Modeling Distant Pointing for Compensating Systematic Displacements

Sven Mayer, Katrin Wolf, Stefan Schneegass, Niels Henze
VIS, University of Stuttgart
Stuttgart, Germany
{firstname.lastname}@vis.uni-stuttgart.de

ABSTRACT
Distant pointing at objects and persons is a highly expressive gesture that is widely used in human communication. Pointing is also used to control a range of interactive systems. For determining where a user is pointing at, different ray casting methods have been proposed. In this paper we assess how accurately humans point over distance and how to improve it. Participants pointed at projected targets on a wall display from 2m and 3m while standing and sitting. Testing three common ray casting methods, we found that even with the most accurate one the average error is 61.3cm. We found that all tested ray casting methods are affected by systematic displacements. Therefore, we trained a polynomial to compensate this displacement. We show that using a user-, pose-, and distant-independent quartic polynomial can reduce the average error by 37.3%.

Author Keywords
distant pointing; mid-air gesture;

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation (e.g. HCI): User Interfaces

INTRODUCTION
Human communication is often supported by gestures. Probably the most common gestures is the pointing gesture selecting an object, place, or person. One of the earliest examples of absolute distant pointing is used in Bolt's seminal 'Media Room' [2]. Users could interact with the system through a combination of distant pointing and speech input. A large body of human-computer interaction research further advanced Bolt's work for various tasks and investigated its usability. With the rise of ubiquitous computing, pointing at real world objects is also a topic worth to be investigated, for example, to switch the light on and off through pointing at the light source [7]. Since the introduction of the Wii Remote and the Kinect, absolute distant pointing at virtual objects is also widely used in consumer products.

CHI'15, April 18 - 23, 2015, Seoul, Republic of Korea

ACM 978-1-4503-3145-6/15/04...$15.00
http://dx.doi.org/10.1145/2702123.2702332

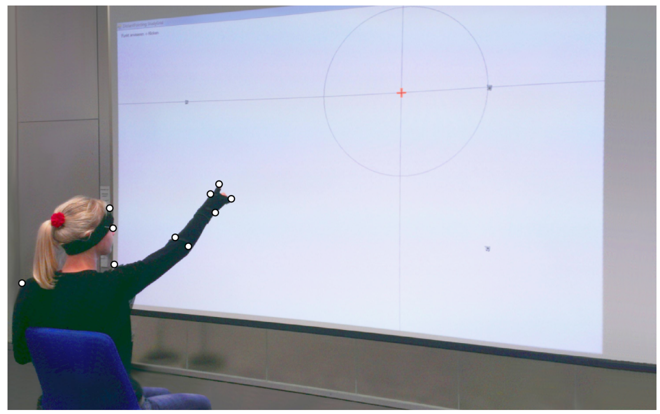

Figure 1. The study setup with the motion capture marker shown as white circles.

Work in psychology shows that pointing is such a fundamental activity that it is already developed in early childhood [6]. Already at this age, children begin to express themselves through pointing gestures. Haviland, however, also emphasizes that '*pointing may seem a primeval referential device, it is far from simple: It is complex*' [6, p. 156]. Foley and Held further show that humans do not point at targets with perfect accuracy [5]. Even if a person tries to point straight at a distant target, a ray cast that virtually extends a person's arm or finger does not necessarily hit the center of the target.

Humans' limited accuracy when pointing does not necessarily pose a problem for human communication. Through context and inference humans are excellent in resolving potential ambiguities when the intended target is not clear. Current computing systems, however, lack this understanding of the situation. Targets in graphical user interfaces, for example, are typically dense and require highly accurate input techniques. Therefore, it is crucial to precisely determine where a user intends to point at. While previous work developed interaction techniques to allow distant pointing (e.g., [13]), increasing the accuracy of absolute distant pointing itself has not gained much attention in human-computer interaction research.

In this paper we analyze and improve the accuracy of selecting targets through absolute distant pointing. First, we present a target selection study to determine precise body postures while pointing using a motion capture system. Analyzing the recorded data, we describe the pointing accuracy when using different ray cast approaches. We show that simple ray casting is limited to an average error of 59.7cm when standing 3m in front of the target. Using the collected data we developed a model that compensates systematic displacements to reduce the inaccuracy. We show that the developed model improves users accuracy by 37.3% which corresponds to an average absolute error of 23.7cm. We discuss the implication on current consumer products and close the paper with an outlook on future work.

