

Range and variability in gesture-based interactions with medical images : do non-stereo versus stereo visualizations elicit different types of gestures?

Citation for published version (APA):

Beurden, van, M. H. P. H., & IJsselsteijn, W. A. (2010). *Range and variability in gesture-based interactions with medical images : do non-stereo versus stereo visualizations elicit different types of gestures?*. Paper presented at conference; IEEE virtual reality 2010 : workshop on medical virtual environments; 2010-03-21; 2010-03-21.

Document status and date:

Published: 01/01/2010

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Range and variability in gesture-based interactions with medical images: Do non-stereo versus stereo visualizations elicit different types of gestures?

Maurice van Beurden & Wijnand IJsselstein
Human-Technology Interaction Group
Eindhoven University of Technology
PO Box 513, 5600 MB Eindhoven, The Netherlands
{m.h.p.h.v.beurden; w.a.ijsselstein}@tue.nl

ABSTRACT

The current paper presents a study into the range and variability of natural gestures when interacting with medical images, using traditional non stereo and stereoscopic modes of presentation. The results have implications for the design of computer-vision algorithms developed to support natural gesture-based interactions in a medical context.

KEYWORDS: gesture-based interaction, image manipulation, medical imaging, stereoscopic 3D.

INDEX TERMS: H.1.2 [Information systems]: User/Machine systems —Human Factors; H.5.2 [Information, Interfaces and Presentation (e.g.,HCI)] User Interfaces— Interaction Styles.

1 INTRODUCTION

Information technology is increasingly deployed in hospital settings, in particular in areas of patient information management (e.g., the Electronic Medical Record, or EMR) and pre-operative planning. This enables easy and ubiquitous access to accurate and up-to-date clinical information, increasing a hospital's efficiency and quality of care. In the operating room (OR), however, the use of digital information has been relatively limited, due to a demanding set of human-machine interface (HMI) requirements imposed by the task environment. First, in OR procedures, time is at a premium. Therefore, access to medical data needs to be easy and fast. Second, the surgeon's attentional focus should be on the patient, and not on the interaction technology nor on instructing assistants to interact with the technology on their behalf, which is a slow and error-prone process (see e.g., [1]). Again, this stresses the need for accurate and intuitive HMI technologies in the OR. Third, sterility of the working environment needs to be guaranteed. Traditional HMI technologies, such as mouse and keyboard, are difficult to sterilize and are known to be potential sources of infection [2]. Taken

together, these requirements specify an HMI that is fast, accurate, intuitive, and does not require hands-on interaction.

The field of Human-Computer Interaction (HCI) is progressing fast in developing HMI technologies that build on natural modes of human communication and interaction, including speech, hand- and body gestures, facial expressions, and gaze. Several of these HMI technologies are being explored for their potential in the demanding OR environment. Examples include the Face MOUSE [3], a laparoscopic positioning system controlled by the surgeon using face gestures, the Non-Contact Mouse [1], where surgeons interact with endoscopic images while using a well-defined set of gestures to perform standard mouse functions (pointer movement and button presses), and Gestix [4] another hand gesture-based system for browsing medical images from an EMR database. These systems invariably aim to support intuitive interactions, yet require surgeons to use a relatively well-specified and limited set of gestures, in order for the computer vision algorithms to achieve an acceptable rate of recognition.

What has remained unexplored to date is the range and individual variability of gestures used when unconstrained interactions are allowed. This is the purpose of the current research. Moreover, we hypothesize that differences may exist in gesturing behavior when medical images are presented on a traditional non-stereo display as compared to a stereoscopic 3D display. In particular, we expect that images that are not displayed in stereo, to elicit gestures that are more in line with traditional interactions using a desktop metaphor (e.g., point and click, double-click). In contrast, we expect stereoscopic 3D to elicit gestures that are more spatial in nature. This includes an increase in two-handed interactions. Previous work [5] has shown that people often express spatial manipulations using two-handed gestures. To address this question in our experiment, we will include both non-stereo

and stereoscopic modes of visualization. Participants were asked to demonstrate the gestures used to perform various manipulation tasks. Since the purpose of the experiment was to find out the range and variability of gestures participants would naturally use, we did not provide any *a priori* constraints or requirements on the gestures participants were allowed to perform. As implementing one specific set of gesture recognition algorithms would have constrained the expressive ability of participants, we chose to use a set of static images, rather than interactive ones, throughout the experiment.

