2010 IEEE International Symposium on Multimedia

User-evaluated Gestures for Touchless Interactions from a Distance

Wim Fikkert
Human Media Interaction,
University of Twente
Enschede, The Netherlands
f.w.fikkert@ewi.utwente.nl

Paul van der Vet
Human Media Interaction,
University of Twente
Enschede, The Netherlands
vet@cs.utwente.nl

Anton Nijholt
Human Media Interaction,
University of Twente
Enschede, The Netherlands
anijholt@cs.utwente.nl

***Abstract*—Very big displays are now commonplace but interactions with them are limited, even poorly understood. Recently, understanding touch-based interactions have received a great deal of attention due to the popularity and low costs of these displays. The direct extension of such interactions, touchless interactions, has not. In this paper we evaluated gesture-based interactions with very big interactive screens to learn which gestures are suited and why. In other words, did 'Minority Report' get it right? We aim to discover to which extend these gesture interfaces are technology-driven and influenced by prototyped, commercial and fictive interfaces. A qualitative evaluation of a gesture interface for wall-sized displays is presented in which subjects experienced the interface while completing several simple puzzle tasks. We found that simple gestures based on the act of pressing buttons was the most intuitive.**

***Keywords*-Gesture interface; very large displays;**

I. INTRODUCTION

Digital displays keep on growing in size and they are finding their way into homes, offices and public spaces more and more. However, in many cases these displays are not yet interactive and they merely present information. An untapped potential is the interactivity of these large displays for which there are diverse application areas: meeting rooms [1], surgery rooms [2] and shopping centres [3]. Examples include informing passers-by in shopping centers in an entertaining manner through playful interactions [4] or by integrating these displays in urban games that mostly focus on devices that are carried around by the players [5]. For an interface to be successful, it is of paramount importance to understand the interactions that take place with it. It was recently shown that touch-sensitive displays can be built cheaply and with ease and, as a result, there has been extensive attention for the interactions with touch-sensitive displays [6]. However, interactions with large displays cannot always be based only on tactile input. Displays might be placed out of reach or behind glass surfaces to prevent vandalism, other users might obstruct the interaction [6] or the display might be too large [7].

The display surface cannot or may not be touched in touchless, gesture-based interactions. These interactions are a direct extension of touch-based interactions and, as a result, they can and should complement one another. But what makes hand gestures are suited for these touchless interactions? The choice for gestures in commercial systems and scientific prototypes is driven by technological developments and by the cost, complexity and availability of the sensors that are used to look at the users gesturing [8]. This often entails that seemingly unnatural gestures must be learned by users to accommodate the sensor, for example, using the flat hand in various poses to navigate a menu [9]. This contradicts other claims that the interaction should come naturally [10]. Fictive interfaces, on the other hand, portray compelling, impressive gesture interfaces in popular movies such as 'Minority Report'[1] and 'Paycheck'[2]. The gestures portrayed in these movies are meant to look futuristic yet recognizable and, as such, are often based on everyday actions, for example, handling a physical tool. However, the extent in which these gestures are really intuitive is mostly ignored [9]; 'intuitive' meaning how easy users can learn, remember and correctly perform these gestures. Our goal is to find out what makes a touchless interaction successful:

H1 Touchless, gesture-based interactions with large displays are best based on intuitive, everyday actions.

To explore this hypothesis we present an evaluation of a gesture set that forms the basis for a large display gesture interface. This gesture set is comprised of gestures from movies, literature and commercial interfaces.

In the remainder of this paper we first describe, in Section II, related work that focuses on gesture-based interactions, from both a user and a technological perspective. Section III then describes the method we used to find out what consitutes natural gesturing in human-computer interfaces. The results of our evaluation are reported in Section IV and we draw conclusions in Section V. Concluding this paper, our current activities in expanding the touchless interactions with a multi-touch panel are described in Section VI.

II. RELATED WORK

Popular science fiction movies depicte several gesture-based interfaces. The basis for the futuristic-looking interactions is formed by familiar actions in these movies.

---

[1] http://www.imdb.com/title/tt0181689, 21-Jul-2010.

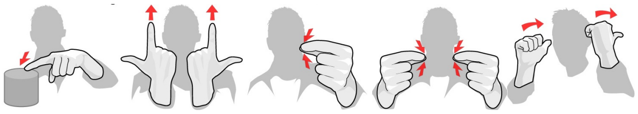
[2] http://www.imdb.com/title/tt0338337, 21-Jul-2010.

978-0-7695-4217-1/10 $26.00
DOI 10.1109/ISM.2010.30

(a) (b)

Figure 1: The large display gesture interface with (a) one of the subjects with the interface and (b) the prototype gloves that were designed based on gesturing from previous observations, literature, movies and commercial systems.

