SimTech Milestone Report: Finger Orientation as an Additional Input Dimension for Touchscreens

Dipl.-Inf. Sven Mayer
Institute for Visualization and Interactive Systems (VIS)
University of Stuttgart

Examiner: Prof. Dr. Niels Henze
Co-Examiner: Prof. Dr. Andrea Barth

1 Introduction

The age of ubiquitous computing has brought a large number of interactive surfaces into our lives. Interactive surfaces are present in various forms, and various contexts from tabletops to mobile devices, and touchscreens continue to remain the main input technique. In current systems, a finger touching a screen or surface is typically reduced to simple two-dimensional coordinates. To overcome the limited expressiveness when using touchscreens, a large body of research proposes to enrich such a touch interaction. In particular, previous work suggests using the finger orientation as a means of input, to address this is the fundamental problem when using touchscreens. Wang et al. [13] have proposed to adapt menus to finger orientation, e.g., for pie and torus menus. Finger orientation was also considered for creating novel interaction techniques for mobile devices. For example, understanding touch as a finger orientation was the underlying concept of *Z-touch* [12] which used finger pitch angle as an input dimension in a drawing application. Similarly, Kratz et al. [5] enabled finger orientation gesture detection with a depth camera above the touchscreen. While *Z-touch* and the work by Kratz et al. [5] detect finger orientation using a vision-based sensing system, Rogers et al. [11] used a capacitive sensor array to detect the pitch and yaw of the finger. Finally, Xiao et al. [14] used an off-the-shelf consumer touchscreen device to detect pitch and yaw as additional input. Xiao et al. [14] reported their method leads to a pitch error of $9.7°$ and a yaw error of $26.8°$ for their prototype. They further reported that during their study they faced ergonomic problems, such as long fingernails and cretin finger orientations which are hard to perform. Therefore, they limit their input space.

This work focuses on three different aspects in using finger orientation as a additional input dimension. First, we investigative two different approaches to detect the finger orientation when interacting with touchscreens. Second, we study the ergonomic constraints which constrain designers in the ways

Figure 1: The prototype with the 8.4-inch tablet and the depth camera.

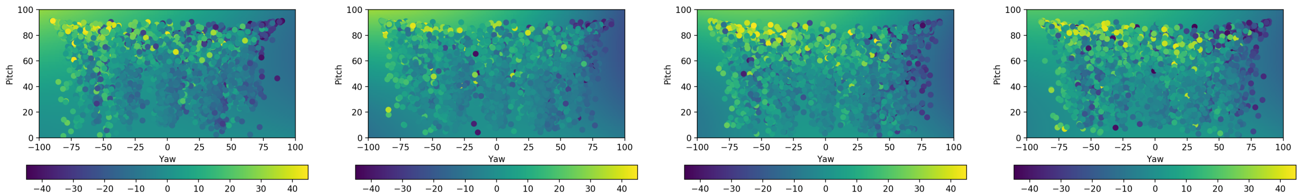
they can implement finger orientation in new systems. Third, we contribute a new method which allows us to study the social implications of new interaction techniques, in detail this enables us to study the impact of finger orientation on face-to-face conversations.

2 Recognition

In the attempt to recognize the finger orientation while interaction with a touch screen we investigated two different approaches. The first approach uses a depth camera which is attached above the touchscreen and can, therefore, observe finger while approaching the touchscreen. Here we extend from the work by Kratz et al. [5] and improved accuracy for both pitch and yaw recognition and extend the recognition to all fingers but the thumb (published at MobileHCI'17 [9]). The second approach uses an off-the-shelf smartphone and uses Machine Learning (ML) to recognizance the fingers orientation. Using a Convolutional Neural Network (CNN), we reduce the estimation error for pitch by 9.8% and yaw by 45.7% in comparison to existing approaches; these results are published at ISS'17 [7].