RELATED WORK

A substantial body of research investigated selection of distant targets. One strand of research focused on the use of relative input devices to steer a cursor (e.g., [3]). In contrast, we are interested in absolute distant pointing as it is not only already used in commercial devices but also in human communication. Another strand of research investigated users' performance with absolute pointing devices that provide visual feedback about the location the user points at. Myers et al., for example, compared users' performance when using laser pointers or similar devices [11]. Vogel and Balakrishnan [13] investigated absolute pointing without a device by steering a cursor that provides feedback. In contrast, we also focus on situations where no visual feedback can be provided.

Distant pointing has been widely addressed in other domains. Kendon [9], for example, provides a general overview about the body posture when a person points. Absolute distant pointing with and without visual feedback has intensively been addressed in psychology and psychophysics. In particular, psychology aims to understand cognitive and physical processes while humans are pointing at distant targets. Foley and Held [5], for example, found that direction of the sighting eye doses appear to have a large influence on the pointing direction. Psychology focuses on qualitative models that increase our understanding of how humans point but provide no quantitative models compensating systematic errors.

In HCI, previous work on absolute distant pointing without visual feedback mainly focus on casting a ray out of a body posture. Using the direction of the ray the intersection with potential targets can be determined. Argelaguet et al. [1] classify ray cast techniques by the origin of the ray. They distinguish between *hand rooted techniques* and *eye rooted techniques*. Corradini and Cohen describe the most common hand rooted technique as '*passing through the base and the tip of the index finger*' [4]. We will refer to this ray casting approach as index finger ray cast (IFRC). There are commonly two different *eye rooted techniques* used. First, the direction of the eyes [12] which is also known as 'gaze ray cast'. The second technique uses the eyes as root and the tip of the index finger as direction of the ray cast. In this case it is common to use the point between the eyes as eye root point, Kranstedt et al. [10] described this as 'Cyclops eye'. We refer to this technique as eye finger ray cast (EFRC). Nickel et al. [12] further investigated *elbow rooted techniques* by using the ray between the elbow and the hand (forearm ray cast (FRC)). Furthermore, previous work also compared different ray casting approaches and assessed their accuracy showing that the technique needs to be selected depending on the task [8].

Overall, a significant body of work investigated the selection of distant targets. In particular, previous work proposed different ray casting approaches and investigated their accuracy. In contrast, we do not only aim to assess users' accuracy but also improve it by compensating systematic displacement.

METHOD

We conducted a study to accurately determine the body posture while pointing at distant targets using a motion capture system. The aim is to use this data to determine the accuracy of different ray casting approaches, determine systematic displacements, and to develop a model to compensate the displacement. In the study participants pointed at targets projected on a large screen in front of them from different distances and with different body postures.

Design & Task

As we aimed to get a spectrum of body postures while pointing we varied the distance between the participant and the projection screen (2m and 3m). In addition, participants pointed at the targets while sitting and while standing. Participants took part in all four conditions resulting in a 2x2 repeated measures design. The Targets were arranged in a 7x5 (column x row) grid resulting in 35 target positions. We show a red cross at one of these positions minimizing the size of the target to the center of the cross. Participants pointed three times at each target resulting in a total of 420 target selections per participant. We counterbalanced the order of the four conditions using Latin square and randomized the order of target positions. The targets were projected on a 4.5m x 3m large screen (see Figure 1). The spacing of the target grid was 0.7m x 0.6m. Thus, the distance between the leftmost and the rightmost target was 4.2m. To reduce carryover effects from one selection to the next, participants had to come back to a starting posture before pointing at the next target.

Apparatus & Measurements

As apparatus, we used a Windows 7 PC connected to a projector and the marker based motion capture system OptiTrack by NaturalPoint. The tracking system delivers the absolute position of the markers attached to the participant at 30 FPS. We calibrated the system as suggested by the manufacturer resulting in millimeter accuracy. Therefore we used 17 cameras which were positioned in the way that each spot was covered by at least 4 cameras. We equipped each participant with 16 markers (cf., Figure 1) to get a precise description of the participants' posture (marker count and positions: 4 hand, 2 wrist, 2 elbow, 4 head, 2 shoulder, and 2 hip). We implemented a tool in C# to project the targets and to record the tracking data, which recorded all markers at 30 FPS. In addition to recording participants' body posture while pointing, we asked them to fill a NASA Task Load Index (NASA-TLX) after each condition to check for fatigue effects.