In Paycheck[2], a 3D hologram representation of a product design is manipulated through two hand-held pens. Using button presses on these pens, (parts of) the virtual design can be moulded, taken apart and combined. Large vertical displays are controlled from a distance with futuristic gloves in 'Minority Report'[1]. Gestures include pointing by ray-casting, zooming by moving the hands relative to/from each other in depth and selecting by encircling a target. This gesture interface has since then been implemented in the g-speak spatial operating system[3]. By combining hand poses and hand movements in (a)synchronous and (a)symmetrical ways, the system performs requested tasks. The gestures in a prototype system, named g-stalt, are tuned to manipulate photos, see Figure 2. Simple gestures such as pointing through ray-casting and select through a *ThumbTrigger* gesture [11] are possibly intuitive but the more complex gestures that are proposed in g-stalt are not, see Figure 2.



(a) (b) (c) (d) (e)

Figure 2: Gestures in the g-stalt prototype[5]: (a) 'get photos of this object', (b) 'bring up more photos', (c) 'grab and move space', (d) 'grab and rotate space' and (e) 'reset view'.

In addition to these fictive interfaces, there is also a large body in scientific research on the matter. Nielsen et al. [12] introduced design criteria for the human-based approach in which gestures need to be easy to perform and remember, intuitive, not physically stressing and logical in terms of functionality. In a paper mock-up interface, it was found that iconic gestures can represent objects and pointing with the index finger or waving with the hand in the general direction of an object selects it. Other tasks such as 'move' and 'select all' required an explicit state-transition gesture that resulted in rather obscure gestures such as stopping an action with a 'halt' emblem. These signal gestures are explicit and potentially intuitive for the users [13]. Several efforts have been made to formulate a gesture set that is based on its users. A Wizard of Oz study was used to discover a gesture set for controlling the SmartKom system [14]. Users pointed with one or more fingers and with one or two hands. Selecting was done by circling around an object or region while new forms of interactions such as 'no' or 'go back' were realized by a kind of waving of the hands.

By showing the system's response to a gesture, a teach-back experiment was used to discover gestures in multi-touch tabletop interactions [15]. Both unimanual and bi-manual gestures were observed; two hands are used for enlarging and zooming into an object but not for shrinking or minimizing. The users preferred to use only one hand in their gestures. Such touch-based interfaces are also popular in movies because they mimic familiar, everyday actions which makes the interface immediately appealing. Tangible and virtual objects alike are manipulated in similar ways by moving and rotating them on the surface. Examples of touch sensitive displays are seen in 'Quantum of Solace'[5] in which a team is analyzing fake money bills and in 'The Island'[6] where an artist impression of a boat is drawn.

Gestures in these and other interfaces are highly idiosyncratic, often hard to learn and designed to directly replace GUI commands. It is not uncommon that users need to learn to very accurately shape their hand to model 3D objects

---

[3] http://oblong.com, 23-Jul-2010.

[4] http://zig.media.mit.edu/Work/G-stalt, 23-Jul-2010.

[5] http://www.imdb.com/title/tt0830515, 21-Jul-2010.

[6] http://www.imdb.com/title/tt0399201, 21-Jul-2010.

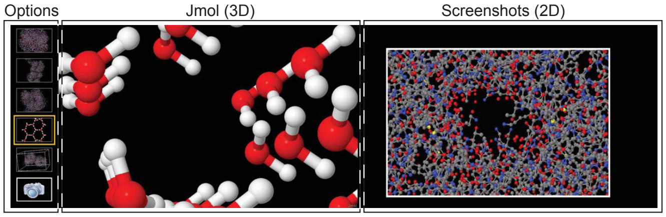


Figure 3: The Graphical User Interface used in this experiment with an overview of the three panels in the GUI.

[16]. The sensor in the interface is not always determines the 'best' gesture for a task; gesture interfaces are also driven by popular interfaces such as the Apple iPhone with its *Fingers apart* gesture for zooming: the SixthSense prototype [17] has a mobile, camera-based gesture interface that uses a projector to visualize information on any surface.

III. METHOD

Repeated gesturing forms the basis for the evaluation of our gesture interface. By performing each gesture multiple times, our users experienced the gesture interface. This gives us qualitative insight into the interaction and, more precisely, the users' perception of employing gestures in it. We used questionnaires on prior experiences, the overall interaction and the interaction per command to evaluate the user experience while operating the interface.

A. *Semantics*

Our subjects performed four randomized pattern-matching tasks. Each task consisted of finding a goal-state that was a certain orientation and zoom-level of a complex 3D mesh. Tasks were completed when the 3D mesh indistinctively matched the goal-state. One of four 3D meshes, which were biochemical structures, could be selected at any one time. The four meshes were similar in their 3D shape from certain angles yet very different from another angle. This forced the subjects to explore the 3D structure of each mesh before being able to find the goal-state. The image of the desired goal-state was provided and could be referred to at any moment. However, we did not inform the subject which mesh belonged to the image of the desired goal-state. The subjects were not required to have any knowledge of the biochemical structures nor of their visualization standards.

