Making Bare Hand Input More Accurate

Chat Wacharamanotham
chat@cs.rwth-aachen.de
RWTH Aachen University
52056 Aachen, Germany

Abstract
Using postures and movements of bare hands for computer input allows harnessing the benefits of the high degree of freedom, bimanual reference frame, and manual dexterity. To enable seamless interaction with both virtual and physical worlds, computers need to reliably recognize the users' intention in order to distinguish purposive input from other hand movements. We aim to investigate the influence of proprioception and interaction context on the postures and movements of the hands. Preliminary studies suggest that some of these relationships are consistent across users. This consistency potentially allows more accurate prediction of users' intention.

CHI 2014, April 26–May 1, 2014, Toronto, Ontario, Canada.
ACM 978-1-4503-2474-8/14/04.
http://dx.doi.org/10.1145/2559206.2559955

ACM Classification Keywords
H.5.2 [Information Interfaces and Presentation Interfaces]: Input devices and strategies.

Context and Motivation
Registering bare hand postures and motions for spatial input (e.g., arranging or assembling virtual objects) has the benefit of manual dexterity, bimanual frame of reference, and high degree of freedom of manipulation. Such input is proposed for interactive surfaces [8] and in the space above desk surfaces [16]. However, a challenge for bare hand spatial input is to classify movements that are intended to control the computer from those that are not. Ideally, this classification should require minimal mode-switching gestures, additional devices, and users' cognitive load. We believe that sensing additional information from the interactive environment can help improving this classification.

In this thesis, we investigate the influence of proprioception and interaction context to the postures and motions of the hands, and how they can be used to accurately classify *users' intention* behind their hand movements.

Box 1: Factors hypothesized to influence hand postures and motions

Proprioception:

- The position and orientation of fingers on the palm
- The position and orientation of two hands with respect to each other

Context:

- External reference frame
- Anticipated shapes and positions of the physical devices aimed to grasp
- The existence and intensity of haptic feedback
- Types of transfer functions and visual feedback

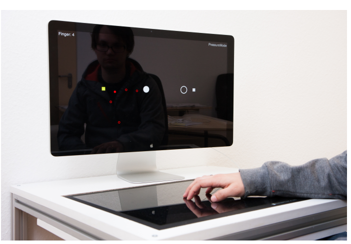

Figure 1: Indirect multitouch setup: horizontal surface for expressive input, vertical surface for ergonomic display.

Related Work

Buxton characterizes three elementary states for graphical input devices: out-of-range, tracking, and dragging [2]. Using bare hands for input increases difficulties in discerning intended input states. To minimize errors, we need to reliably recognize users' intention to dwell in or make a transition among these states. Previous works use explicit gestures [12], fixed invisible boundaries in midair [10], or head directions [9] for state transitions in midair input. These solutions either require equipping users with additional devices or users' awareness of gestures and invisible boundaries.

Hand postures and motions have been reported to be consistent within particular tasks such as describing 3D objects (miming motion and hand posture) [6], aiming a target on touchscreens (visible features of fingernails) [5], and engaging physical work environments [4]. These attributes, however, are influenced by the context of the interaction. For example, Kattinakare et al. found that distances and movement constraints influenced the accuracy of stylus movements near surface [7]. Cockburn et al. found that systematic interactions between the allowed degree of freedom and the extent of visual feedback [3]. In prehension[1] of physical objects, Barrett et al. reported that hand opening and kinematic profile were influenced by the orientation, shape, and size of the objects [1]. HCI has yet to use these relationships to improve the accuracy in predicting users' intention.

Statement of Thesis

We hypothesize that the *postures and motions* of users' hands are consistently influenced by *proprioception*[2] and *context*, which can be determined or tracked with present day sensors. Understanding how these factors influence users' hands will allow us to predict users' intention more accurately.

[1]the action of grasping

[2]the perception of stimuli generated within the user herself

Research Goals

We aim to investigate how hand postures and motions are influenced by features in Box 1, and how these features interplay. We are interested in how these relationships allow us to improve accuracy in (1) classifying intended manipulation actions from spurious movements, (2) accurately determining the magnitude of the manipulation, (3) compensating for systematic errors influenced by closed-loop visual or haptic feedback.