2.1 Depth Camera Approach

We developed a prototype with a depth camera mounted on a tablet that can determine the finger orientation. We conducted a study with 12 right-handed participants (3 female, $M = 25.83 years$, $SD = 3.31$) to record ground truth data for all fingers but the thumb to evaluate the accuracy of our prototype using the PointPose [5] algorithm to estimate the finger orientation. By applying 2D linear correction models, we show a reduction of RMSE by 45.4% for pitch and 21.83% for yaw.



(a) Index finger (b) Middle finger (c) Ring finger (d) Little finger

Figure 2: The scatterers are showing the samples from the study. The underlining plain represents the correction model for the yaw correction based on pitch and yaw of the depth camera.

2.1.1 Prototype

For our prototype, shown in Figure 1, we use a Samsung Galaxy Tab Pro 8.4 which offers $2560 \times 1600\,px$ on an 8.4-inch screen resulting in 359.39 *PPI*. As a depth camera, we use an Intel RealSense F200. The camera has a minimum sensing distance of $20\,cm$ and a resolution of $640 \times 480\,px$ at 120 *FPS*. We used this camera due to its small minimum distance in comparison to other depth cameras. However, we needed to overcome the $20\,cm$ between the tablets screen and the depth camera. Therefore, we 3D printed a mount for the tablet and laser cutted a connection plate to attach the camera to the tablet. We firmly connected the parts using metal screws.

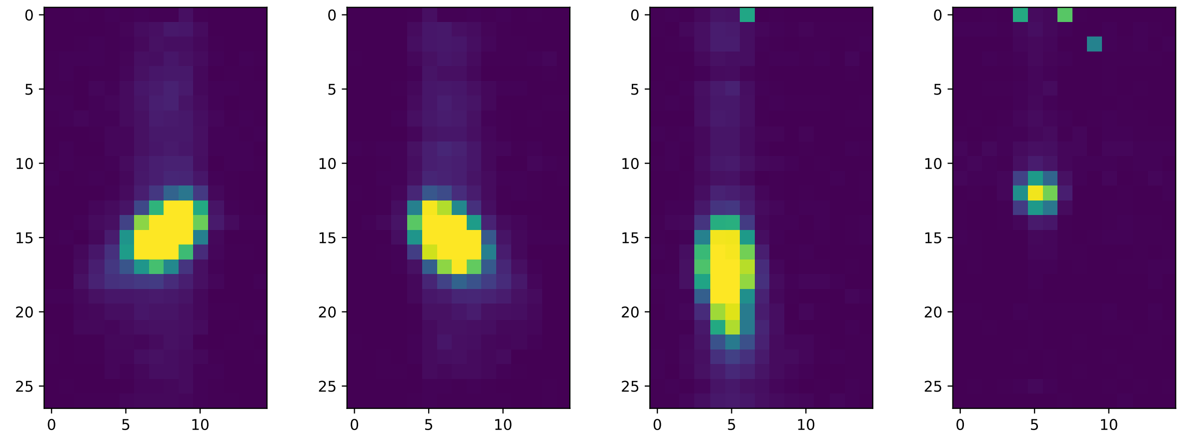
2.1.2 Study

We designed the study using a repeated-measures design with four independent variables: TARGETS, YAW, PITCH, and FINGER. We randomized the order of FINGER, and within FINGER we randomized TARGETS and YAW. To cover a broad range of possible positions, we used 20 TARGETS arranged in a 4×5 grid on the tablet. The targets further represented five PITCH input angles: $15°$, $30°$, $45°$, $60°$ and $75°$. Further, we used five YAW input angles: $-60°$, $-30°$, $0°$, $30°$ and $60°$.

The study procedure was as follows: the experimenter marked each finger with two red dots on the left and right side and on top of the finger to later calculate the finger orientation. Then, we explained that they have to touch the center of the red crosshair while aligning the finger with the longer red line indicating the yaw angle; while changing the pitch to cover all pitch angles (see Figure 1).

2.1.3 Results and Summary

Using the red marker we analyzed the data and calculated the average error of PointPose [5]. We modeled the error with a full second order two-dimensional polynomial. For the final model we used the training and test data to fit the model, we achieved an RMSE reduction by 45.4% for pitch and 21.83% for yaw. Further the fitness of the pitch correction functions for the four fingers functions is on average $R^2 = .5$, and for the four yaw functions the fitness is on average $R^2 = .66$ (see Figure 2).