1) Setup: The large display, sized 400 × 125 cm, used in this experiment is depicted in Figure 5. The display was created with two projectors that displayed a total resolution of 3840 × 1200 pixels. The projectors were mounted on the ceiling in such a way that the user did not cast shadows on the projection screen. The user was allowed to walk in front of the screen but was not allowed to come closer than 1.5 meters to the screen. An infrared camera was used to look at the user's pointing behavior. We darkened the room in which the experiment took place to remove the presence of sunlight that might hinder the computer-vision recognition process. The room was uniformly lit with fluorescent lights. The state-of-the-art for detecting, recognizing, interpreting and, equally important, reacting to the subjects' gesturing in an unobtrusive way uses mostly camera-based solutions and is far too immature for our purposes. Our gestures are fine-grained gestures in which minor changes in hand shape and bending of the fingers convey the gesture's meaning. We designed and built a pair of wearable, wireless gloves that allowed us to evaluate the selected gestures, see Figure 1b. We attached the laser pointer to the back of the hand so that the whole hand can be used for pointing [18]. In this prototype we used red lasers so that the user was directly aware of where he was pointing. Each glove is equipped with two buttons that the subject can use to perform a *ThumbTrigger* gesture. The buttons are sewn on elastic rings that could be placed anywhere on any finger. By allowing the positioning of these rings we wished to evaluate the most comfortable spots on the left and right hands where the user would use *ThumbTrigger* or *Pinch*. We could detect pressing, holding and releasing each button separately and for both gloves simultaneously. Button presses were registered via a wireless Bluetooth connection; users felt hindered by tethered sensors in preliminary tests.

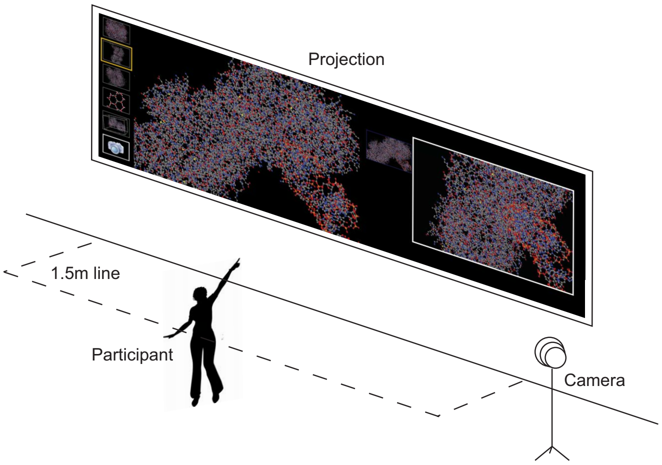


Figure 5: The setup used in this prototype. The whole display was in view of the camera. The subject could walk around but was not allowed to approach the display.

2) Graphical user interface: The graphical user interface that we used in this experiment is depicted in Figure 3. It consists of three borderless panels: an options menu, a 3D mesh and a collection of 2D screenshots. The screenshots represent past stored states that the user could revert to, changing the 3D mesh to the past state. The menu allowed users to load a specific structure, toggle a bounding box around the structure and to create a screenshot of the 3D mesh. By selecting a biochemical structure from the menu, its 3D mesh was loaded in the middle panel with a default orientation and zoom-level. The structure could be rotated and zoomed in and out. These complex biochemical struc-

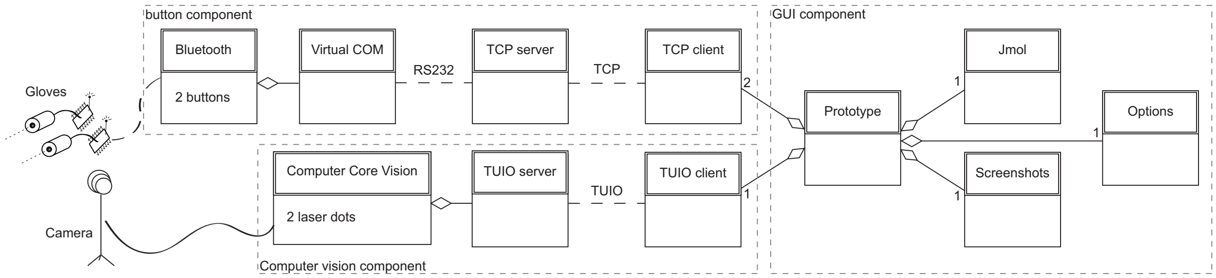


Figure 4: The software layout of our prototype; separate components ran autonomously on different computer systems.

tures ensured that our subjects would first spend some time searching for the correct structure and then some more time searching for the correct orientation and scale: lengthening the time spent interacting.

To enable the user to switch easily between previously visited locations we facilitated the use of screenshots that could be ordered as the subject saw fit. To ensure that all available commands were indeed repeatedly given by each subject, we requested the creation of at least two and deletion of at least one screenshot per goal that was offered. The screenshots were presented in the right-most panel, by default sized to 10% of the display height. Screenshots could be removed when no longer needed and resized to help recall the details of each screenshot. We represented the goal-state as a screenshot that could not be loaded or removed.

B. *Software*

Our software implementation consists of three main components, see Figure 6. The lasers were detected using the Computer Core Vision (CCV) open-source package[7] that allowed us to detect and track multiple laser dots. The camera coordinate system was mapped to our interface in a pre-trail calibration step. The laser dots locations were passed on via the TUIO protocol [19]. The prototype interface responded only when a button was pressed or released, depending on the location of the laser dots, and the button(s) that was/were pressed. We used Jmol[8] to visualize the 3D meshes.