2 METHOD

2.1 Design

In this explorative study we manipulated the presentation mode (non stereo vs. stereo) in a between- subjects experimental design. Within each presentation mode participants performed seven different tasks; positioning, selecting, activating, rotating, zooming in, zooming out, and deactivating. The tasks are performed using four types of content; three medical images, and an overview image where the three medical images are arranged vertically. The dependent variable was the type of gesture performed by the participant to carry out the specific task.

2.2 Participants

Twenty-four participants between 20 and 34 years, all with normal or corrected to normal vision took part in this study. All participants had stereovision better than 40 min of arc, as tested with the Randot® stereotest. Participants were either students or employees at the Eindhoven University of Technology with no or little knowledge of using gestures as interaction technology. Students were compensated with 5 Euros for their participation

2.3 Setting, Apparatus and Materials

The stimuli of the task were displayed on an HHI Free2C 3D Display, a high-resolution autostereoscopic display with stereo video head tracking. All gestures were recorded with three camera's (one from the left, one from the right and one from above) to ensure the gestures would be clearly visible for later analysis. Three medical images were used: a scan of a heart, a hip with blood vessels, and an image of a spine. These images were obtained from a public domain website and did not contain any identifiable patient information. The stereoscopic images were presented on an autostereoscopic display, with a maximum disparity of approximately 60 min of arc.

2.4 Tasks and Procedure

The participants performed seven tasks as described above. The tasks were explained with

short scenarios in order to avoid using technical terms such as rotation or zoom-in or zoom-out. For example, we formulated rotation as "When you want to see the back or the side of the volume, how should you do that?" This was done to avoid *a priori* associations with desktop metaphors, mobile phones with touch screens, or other technical products users may be familiar with.

On arrival at the lab, users were made aware that they were being recorded during the experiment. After a stereo acuity test we explained the context of the experiment and mentioned that any kind of gesture was acceptable in this experiment. To familiarize participants with the task of gesturing in relation to what was being displayed, the experiment started with 4 colored squares, were the task was to point to the colors mentioned by the experimenter. The proper experiment then started with an overview of the three medical images, and participants had to move one of the images to the bottom or the top of the screen, followed by a selection of a specific image. After selecting an image, it was subsequently shown and users performed five tasks. The first was activating the volume, followed by rotating the image, and zooming in and out of the volume. The last task was to deactivate the volume. After completing these tasks the overview image was presented again and users had to select the next image. This procedure was repeated for all three medical images.

The order of the images was counterbalanced for each participant. Depending on the condition the participant was in, the images were either presented in monoscopic or in stereoscopic viewing mode. The gestures performed by the participants did not result in changes in the image, i.e., the images remained static during the experiment. The duration of this experiment was around 15 minutes.

3 RESULTS

Noldus Observer XT 9 was used for the analysis of the video streams. The classification of the gestures was performed by two observers. The gestures are classified using functional categories. "Dynamic pointing" is a pointing finger moving toward the screen and back. A "double click" gesture is typically performed with a finger making two movements toward the screen. The "point and drag" gesture is performed by pointing at an object and moving it towards the new position. The "sweep" gesture is a horizontally movement of the hand or finger in front of the screen, where the start and end position is the same. The "pinch" gesture is moving two fingers towards each other. The "reverse pinch" is the other way around. "Wipe away gesture" is performed by moving the hand horizontally in front of the screen, or from top to