(a) $23°, -42°$ (b) $23°, 41°$ (c) $17°, 4°$ (d) $80°, 3°$

Figure 3: For different capacitive images form a Nexus 5 with different finger orientations. The labels represent the finger orientations pitch, yaw.

First, we build a prototype to determine the accuracy of the algorithm by Kratz et al. [5]. Second, we showed a reduction in RMSE when applying our offset correction model. Further, we showed that the algorithm proposed by Kratz et al. [5] also could determine the pitch and yaw of the middle, ring, and little finger with an equal accuracy. We build a prototype which can be applied to any flat surface to determine the yaw and pitch of the finger. This is especially handy when working with paper prototypes to investigated new interaction techniques that take the finger orientation into account.

2.2 Machine Learning Approach

Approaches for determining a finger's orientation using off-the-shelf capacitive touchscreens proposed in previous work already enable compelling use cases. However, the low estimation accuracy limits the usability and restricts the usage of finger orientation to non-precise input. We provide a ground truth data set for capacitive touch screens recorded with a high-precision motion capture system. Using this data set, we show that a Convolutional Neural Network can outperform approaches proposed in previous work. Instead of relying on hand-crafted features, we trained the model based on the raw capacitive images. Thereby we reduce the pitch error by 9.8% and the yaw error by 45.7%.

2.2.1 Study

In a study, we collect capacitive images and respective finger orientation angles as ground truth using a motion capture system. We followed the approach by Holz and Baudisch [4] to collect ground truth data of the finger orientation. Here participants continusily entered touch samples while their finger is tracked in space using the motion capture system. We recruited 35 participants (7 female) through our university's mailing list. Due to due to technical issues we excluded two participants. For the remaining 33 participant (7 female) the age ranged from 20 to 33 years ($M = 22.9$, $SD = 3.4$).

2.2.2 Results and Summary

In our study, we collected a data set of capacitive images which represents the change in capacitance caused by fingers touching the display in different finger orientations. These images are labeled with the pitch and yaw angles of the finger through a high-precision motion capture system. In our study we collected $457,268$ labeled samples which the used to train our models. The feature-based approach by Xiao et al. [14] is based on a Gaussian process (GP) and a simple heuristic to estimate the finger orientation using a set of engineered features. In contrast, our model uses the representation approach using a Convolutional Neural Network (CNN) trained with raw capacitive images. We reduces the estimation error for pitch by 9.8% and yaw by 45.7% in RMSE when comparing to our *pseudo implementation* of the Xiao et al. [14] approach using their features. Besides the data set, one outcome of this work is an improved finger orientation estimation model that can be readily deployed to Android and iOS devices. We are publicly releasing the model for others to deploy them on their mobile devices, as well as the data set for further usage.

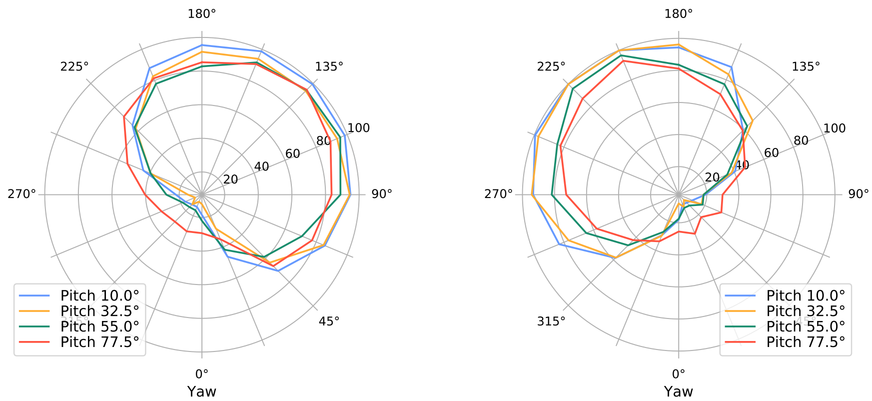
Using the latest ML algorithms, we showed a noticeable improvement over previously presented approaches. The next steps to improve finger orientation estimation on mobile devices would include using a touchscreen that provides a higher capacitive image sampling resolution. Moreover, using touchscreens with sensing capabilities above the display would enable a reconstruction of the full finger and provide more vital information to estimate the finger orientation.