C. *Commands*

We selected gestures based on our previous experiments [9]. In our first experiment we performed a Wizard of Oz study by allowing users to gesture freely in order to control zoom and pan in a map application on a large display. Gestures that we observed differed mostly in the preparation and retraction phases of the whole gesture. Subjects explicitly changed their hand shape their hand shape from rest to a flat hand or pointing hand for panning and two flat or pointing hands for zooming [13]. These gestures, and others from literature, movies and commercial interfaces where then evaluated in a large-scale, online user study. 100+ subjects rated video representations of 26 distinct gestures for 7 commands based on intuitiveness, ease of use and comfort. Pixel-precise ray-casting was expected for pointing at targets, selections were made in a way that mimicked button presses, moving hands and fingers apart was preferred for resizing while our suggestions for opening a context-sensitive menu were not liked. These gestures were then reevaluated and confirmed in a partially working prototype were all gestures were performed by each subject. Subjects found gesturing with one hand more comfortable although indicating distance, for resizing, was the exception. The best-scoring gestures were selected for the experiment described in this work. Commands are issued through gesturing. Users explicitly start and stop each gesture by changing their handshape. We describe the selected gestures below. Each gesture was evaluated with a questionnaire for ease of use, comfort and intuitiveness. We also asked the subjects regarding the glove design. Gestures are modeled as state-changes in a interactions-state diagram [9].

1) Out-of-range and tracking: Ray-casting is used to detect whether a subject was in the out-of-range or in the tracking state [9]. When pointing at the display, both hands were tracked separately. Subjects could point to one or two panels with one or both hands simultaneously.

2) Select and deselect: Both *ThumbTrigger* [11] and *Pinch* were included in our evaluation, see Figures 6a and 6b respectively. *Pinch* is a variation on *ThumbTrigger* that was not present in our previous evaluations. It was included here because slight user-dependent variations were observed in the execution of *ThumbTrigger* that increased comfort levels while gesturing. Sensors used to detect these gestures were placed on the index and middle fingers by default but could be altered to increase comfort. It was possible to select and deselect in the menu and screenshot panels. The six menu options could be selected or a screenshot could be selected to restore it to the 3D panel. Selections could be undone by selecting another option or screenshot.

3) Rotate: A special selection case is rotating in a 3D visualization. The subject dragged their laser dot with *ThumbTrigger* with one hand on the 3D panel to rotate the biochemical structure to the desired orientation. We selected ArcBall for rotating around the x and y axes. In our exploratory trials we found that ArcBall rotation around

---

[7] http://ccv.nuigroup.com/, 23-Jul-2010.

[8] http://jmol.sourceforge.net/, 23-Jul-2010

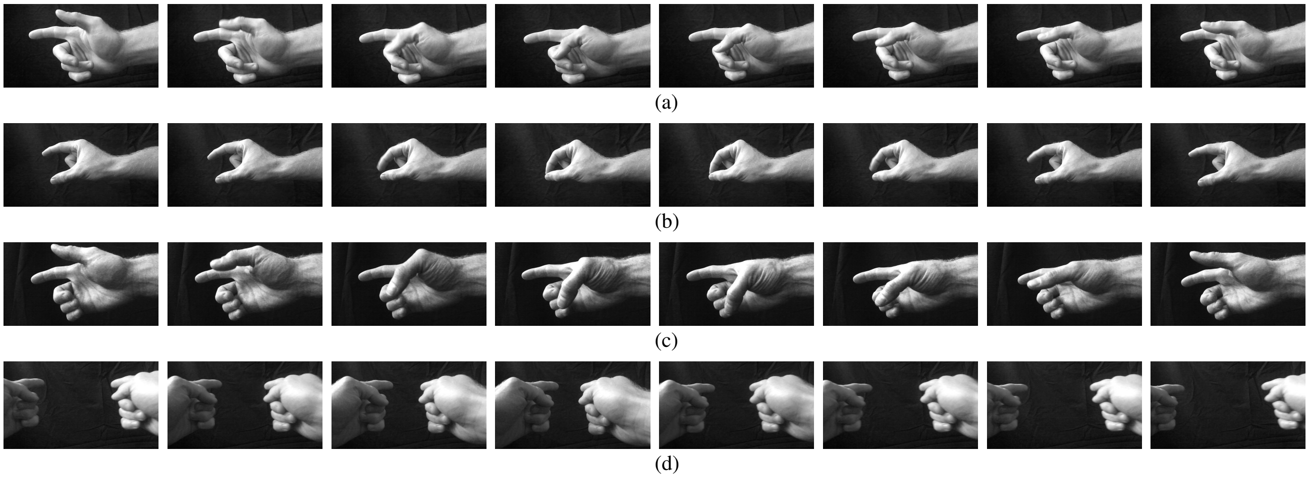

(a)
(b)
(c)
(d)

Figure 6: (a) *ThumbTrigger*: thumb touches middle finger [11], (b) *Pinch*: fingertips touch, (c) *PinkieTrigger*: thumb touches pinkie finger and (d) *Hands apart*: holding *ThumbTrigger* while distance between hands indicates size.

the z axis is desired, we facilitated this using *PinkieTrigger* while moving the laser dot horizontally, see Figure 6c.

