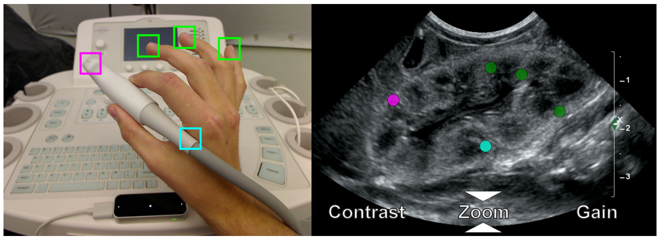


Fig. 2. We use green indicators to identify tracked fingertips. Blue is used to indicate the palm position and magenta for tool tips, when recognized. Note that the coloured boxes on the left image correspond to the coloured circles on the right image, which is what is actually displayed to the user.

A. *Gesture Interaction Design*

To minimize cognitive load, our system uses a vocabulary of 3 communicative gestures for finger tips or tool tips, and

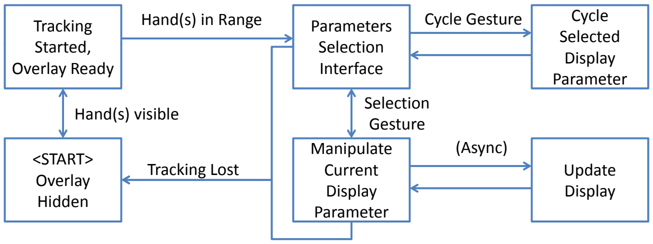


Fig. 3. State transition diagram. Note that the Cycle Selected, and Update Display states loop back automatically, whereas the Selection Gesture is needed to toggle between the Parameters Selection and Manipulation states.

1 manipulative gesture for hands. A state transition diagram outlining the operation of the system is shown in Figure 3.

In its initial state, the interface overlay is hidden, and only the display is shown. As users bring their hands closer to the interaction volume, the system begins to track their hands, fingers and tools, and a partial overlay is shown as well as tracking indicators (represented as small colored circles) to provide visual feedback. When at least one hand is fully in the interaction volume, the parameter selection interface is shown and the currently highlighted display parameter is indicated. Drawing a small clockwise or counter-clockwise circle in the air (with a radius as small as 10 mm in size) will either cycle the currently highlighted parameter forward or backward through the list. To actually manipulate the currently highlighted parameter, the user performs a mid-air tap as if they were clicking a mouse. In parameter manipulation mode, the system uses the Y position (height) of the hand closest to the monitor to raise or lower the selected display value, which is updated on the display asynchronously in a separate thread to preserve system responsiveness. A tap performed by any finger tip or tool tip exits the adjustment mode and returns the user to the parameter selection state. Figure 4 shows gesturing possibilities in several hand configurations.

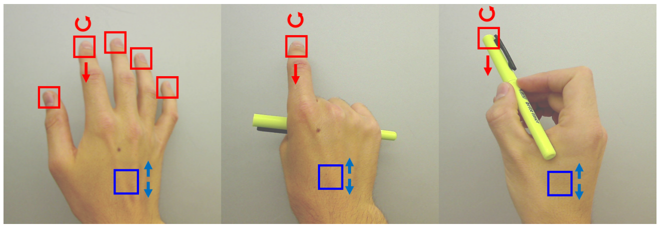


Fig. 4. Left: Any finger may be used to perform mid-air tap gestures or circle gestures (shown in red). Middle: Only the index finger movements will be tracked for circle and tap gestures, the tool is hidden. Right: The tip of the felt pen is tracked for gestures while the remaining fingers are not, being closed. In all cases, the hand palm position (blue) is tracked independently, and may be moved for manipulative gestures.

Because the system does not restrict the tooltip or fingertip gestures to the hand being used for parameter manipulation (i.e., the hand closest to the display) highly precise adjustments can be made by using tapping gestures on one hand (with a tool or finger) while using the closer hand for parameter adjustment. For one-handed gesturing, the largely independent tracking of the palm and fingertips or tooltip tracking still allows for highly precise manipulations.

B. *Rule Based Filtering of Gestures*

A drawback of our approach described thus far is that unintentional fingertip or tooltip movements on one or both hands may be tracked and interpreted as input. Therefore, a set of custom rule-based filtering techniques were implemented to reduce unintentional gestures. For circle gestures, as soon as a successful circle is recognized (within radius requirements of about 10mm - 80mm and time < 1.0s), the time window of the gesture is established. Any other gestures with a time windows that overlaps a successful gesture are suppressed. This avoids the common issue of users accidentally moving multiple fingers in a similar manner while performing gestures. Quick gestures with a small time window, such as tapping, require a secondary time window of fixed size to be imposed to artificially suppress multiple fingers or tools unintentionally being involved in a tap.