Procedure & Participants

We recruited participants with various professions including mechanical engineers and judicial assistants. In total, 12 participants took part in the study (6 female, 6 male). The age of the participant was between 16 and 27 (M = 24.4, SD = 2.7). The body height was between 167 and 194cm (M = 177.6, SD = 9.9). All of them were right handed and none of them had any locomotor coordination problems.

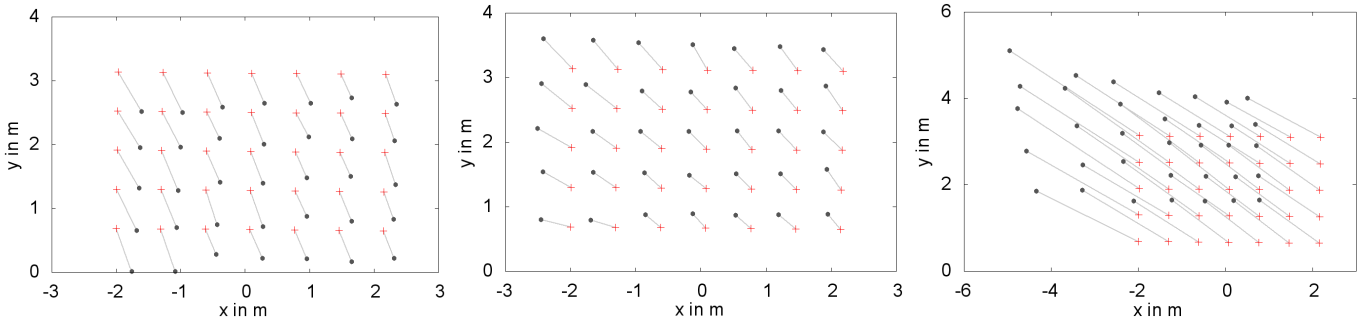

Figure 2. The average intersection point (gray dot) and the targets (red cross) for the ray casting methods EFRC (left), IFRC (center), and FRC (right).

After welcoming a participant we explained the procedure of the study and asked to fill an informed consent as well as a demographic questionnaire. Afterwards, we asked them to stand or sit at a specific position and point at the targets using their dominant hand. To compensate for natural hand tremor they had to hold the pointing position for at least one second. To ensure this time span, the participant had to click with the non-dominant hand on a button of a remote control when they started holding. The target disappeared after one second. We instruct the participant to point as they would naturally do in other situations. We intentionally did not restricted participants pose to record a range of pointing postures. While the experiments we observe it procedure from behind of the participant.

ANALYSIS & MODELING

We first analyzed the NASA-TLX to determine if we have to consider fatigue effects. The average NASA-TLX score was M = 31.0 (SD = 13.2) after the first, M = 29.0 (SD = 15.3) after the second, M = 29.9 (SD = 16.6) after the third, and M = 27.2 (SD = 15.3) after the fourth pass. As a repeated measures one-way ANOVA did not reveal a significant effect ($F_{3,33}$ = 1.740, p = .178), we assume that the effect of participants' fatigue is negligible.

Throughout analysis and modeling, we used the three ray casting methods EFRC, IFRC, and FRC that have been used in previous work. In total we collected 5040 pointing gestures. As a first step we filtered the data to remove outliers using the distance between position where the ray cast intersects with the projection screen and the position of the target. We removed outliers for each ray casting method, condition, and target individually that are more than two standard deviations away from the average. Thereby, we removed 38 trials for EFRC, 12 trials for IFRC, and 39 trials for FRC

Accuracy of Ray Casting

We determined the distance of the point where the ray cast intersects with the projection screen for the three ray casting EFRC, IFRC, and FRC. Table 1 shows the average distances for the three methods and the four conditions. The average distance is 58.1cm for EFRC, 61.4cm for IFRC, and 278.9cm for FRC. For EFRC and IFRC the distance is smaller for 2m than for 3m and also smaller for standing than for sitting. For FRC the average distance between ray cast intersection is more than four times higher than for the other methods and lower for standing than for sitting.