4) Resizing – shrinking and enlarging: Given the display's scale we focused on *Hands apart* for the resize command, see Figure 6d [9]. Subjects performed *ThumbTrigger* with both hands to signal the gesture's start and end. The thumbs were pressed down for the duration of this gesture. Subjects could resize targets, structure or screenshot, on the the 3D and screenshot panels.

5) Restore and Remove: The 3D mesh with the orientation and size depicted in a screenshot could be restored in the 3D panel by performing *PinkieTrigger* on a screenshot. Bimanual *PinkieTrigger* removed a specific screenshot, it was not possible to undo that action.

IV. RESULTS

A. Sample

Twenty-three subjects participated in this within-subjects design, all studied at or worked for our university. The average age was 29 years old (ranging 24-47 years, $\sigma = 5$ years). All subjects completed the experiment. Five subjects were female, 18 were male, all where right-handed. Eight subjects held a Bachelor's degree, 13 subjects held a Master's degree and two a PhD degree. All subjects except one were unfamiliar with the four 3D meshes that were used in the prototype.

On a seven-point Likert-scale, subjects were moderately familiar with pen-based devices such as a PDA and tablet PC ($\mu = 4.3$, $\sigma = 2.2$) and they also mentioned the Nintendo DS, cellphones and the Apple iPhone in this category. The subjects were not familiar with the iPhone ($\mu = 2.9$, $\sigma = 2.3$) but more so with other multi-touch systems ($\mu = 3.4$, $\sigma = 1.8$). They mentioned the touch tables at our research group, the iPhone itself and the track pad on their notebook. Our subjects were moderately familiar with the Nintendo Wii and its controllers ($\mu = 3.8$, $\sigma = 1.6$) but less so with other gesture interfaces ($\mu = 2.5$, $\sigma = 1.7$) for which they mentioned the Playstation EyeToy, data gloves, photo play and Firefox mouse gestures. Our subjects were not so familiar with video clips of gesture interfaces ($\mu = 3.5$, $\sigma = 1.5$). 'Minority Report' was mentioned explicitly nine times while other sources were 'The Island' (2), 'Paycheck', 'Star Trek' (2), 'Iron Man' but also Oblong's G-Stalt[3] and Microsoft's Surface. Other gesture interfaces that were named included: 'endoscopic operation robot in surgery', 'EMG-based guitars' and 'Microsoft Natal'. A D'Agostino-Pearson K^2 analysis showed that there are normal distributions for these ratings except for experience with other gesture interfaces ($K^2 = 9.860$, $p < .01$) than Nintendo's Wii. This deformation is a result of a high values for skewness and for kurtosis.

B. Experiences during the experiment

1) Questionnaire overall: Trials lasted on average 37 minutes ($\sigma = 4$minutes) for the four pattern-matching tasks, 7 minutes ($\sigma = 2$ minutes) per task. A D'Agostino-Pearson K^2 analysis showed that the ratings for the whole interaction do not follow a normal distribution. The overall experience was positive, see Figure 7. Our subjects understood how the lasers were used for pointing, the pointing accuracy, operation speed and comfort while interacting were high, and there was limited fatigue in the hands and arms while interacting. The rating for the 'fun-factor' was high as well. The smoothness of the interaction scored somewhat lower.

One subject commented that the interaction could have been smoother. Three subjects mentioned that 'Getting used to [the interface] is difficult because the lasers have the same color', indicating they had trouble to estimate which dot originated from which hand. It was also mentioned that 'Inaccuracy was not so much a bother because you get visual feedback from the interface *and* the lasers'.

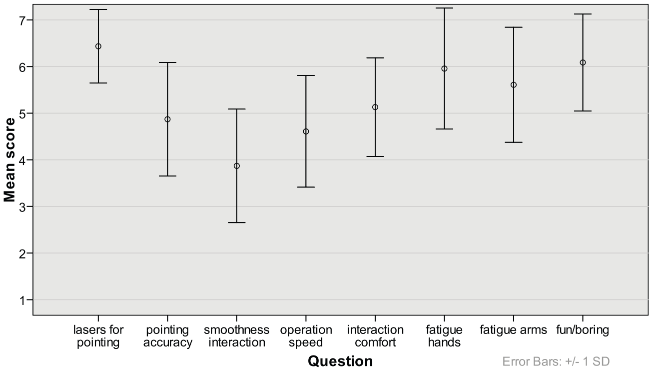


Figure 7: Overall interaction ratings.