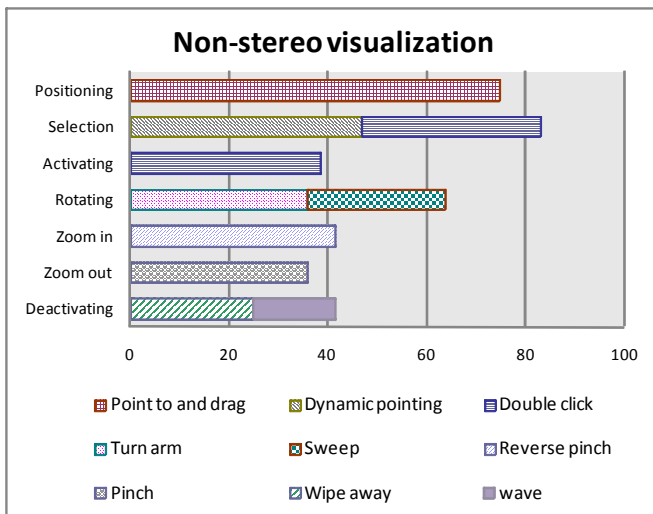


Figure 1: Gestures used for interacting with non-stereo visualizations

the bottom of the screen. The difference between the Wipe away and the sweep gesture is that the wipe away gesture is performed in one single movement, i.e. the hand does not return to the starting position.

Using these functional categories for the classification of gestures revealed that for certain types of actions there is considerable similarity in the type of gestures used by participants. However, it should be noted that we defined quite broad gesture categories, thus the gesturing behavior within one category still varied somewhat per participant. Both the number of hands or fingers, as well as the execution of the gestures varied between participants. For positioning, selecting, and activating the majority of users performed the gesture using one finger. Typical gestures that were performed with one full hand are: wave, sweep, grasp object and bring towards you, and wipe away gestures. The pinch and reverse pinch gestures were performed either with two fingers of one hand or one full hand. Further, the results showed that in non-stereo visualizations the dynamic pointing and double click gestures are performed both with one finger at one hand or one full hand, whereas in the stereo visualizations mainly one finger at one hand was used.

In Figures 1 and 2, the most frequently used gestures are plotted for each task. The results revealed that for positioning and selecting similar gestures were used in both non-stereo and stereo visualizations. However for the tasks: activating, rotating, zooming-in, zooming-out and deactivating, more variability was observed in gesture behavior.