To determine reasons for the large deviations, we further analyzed the displacement for the individual targets. Figure 2 exemplarily shows the average intersection for the standing 3m away condition. For all three methods, the displacement is similar for all targets. The average intersection point is 23.4cm to the right and 49.7cm below the target for EFRC, 34.7cm left and 31.2cm above the target for IFRC, and 200.4cm left and 140.1cm above the target for IFRC.

Model for Improving Pointing Accuracy

As we found that the accuracy of the three ray casting methods is limited, we investigated approaches to compensate systematic displacements. In a first step we transformed each pointing gesture in the two angles α_{lr} (horizontal deviation) and α_{bt} (vertical deviation) to get a distance-invariant measure of the individual trials. Thereby, we can derive the according two correction angles Δ_{lr} and Δ_{bt} that describe the deviation between pointing ray and a ray to the target.

After transformation to angles, we fit a function that removes systematic displacement and thereby improves the accuracy. This requires one function for the horizontal deviation and one function for the vertical deviation. We generated 4 models by fitting the data to 4 different functions using *ordinary least squares*. The first function f_1 is a one-dimensional polynomial complete function of second degree. For the model we fit α_{lr} to Δ_{lr} and α_{bt} to Δ_{bt}. The functions f_2 to f_4 are complete two-dimensional polynomial functions. f_2 is of degree 1, f_3 of degree 2, and f_4 of degree 4. For these three functions we fit both α values to the Δ values.

		EFRC	IFRC	FRC
sitting	2m	53.8 (45.0)	57.9 (31.4)	353.4 (445.0)
	3m	69.8 (63.8)	72.6 (38.7)	334.4 (265.2)
standing	2m	48.6 (45.8)	55.4 (20.7)	222.9 (182.5)
	3m	60.1 (59.7)	59.7 (27.4)	204.9 (71.8)

Table 1. Mean distances between ray cast and target. SD in brackets, all distances are in cm.

		EFRC	IFRC	FRC
sitting	2m	37.9 (25.8)	35.6 (12.5)	53.6 (21.6)
	3m	41.4 (27.5)	47.2 (15.7)	59.2 (30.9)
standing	2m	36.7 (12.5)	35.0 (10.0)	44.7 (21.2)
	3m	40.4 (12.8)	36.0 (9.2)	45.1 (20.2)

Table 2. Mean distances between ray cast and target when using the model with f_4. SD in brackets, all distances in cm.

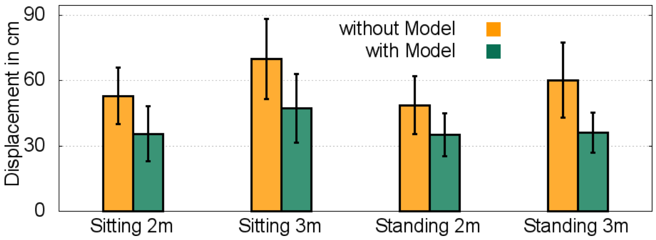

Figure 3. The improvements of the model when using index finger ray cast (IFRC) and fitting function f_4 (error bars show the standard error).

Evaluating the Model's Performance

We tested the four functions using leave-one-out cross-validation. For each participant, we fitted a model using the data of the 11 remaining participants. Afterwards, we determined the remaining error for the participant's trials. Thereby, we determined the performance of the four function if used as a user-independent model to compensate systematic displacement. For all three ray casting methods, the performance of the four functions follows the same trend. For IFRC, for example, The two linear functions already reduce the mean error to 40.8cm for f_1 and 40.6cm for f_2. The two-dimensional polynomial f_3 reduces the mean error to 40.4cm. The two-dimensional polynomial f_4 results in the smallest mean error (38.5cm): $f_4(x, y) = ax^4 + by^4 + cx^3y + dxy^3 + ex^3 + fy^3 + gx^2y^2 + hx^2y + ixy^2 + jx^2 + ky^2 + lxy + mx + ny + o$. The coefficients for the correction functions ($f_{4,lt}$ & $f_{4,bt}$) are show in Table 3 when using α_{lr} as x and α_{bt} as y.