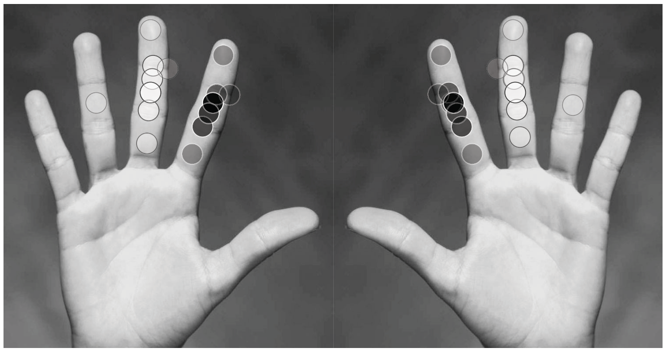


Figure 8: Button placement on the hands, black and white dots represent the two buttons.

Figure 8 shows where the buttons were placed on the subjects' hands. The default placing in the middle of the index and middle fingers at the beginning of each trial could be adjusted freely. Three subjects decided to change the buttons when so asked. Others did so of their own accord, mostly because the rings were either too wide or too narrow, which was especially true for the five female subjects with their slender fingers. Our subjects did not mention that this caused the gesture to be uncomfortable. We found no significant difference for the comfort between men and women. Our female subjects rated the perceived operation speed significantly higher as the males did ($p = .02$).

2) Questionnaire per command: The ratings to learn and remember a gesture and for the comfort in gesturing scored similarly for all seven commands. A D'Agostino-Pearson K^2 analysis shows that the ratings per command do not follow a normal distribution: for learning and remembering a gesture we found $K^2 = 41.9$ ($p < .01$) and for the gesture comfort we found $K^2 = 42.5$ ($p < .01$). This deformation is caused by high values for kurtosis (1.7 and 1.4 respectively) and high negative values for skewness (-1.4 and -1.3 respectively). A Kruskal-Wallis H analysis shows that there is a significant difference between ratings for the seven commands with respect to how easy they were to learn and remember ($\chi^2 = 36.466$, $p < .01$) but not for the comfort of performing the gesture ($\chi^2 = 8.125$, $p = .23$). We performed an independent samples analysis on the seven commands using a Mann-Whitney U analysis for the question how easy it was to learn and remember the gesture. Rotating and resizing the structure (3D) both scored significantly higher than moving a screenshot and selecting options (2D). Moving and resizing a screenshot (2D) scored significantly lower than restoring and deleting a screenshot. Resizing, restoring and deleting a screenshot scored significantly higher than selecting options.

Two subjects commented that, for rotating the structure in 3D, there was an irritant jitter that they attributed to low resolution of the pointing device. Four subjects mentioned that the response time was too high. Another subject found this approach intuitive because it is based on the traditional mouse-based control of 3D space. Three subjects mentioned that small changes in pointing could lead to big changes in resizing the structure. One subject mentioned the iPhone as the source of this gesture.

For moving screenshots in 2D, four subjects mentioned that the calibration is very important and that it should be better calibrated because 'coupling [the laser dot to the screenshot] was clumsy'. Another subject mentioned that this method was very easy to understand: 'point and click, how much easier can it get?'. One subject found that resizing screenshots should have worked the same way as in 3D, three found it hard to find the correct spot for both laser dots to start resizing.

Restoring a screenshot to 3D with *PinkieTrigger* was deemed unintuitive by one subject who preferred to just drag the screenshot to the 3D panel. Seven subjects mentioned that they did not use this gesture much. For selecting options from the menu, one subject mentioned that he had accidentally selected options while performing resizing in 3D because she came too close to the menu with a laser dot. Another subject mentioned that additional feedback mechanisms, for example, an audible click, would be nice. With respect to the two-handed *PinkieTrigger* gesture for deleting, three subjects mentioned they felt that they had not used it much. One subject mentioned that, for all comfort questions, he would have scored higher if our gloves would have been smaller 'like a [..] ring'.

3) Observations: Our subjects mostly stood with their upper arms along their body and both their lower arms pointing towards the screen, even when they were only actively using one or even neither of the hands. They mentioned that this was the most comfortable way for them to stand. It was rare for subjects to walk around in front of the display although it was explicitly explained that it was allowed. All subjects switched their standing leg to increase comfort. Most subjects first loaded all four molecules to explore their shapes before starting to fit them to the requested goal. Subjects frequently mixed up the meshes.

Almost none of the subjects noticed that they switched hands for pointing, between their left and right hand. When asked why they did so, they were surprised to find out that this was the case and mentioned that it was more comfortable for pointing. One subject commented that she was ‘very right-handed’ although we observed that she too was switching her left and right hands for pointing. We did not observe any subject always using the left hand to point to the left side of the display, or vice versa. All subjects mentioned, when asked, that they liked the visual feedback that the laser dots provided to them. They also argued that it was clear that when they did not press a button, the interface would not respond. All but one subject mentioned that they liked pointing with the whole hand because it was more comfortable to keep their fingers relaxed. One subject had significant difficulties in perceiving depth in the 3D panel, 3D projection was suggested for improved depth perception.