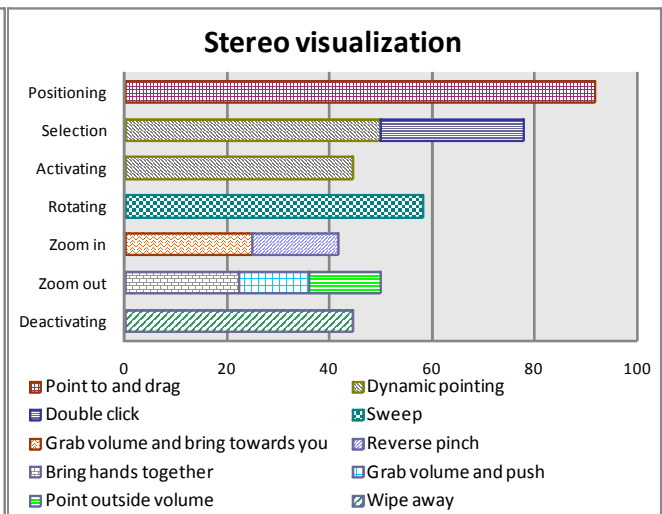


Figure 2: Gestures used for interacting with stereo visualizations

To activate an image without stereo, a “double click” gesture was preferred by 38.9% of the users, whereas in a stereoscopic presentation “dynamic pointing” was seen as the most natural way to activate the content by 44.4% of the participants. A “sweep” gesture was preferred in 58.3% of the participants for rotating an image in the stereo condition. For the non-stereo content an “arm turn” (36.1%) was used the most, followed by the Sweep gesture (27.8%). The “reverse pinch” gesture was used most frequently (41.7%) to “zoom-in” to an image when it was not presented in stereo. In the stereo mode the results showed more variability between users; 25% of the users preferred “grab volume and bring towards you” and 16.7% of the users preferred the “reverse pinch” gesture. Also, zooming out in stereo revealed more variability in gestural behavior than without stereo. In the stereo condition: 22.2% of the users preferred “bring hands together”, 13.9% “grab volume and push” and 13.9% “point outside volume”, whereas in the non-stereo condition the “pinch” gesture was preferred. The last task was deactivating the volume. In the stereo condition a “wipe away” gesture was preferred by 44.4% of the cases. In the non stereo condition the “wipe away” gesture is preferred in 25% and a wave gesture in 16.7% of the cases.

4 DISCUSSION

When users were unconstrained in the gestures they performed for the various tasks, we observed both considerable similarity as well as variability in gestures used by the participants, depending on the task action required. In line with our initial hypothesis we found a number of differences in gestures between non-stereo and stereo modes of visualization. For visualizations without stereo the gestures used for Activating, Zooming-in and Zooming-out were comparable

with traditional interaction styles using a desktop metaphor (double click, resizing windows). For stereo visualizations however those gestures were more spatial in nature, such as “sweeping” a volume or “grabbing” a volume. Contrary to our expectations, however, gestures in stereo were not performed more frequently with two-hands as compared to gestures in the non-stereo mode of visualization.

When designing and implementing natural gesture-based interactions in a medical context we can take advantage of the fact that for some interactions (positioning and selecting), the gestures are relatively uniform. However, other, more complicated actions, such as rotating and zooming, showed more variability, thus making a “one-size-fits-all” implementation of such actions less intuitive for at least some of the users. Of course, in limited, well-specified tasks (e.g., browsing an EMR, without zooming or transforming the image), the natural set of gestures may be more limited than in our study. Alternatively, gesture recognition software could incorporate a learning algorithm making it more robust to some of the between- and within-user variability. Moreover, our results also demonstrated some differences in gesture-based interactions in relation to non-stereo versus stereo visualizations of the same content. Although this may partly be due to individual variation, it should be taken into account as a potentially relevant parameter in the design of future medical HMIs utilizing 3D displays.

ACKNOWLEDGEMENT

Support from the EC HELIUM3D project is gratefully acknowledged (www.helium3d.eu). In addition the authors like to thank Inge Brandt for her support in executing the experiment and observing the video files.

REFERENCES

- [1] C. Graetzel, T.W. Fong, S. Grange, and C. Baur, A non-contact mouse for surgeon-computer interaction, *Technol Health Care*, 12(3):245–57, 2004.
- [2] M. Schultz, J. Gill, S. Zubairi, R. Huber, and F. Gordin, Bacterial contamination of computer keyboards in a teaching hospital, *Infect Control Hosp. Epidemiol.*, 4(24):302–3, 2003.
- [3] A. Nishikawa, T. Hosoi, K. Koara, D. Negoro, A. Hikita, S. Asano, H. Kakutani, and F. Miyazaki, FAcE MOUSE: A novel human-machine interface for controlling the position of a laparoscope, *IEEE Trans. on Robotics and Automation*, 19(5):825– 841, 2003.
- [4] J.P. Wachs, H.I. Stern, Y. Edan, M. Gillam, J. Handler, C. Feied, and M. Smith, A gesture-based tool for sterile browsing of radiology images, *J Am Med Inform Assoc.* 15:321–323,2008.
- [5] A. Hauptmann, Speech and gestures for graphic image manipulation, *Proc. CHI'89*, pp. 241-245, 1989.