A unique drawback of the chosen Leap sensor is that it occasionally causes partially occluded fingertips or tooltips to rapidly vibrate or jump in and out of a tracked state. These erroneously tracked 3D positions can sometimes be interpreted as a gesture, and thus must be recognized and suppressed until the tracked point displays more stable behavior.

IV. EVALUATION

To evaluate our interaction design, a user study was conducted in which users were required to perform a list of 6 arbitrarily chosen parameter adjustments (to Gain, Zoom, and Contrast) on our ultrasound machine's display interface using only mid-air hand gestures. Participants were first instructed on how the interface worked, and were made to practice 10 successfully tracked circle gestures and 10 tapping gestures with their fingers before beginning. The participants repeated the exact same task list using their dominant hand in 3 different configurations:

- Open Hand: The hand is empty and completely open in a relaxed state. All fingers are tracked in this configuration and users are allowed to use any finger to perform the gestures.

- Pointing Finger: The empty hand is closed except for the index finger which is pointing forwards. This pose mimics holding a tool in the hand but not gesturing with it.

- Tool in Hand: For safety, a pencil was chosen as the tool to gesture with for the experiment rather than a piece of medical equipment. Users were instructed to hold it as naturally as possible while leaving at least 30mm protruding at the tip.

The list of arbitrary parameter adjustment tasks was designed to take about 1 minute to complete. Upon completing the task in all 3 modes (Trial 1), the users were required to repeat the entire round of tasks two more times (Trials 2 and 3). The experiment was conducted with 12 participants (4 female, 8 male) between the ages of 20 and 55. After completing the experiment, users completed a 5-point Likert scale questionnaire.

A. *Results and Discussion*

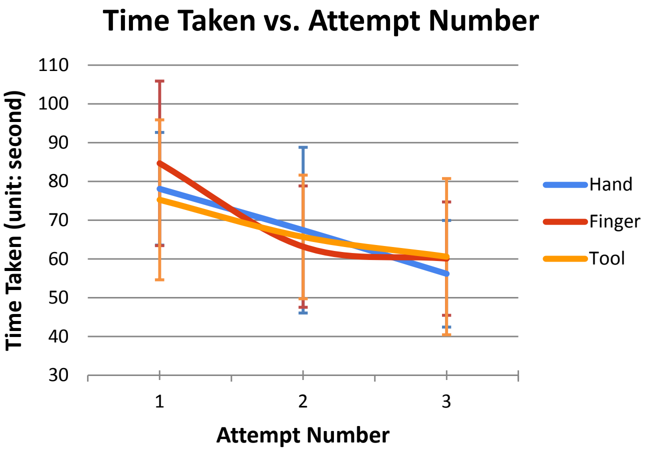
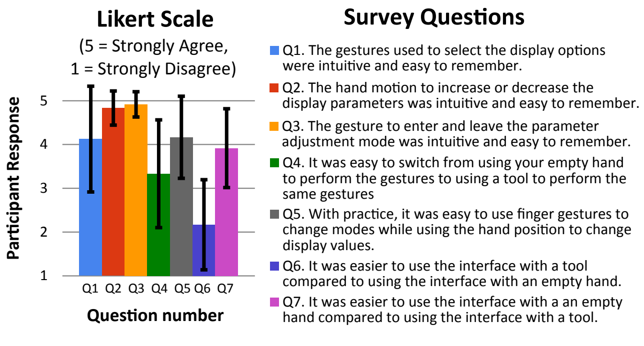


Fig. 5. Comparison of Time taken vs. Attempt number.

The task completion times for all 3 trials were recorded and are shown in Figure 5. The results of the Questionnaire are summarized in Figure 6. The error bars in both graphs indicate the standard error. A two-way repeated measures ANOVA test examining the effect of trial round on completion time rejects the null hypothesis at the 5% significant level (p-value: $0.00004 < 0.05$) which indicates that the user completion time varies from one trial round to another. A two-way repeated ANOVA examining the effect of hand configuration on completion time does not reject the null hypothesis that the three interaction modes do not significantly differ from each other in this experiment with respect to user completion time at the 5% significance level (p-value: $0.85158 > 0.05$).