We focus our investigation in three desktop workspace settings that allow ergonomic interactions while being augmented with expressive input using bare hands.

Indirect multitouch surface: This setting consists of a horizontal multitouch surface for input coupled with a vertical screen for output (Fig. 1). Here, users benefit from both the expressiveness of the multitouch input and an ergonomic upright sitting position that allows the arms to rest on the desk surface [11].

Near-surface finger input: The second setup extends typical desktop computer workspace with interactive midair layers. Users interacts with both desktop devices (mouse, keyboards, or touchscreens) in combination with midair input (Fig. 2). Each midair layer has a limited thickness and they can be stacked to add additional degrees of freedom for input, e.g. lifting the finger upward to reveal auxilary information layers [10].

3D indirect input: In this setup, the finger positions and orientation in 3D are mapped to the virtual world shown

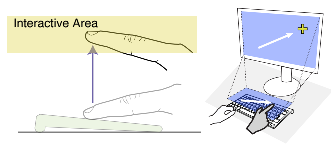

Figure 2: The area above desktop surface provide an additional input dimension.

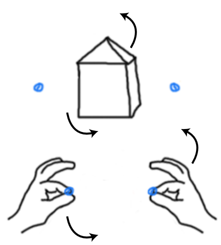

Figure 3: 3D in-air rotation setup. The user grab and rotate object with bare hands.

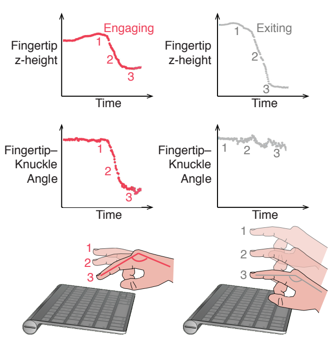

Figure 4: Hand posture changes can be use to classify air between tapping and leaving.

on the screen to control the 3D position orientation of an object (Fig. 3) [16].

Methods

For each of the settings above, we planned a series of studies:

Phase 1 elicits hand behavior by measuring how the hands behave when asked to perform a set of interactions with computer and with physical environment with statically predefined factors in Box 1.

Phase 2 uses the data from phase 1 to design classifiers for users' intention and evaluate their precision and recall in tasks that require frequent input state transitions.

Phase 3 investigates how dynamic factors such as transfer functions and closed-loop visual feedback influence the hand posture and motion.

Results to Date

Indirect multitouch input: From literature, we identified four tracking-dragging switching methods: lift-and-tap, pressure switch, pressure hold (quasi-mode), and hold. We elicited the changes of touch ellipse over time for these methods from five users (Phase 1). We then implemented recognizers for these methods and compared them in single-finger, multiple-finger, and bimanual object manipulation tasks (Phase 2). We found that the lift-and-tap technique allows users to maintain and switch to intended input states more reliably than other methods throughout all conditions [11].

Near-surface finger input: We elicited how hands behave when the users access, maintain, and leave near-surface area (Phase 1). We found that users can maintain the finger reliably within an area of 4 cm thickness, even when the movement is as large as 10 cm without arm support. The results also show that when the finger is lifted to access the near-surface area, each user has a consistent personalized height to lift the finger up. We additionally found that when the user leaves the near-surface space towards keyboards, their hand shape stays consistently flat (Fig. 4). From these results, we proposed two algorithms for recognizing users' intention (First half of Phase 2): (1) an algorithm that analyzes the velocity profile of the finger and dynamically places the center of the tracking area at the height determined by the first movement stroke to minimize the drifting outside the interaction area. (2) an algorithm that distinguishes leaving near-surface from air tapping actions [15].

3D in-air rotation: In a preliminary controlled study, the users were asked to (1) specify a rotation axis on a 3D object and (2) to rotate the object with predefined axis constraint either along X, Y, or Z (Phase 1). We found that users were accurate when specifying rotation axes along the screen width. The accuracy significantly drops when they have to specify the axes in directions parallel to the screen height or perpendicular to the screen surface. We surmised that this error is caused by a slight upward perspective view. This suggests that there is an interplay between the continuous visual feedback and the proprioception of bimanual reference frame (pending publication).