V. CONCLUSION

Gesturing to issue commands towards our very big display is enjoyable. This input modality was experienced as accurate, fast and comfortable. Even though our subjects tended to keep their arms tensed during the trails, they experienced little fatigue in the hands and arms. By switching hands for gesturing and legs for standing we believe that the subjects implicitly rested their body during the interactions. The shape of the *ThumbTrigger* and *PinkieTrigger* gestures was fitted to our subjects’ own comfort. They placed the buttons that we used to detect the thumb pressing against another finger so that it was most comfortable for them. This mostly meant that the subject had to minimally bend his finger so that he could give a command with minimal effort. Women’s slender fingers sometimes hindered them to place the buttons as they desired. However, this did not influence our findings. All subjects felt comfortable to wear gloves even though they had to be tightly strapped to the their hands and wrists. By giving them such an explicit means to interact through buttons on a small wearable device, the interface is transparent and provides an explicit signal to other users who controls the display.

Which gestures are suited and why? We found *Raycasting* combined with *ThumbTrigger* and *PinkieTrigger* to be the best suited gestures for our interface. These gestures proved to be fun and comfortable for giving commands to a large display from a distance. In addition, they were easy to learn and remember for the duration of our trials.

So are touchless, gesture-based interactions with large displays best based on intuitive, everyday actions? Yes. Basic actions such as pointing and grabbing on-screen objects are fun, comfortable and easy to learn. However, none of the more complex, elaborate gestures that are portrayed in ‘Minority Report’ and in several other systems were judged the same. Gestures that are simple, easy to learn, remember and perform gestures are suited for gesture interfaces. These gestures are based on everyday actions and are based on the most familiar actions in operating machines: pressing buttons. However, newer interfaces already influence the way users expect interfaces to work; zooming other than with *Fingers apart* or *Hands apart* is very confusing.

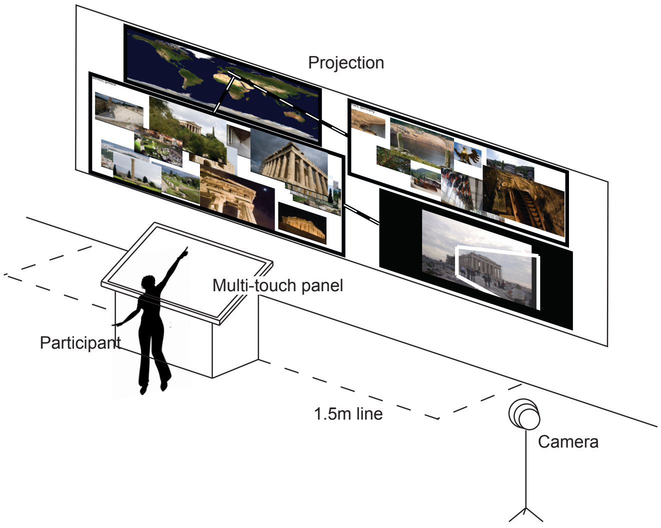


Figure 9: The gesture interface combined with a multi-touch panel. Users have to stand in front of the multi-touch panel to access detailed information.

Prolonged exposure to this type of interface would also help understanding when and how users experience fatigue, especially with Microsoft’s Kinect[9] gaming system becoming available in late 2010. In this controller-free interface, the user himself becomes the controller. Video announcements show that with the Kinect system, gamers are required to be more active than with its main rival, the Wii platform. Fatigue during prolonged gaming sessions will become a point of concern and interest [20].

VI. WORK-IN-PROGRESS

Multi-touch sensitive surfaces are commonplace and its interactions are well understood [15]. We are now in the process of extending the gesture interface that we described in this paper with an additional multi-touch panel, see Figure 9. The combination of using the hands gesturing to control a wall-sized display from a distance and a touch-sensitive panel from close by allows users to interact with detailed data at a close range whilst also having an interactive overview available. We are finding out how users experience having direct access to both the details—at their fingertips—and the overview—at a distance—in a complex, puzzle task.

Users organize photos based on the location they were taken. They will have access to 3D mapping of the photos based on Panoramio[10], descriptions of the list of available locations and a world map. Each photo is matched to the

---

[9] http://www.xbox.com/en-us/kinect, 25-Jul-2010.

[10] http://www.panoramio.com, 22-Jul-2010.

location it was taken and, in addition, they have to be placed in 3D relation to other photos that were taken in the same location. We will use tourist locations such as the centre of large European cities to construct our data set. The task is again designed so that users are required to repeatedly perform the same interactions. We are interested to find out when and why users will use the large display, the multi-touch and the combination of these displays.

REFERENCES

[1] P. van der Vet, O. Kulyk, I. Wassink, W. Fikkert, H. Rauwerda, B. van Dijk, G. van der Veer, T. Breit, and A. Nijholt, "Smart environments for collaborative design, implementation, and interpretation of scientific experiments," in *Workshop on AI for Human Computing (AI4HC)*, ser. International Joint Conference on Artificial Intelligence (IJCAI), vol. 20. AAAI Press, Jan. 2007, pp. 79–86.