These results demonstrate that users can significantly improve their performance with practice over a small number repeated uses. The results also seem to suggest that there is no major difference in observed user performance with respect to the 3 different hand configuration modes. This is encouraging because it suggests that there is no exceptional loss in performance when gesturing with a tool instead of putting it down. However, due to the high variance measured, more user studies need to be conducted in the future to test for more subtle differences.



🟦 Q1. The gestures used to select the display options were intuitive and easy to remember.
🟥 Q2. The hand motion to increase or decrease the display parameters was intuitive and easy to remember.
🟧 Q3. The gesture to enter and leave the parameter adjustment mode was intuitive and easy to remember.
🟩 Q4. It was easy to switch from using your empty hand to perform the gestures to using a tool to perform the same gestures
⬛ Q5. With practice, it was easy to use finger gestures to change modes while using the hand position to change display values.
🟪 Q6. It was easier to use the interface with a tool compared to using the interface with an empty hand.
🟪 Q7. It was easier to use the interface with a an empty hand compared to using the interface with a tool.

Fig. 6. Results of Questionnaire.

The questionnaire revealed that most users generally found the gesture interface design to be intuitive and easy to remember, although they expressed a preference for interacting with an empty hand rather than a tool, though their performance between the two were similar. However, this peculiar result is consistent with results from [9] who found a similar pattern, and suggests that user opinion may change over time as they become more familiar with using tools for gesturing.

V. CONCLUSION AND FUTURE WORK

In this work we presented a novel, markerless gesture interface using a leap motion sensor and robust filtering rules. Our approach seamlessly integrates gesturing with hands and unmarked tools in a single interface for precise control of display parameters via mid-air gestures. Application areas include controlling medical displays, or more streamlined interaction when using devices like computer styluses and tablets. The results of our user study suggested that users can quickly make a significant improvement on their performance with practice, and no exceptional loss in performance between different hand configurations was observed. In future work we will compare alternative filtering rules for robust detection of gestures. Furthermore, we will study the benefits of gesture based interfaces for ultrasonic devices and computer assisted surgery.

VI. ACKNOWLEDGMENTS

The authors like to thank Alberta Innovates (iCORE) and NSERC for their financial support.

REFERENCES

[1] Invisible touchcontrol of a dicom viewer with finger gestures using the kinect depth camera. *Journal of Forensic Radiology and Imaging 1*, 1 (2013), 10 – 14.

[2] Cao, X., and Balakrishnan, R. Visionwand: interaction techniques for large displays using a passive wand tracked in 3d. In *UIST* (2003), 173–182.

[3] Fikkert, W., Vet, P., Veer, G., and Nijholt, A. Gestures for large display control. In *Gesture in Embodied Communication and Human-Computer Interaction*, vol. 5934 of *LNCS*. 2010, 245–256.

[4] Gallo, L., and Ciampi, M. Wii remote-enhanced hand-computer interaction for 3d medical image analysis. In *Intl. Conf. Current Trends in Information Technology (CTIT)* (2009), 1–6.

[5] Gallo, L., Placitelli, A., and Ciampi, M. Controller-free exploration of medical image data: Experiencing the kinect. In *24th Intl. Symp. Computer-Based Medical Systems (CBMS)* (2011), 1–6.

[6] Guo, F., Kimber, D., and Rieffel, E. Featured wand for 3d interaction. In *ICME* (2007), 2230–2233.

[7] Microsoft Kinect. `http://www.xbox.com/en-US/kinect/`.

[8] LEAP Motion. `https://www.leapmotion.com/`.

[9] Nancel, M., Wagner, J., Pietriga, E., Chapuis, O., and Mackay, W. Mid-air pan-and-zoom on wall-sized displays. In *CHI* (2011), 177–186.

[10] Schwaller, M., Brunner, S., and Lalanne, D. Two handed mid-air gestural hci: Point + command. In *Human-Computer Interaction. Interaction Modalities and Techniques*, vol. 8007 of *LNCS*. 2013, 388–397.

[11] Song, P., Goh, W. B., Hutama, W., Fu, C.-W., and Liu, X. A handle bar metaphor for virtual object manipulation with mid-air interaction. In *CHI* (2012), 1297–1306.

[12] Vogel, D., and Balakrishnan, R. Distant freehand pointing and clicking on very large, high resolution displays. In *UIST* (2005), 33–42.

[13] Wachs, J. P., Kölsch, M., Stern, H., and Edan, Y. Vision-based hand-gesture applications. *Commun. ACM 54*, 2 (2011), 60–71.