Research Situation

In our program, students are expected to contribute in teaching and different projects at the beginning of their Ph.D. and then focus on research in the later years. I am in my fifth year of a Ph.D. program in Computer Science and began my research-focus phase since October 2013. I expect one and a half years of work as planned in Box 2.

Box 2: Planned work
Near-surface input

- Verify the dynamic height and leaving recognition algorithm in with users. (Completing phase 2)
- Extend the study to multiple midair layers and longitudinal study. (Phase 3)
- Elicit the influence of the different shape of objects on hand shapes during reaching. (Phase 3)

3D in-air rotation:

- Study how much the error in 3D rotation systematically varies according to the change in parameters in the perspective projection. (Expanding phase 1)

I conducted all phases mentioned above in designing an input technique that allows users with hand tremor to accurately select targets on touchscreen [13, 14].

Expected Contributions

This thesis will contribute quantitative understanding of how proprioception and context influence hand posture and motions of bare hand input. This knowledge will lead to interaction techniques, algorithms, and guidelines to make users' intention prediction more accurate.

Acknowledgements

This work was funded in part by the German B-IT Foundation.

References

[1] T. M. Barrett, E. Traupman, and A. Needham. Infants visual anticipation of object structure in grasp planning. *INFANT BEHAV DEV*, 31(1):1–9, 2008.

[2] W. Buxton. A Three-State Model of Graphical Input. In *Proc. INTERACT '90*, 449–456.

[3] A. Cockburn, P. Quinn, C. Gutwin, G. Ramos, and J. Looser. Air Pointing: Design and Evaluation of Spatial Target Acquisition with and without Visual Feedback. *INT J HUM-COMPUT ST*, 69(6):401 – 414, 2011.

[4] Y. Guiard. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *J MOTOR BEHAV*, 19(19):486–517, 1987.

[5] C. Holz and P. Baudisch. Understanding Touch. In *Proc. CHI '11*, 2501–2510.

[6] C. Holz and A. Wilson. Data miming: inferring spatial object descriptions from human gesture. In *Proc. CHI'11*, 811–820.

[7] R. S. Kattinakere, T. Grossman, and S. Subramanian. Modeling steering within above-the-surface interaction layers. In *Proc. CHI '07*, 317–326. 2007.

[8] K. Kin, T. Miller, B. Bollensdorff, T. DeRose, B. Hartmann, and M. Agrawala. Eden: a professional multitouch tool for constructing virtual organic environments. In *Proc. CHI '11*, 1343–1352.

[9] L.-P. Morency and T. Darrell. Head gesture recognition in intelligent interfaces: the role of context in improving recognition. In *Proc. IUI '06*.

[10] S. Subramanian, D. Aliakseyeu, and A. Lucero. Multi-layer Interaction for Digital Tables. In *Proc. UIST '06*, 269–272.

[11] S. Voelker, C. Wacharamanotham, and J. Borchers. An evaluation of state switching methods for indirect touch systems. In *Proc. CHI '13*, 745–754.

[12] D. Vogel and R. Balakrishnan. Distant freehand pointing and clicking on very large, high resolution displays. In *Proc. UIST '05*, 33–42. 2005.

[13] C. Wacharamanotham, J. Hurtmanns, A. Mertens, M. Kronenbuerger, C. Schlick, and J. Borchers. Evaluating Swabbing: A Touchscreen Input Method for Elderly Users with Tremor. In *Proc. CHI '11*, 623–626.

[14] C. Wacharamanotham, D. Kehrig, A. Mertens, C. Schlick, and J. Borchers. Designing a touchscreen web browser for people with tremors. In *Wksp on Mobile Accessibility, CHI '13*.

[15] C. Wacharamanotham, K. Todi, M. Pye, and J. Borchers. Understanding finger input above desktop devices. In *Proc. CHI '14*, (To appear).

[16] R. Wang, S. Paris, and J. Popović. 6D hands: markerless hand-tracking for computer aided design. In *Proc. UIST '11*, 549–558.