The average distance for the four conditions are shown in Table 2. Compared to the displacement of the standard ray casting methods (see Table 1) the accuracy is improved for all methods. IFRC results in the smallest error with and without compensating systematic displacement. On average over all conditions the model reduces the error by 37.3% when using IFRC. Figure 3 contrasts displacement for IFRC. One of the origins of the remaining error is the free choice of pointing posture. Forcing specific postures can reduce this error.

CONCLUSION

In this paper, we aimed to improve the accuracy of absolute distant pointing. In a study we asked participants to point with their dominant hand at project targets from 2m and 3m while sitting and while standing. Testing three commonly used ray casting methods, we found that even the most accurate ray casting method (index finger ray cast) had an average error of 61.3cm. We found that all tested ray casting methods are affected by systematic deviations. Therefore, we trained a polynomial to compensate systematic displacement. We show that using a user-, pose-, and distant-independent quartic polynomial can reduce the average error by 37.3%.

We aimed to find a user- and pose-independent model and did not force participants to point in a specific way. Considering each of these aspects could further improve the pointing accuracy. If a system, for example, recognizes if the user is sitting or standing, it could select a corresponding model and thereby further improve accuracy. Furthermore, most systems that use absolute pointing for input provide the user with visual feedback. The effect of such a model on pointing with visual feedback needs to be investigates in the future.

coef.	lr	bt	coef.	lr	bt
a	0.0296	−0.0439	i	−2.6181	−1.1506
b	0.0190	0.1070	j	−144.4819	−72.1956
c	−0.0258	−0.0070	k	239.7431	310.0211
d	−0.0634	0.0212	l	77.4749	151.2857
e	−7.7225	−2.2891	m	2863.6584	−1495.0381
f	−3.0723	−19.5427	n	4786.0898	−8136.1496
g	−0.1239	0.0598	o	528615.8408	−522112.5319
h	−2.4860	−0.6280			

Table 3. The coefficients for the correction function f_4 (in 10^{-5}). The coefficients are rounded with in the 95% confidence bounds.

ACKNOWLEDGMENTS

This work is partly supported by DFG within SimTech Cluster of Excellence (EXC 310/2).

REFERENCES

1. Argelaguet, F., Andujar, C., and Trueba, R. Overcoming eye-hand visibility mismatch in 3d pointing selection. In *Proc. VRST* (2008).
2. Bolt, R. A. Put-that-there: Voice and gesture at the graphics interface. In *Proc. SIGGRAPH* (1980).
3. Boring, S., Jurmu, M., and Butz, A. Scroll, tilt or move it: using mobile phones to continuously control pointers on large public displays. In *Proc. OzCHI* (2009).
4. Corradini, A., and Cohen, P. R. Multimodal speech-gesture interface for handfree painting on a virtual paper using partial recurrent neural networks as gesture recognizer. In *Proc. IJCNN* (2002).
5. Foley, J., and Held, R. Visually directed pointing as a function of target distance, direction, and available cues. *Perception & Psychophysics 12*, 3 (1972).
6. Haviland, J. B. *Pointing: Where Language, Culture, and Cognition Meet*. Sotaro Kita, 2003, ch. Pointing Is the Royal Road to Language for Babies.
7. Holzapfel, H., Nickel, K., and Stiefelhagen, R. Implementation and evaluation of a constraint-based multimodal fusion system for speech and 3d pointing gestures. In *Proc. ICMI* (2004).
8. Jota, R., Nacenta, M. A., Jorge, J. A., Carpendale, S., and Greenberg, S. A comparison of ray pointing techniques for very large displays. In *Proc. GI* (2010).
9. Kendon, A. *Gesture: visible action as utterance*. Cambridge University Press, 2008.
10. Kranstedt, A., Lücking, A., Pfeiffer, T., Rieser, H., and Staudacher, M. Measuring and reconstructing pointing in visual contexts. In *Proc. SemDial* (2006).
11. Myers, B. A., Bhatnagar, R., Nichols, J., Peck, C. H., Kong, D., Miller, R., and Long, A. C. Interacting at a distance: measuring the performance of laser pointers and other devices. In *Proc. CHI* (2002).
12. Nickel, K., and Stiefelhagen, R. Pointing gesture recognition based on 3d-tracking of face, hands and head orientation. In *Proc. ICMI* (2003).
13. Vogel, D., and Balakrishnan, R. Distant freehand pointing and clicking on very large, high resolution displays. In *Proc. UIST* (2005).