[2] J. Wachs, H. Stern, Y. Edan, M. Gillam, C. Feied, M. Smith, and J. Handler, "Gestix: A doctor-computer sterile gesture interface for dynamic environments," in *Soft Computing in Industrial Applications*, ser. Advances in Soft Computing. Springer Berlin / Heidelberg, 2007, vol. 39/2007, pp. 30–39.

[3] H. Brignull and Y. Rogers, "Enticing people to interact with large public displays in public spaces." in *Proceedings of the IFIP International Conference on Human-Computer Interaction (INTERACT '03)*, 2003, pp. 17–24.

[4] T. Bekker, J. Sturm, and B. Eggen, "Designing playful interactions for social interaction and physical play," *Personal and Ubiquitous Computing*, 2009.

[5] A. D. Cheok, K. H. Goh, W. Liu, F. Farbiz, S. W. Fong, S. L. Teo, Y. Li, and X. Yang, "Human pacman: a mobile, wide-area entertainment system based on physical, social, and ubiquitous computing," *Personal Ubiquitous Comput.*, vol. 8, no. 2, pp. 71–81, 2004.

[6] P. Peltonen, E. Kurvinen, A. Salovaara, G. Jacucci, T. Ilmonen, J. Evans, A. Oulasvirta, and P. Saarikko, "It's mine, don't touch!: interactions at a large multi-touch display in a city centre," in *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*. New York, NY, USA: ACM, 2008, pp. 1285–1294.

[7] M. Nacenta, D. Aliakseyeu, S. Subramanian, and C. Gutwin, "A comparison of techniques for multi-display reaching," in *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '05)*. New York, NY, USA: ACM, 2005, pp. 371–380.

[8] A. Agarawala and R. Balakrishnan, "Keepin' it real: pushing the desktop metaphor with physics, piles and the pen," in *Proceedings of the SIGCHI conference on Human Factors in computing systems (CHI '06)*. New York, NY, USA: ACM, 2006, pp. 1283–1292.

[9] W. Fikkert, P. van der Vet, G. van der Veer, and A. Nijholt, "Gestures for large display control," in *Gesture in Embodied Communication and Human-Computer Interaction, 8th International Gesture Workshop, GW 2009*, ser. Lecture Notes in Computer Science, vol. 5934/2010. Springer Berlin / Heidelberg, 2010, pp. 245–256.

[10] S. Oviatt and P. Cohen, "Perceptual user interfaces: multimodal interfaces that process what comes naturally," *Communications of the ACM*, vol. 43, no. 3, pp. 45–53, Mar. 2000.

[11] T. Grossman, D. Wigdor, and R. Balakrishnan, "Multi-finger gestural interaction with 3d volumetric displays," in *Proceedings of the 17th annual ACM symposium on User interface software and technology (UIST '04)*. New York, NY, USA: ACM Press, 2004, pp. 61–70.

[12] M. Nielsen, M. Störring, T. Moeslund, and E. Granum, "A procedure for developing intuitive and ergonomic gesture interfaces for hci," in *Gesture-Based Communication in Human-Computer Interaction*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, Feb. 2004, vol. 2915/2004, pp. 409–420.

[13] M. Wu, C. Shen, K. Ryall, C. Forlines, and R. Balakrishnan, "Gesture registration, relaxation, and reuse for multi-point direct-touch surfaces," in *First IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TableTop '06)*, 2006, pp. 185–192.

[14] N. Beringer, "Evoking gestures in smartkom - design of the graphical user interface," in *Gesture-Based Communication in Human-Computer Interaction*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2002, vol. 2915/2004, pp. 409–420.

[15] J. Wobbrock, M. Morris, and A. Wilson, "User-defined gestures for surface computing," in *Proceedings of the 27th international conference on Human factors in computing systems (CHI '09)*. New York, NY, USA: ACM, 2009, pp. 1083–1092.

[16] M. Kavakli, M. Taylor, and A. Trapeznikov, "Designing in virtual reality (desire): a gesture-based interface," in *Proceedings of the 2nd international conference on Digital interactive media in entertainment and arts (DIMEA '07)*. New York, NY, USA: ACM, 2007, pp. 131–136.

[17] P. Mistry, P. Maes, and L. Chang, "Wuw - wear ur world: a wearable gestural interface," in *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems (CHI EA '09)*. New York, NY, USA: ACM, 2009, pp. 4111–4116.

[18] K. Cheng and M. Takatsuka, "Hand pointing accuracy for vision-based interactive systems," in *Human-Computer Interaction (INTERACT '09)*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2009, vol. 5727/2009, pp. 13–16.

[19] M. Kaltenbrunner, T. Bovermann, R. Bencina, and E. Costanza, "Tuio: A protocol for table-top tangible user interfaces," in *Proceedings of the 6th International Workshop on Gesture in Human-Computer Interaction and Simulation (GW 2005)*, 2005.

[20] A. Nijholt, D. Reidsma, and R. Poppe, "Games and entertainment in ambient intelligence environments," in *Human-Centric Interfaces for Ambient Intelligence*, H. Aghajan, R. Delgado, and J. C. Augusto, Eds. Academic Press, Oct. 2009, pp. 393–413.