3 Ergonomic Constraints

Understanding of the physiological restrictions of the hand is required to build effective interactive techniques that use finger orientation as an additional input. We conducted a study to derive the ergonomic constraints for using finger orientation as an effective input source. In a controlled experiment, we systematically manipulated finger pitch and yaw while performing a touch action. We found that finger pitch and yaw do significantly affect the perceived feasibility and 21.1% of the touch actions were perceived as impossible to perform. Our results show that the finger yaw input space can be divided into the *comfort* and *non-comfort* zones. We published these results at MobileHCI'17 [6].

3.1 Study

In a repeated measures experiment, we asked 18 participants (9 female, $M = 25.9$ years, $SD = 2.7$) to perform touch actions with their index finger. We asked them to rate the feasibility (RATING) of the touch action. Feasibility, in this context, was defined as the effort required to perform the touch action. The experiment was conducted with three independent variables: PITCH and YAW of the index



(a) Left hand rating

(b) Right hand rating

Figure 4: The average feasibility RATING (from 0 = '*easy*' to 100 = '*hard*') for the different PITCH inputs.

finger, as well as HAND. We used $10°$, $32.5°$, $55°$, and $77.5°$ for PITCH. For YAW, we covered the full $360°$ range resulting in $0.0°$ to $337.5°$ with $22.5°$ steps. All combinations were tested with the index finger of the right and the left HAND.

3.2 Results

We collected 2304 ratings, out of these 485 (21.1%) were marked by the participants as not feasible to perform. All inputs that the participants marked to be not feasible where considered to be *hard* (100 points) for the analysis. Our results show that finger orientation has a significant effect on perceived feasibility of touch actions. As expected, participants perceived actions performed with the dominant HAND as more feasible than those performed with the non-dominant hand, see Figure 4.

3.3 Summery

We conducted a study to investigate the ergonomics of finger pitch and yaw as an additional touch input parameter. We asked participants to rate the perceived feasibility of performing touch actions with different finger orientations. We systematically manipulated the finger pitch and yaw while performing a touch action. We varied the input orientations using 4 different pitch and 16 different yaw angles. All combinations were performed with the index finger of the left and the right hand. The results show that not all orientations are equally feasible, with some orientations being infeasible to perform. Furthermore, we show that the input space for each hand can be divided into two zones; the *comfort* and *non-comfort* zones. Only $135°$ out of $360°$ of all yaw orientations are within the *comfort* zone and perceived as feasible.

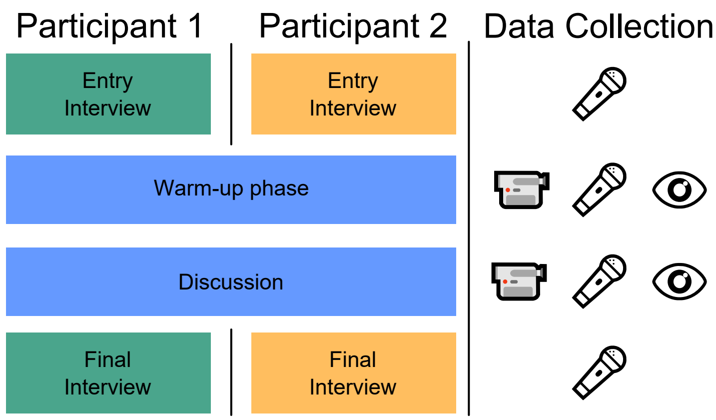


Figure 5: The study procedure in our new mixed-method approach. Data collection methods are: video, audio and eye tracking.

4 Social Implications

While the proliferation of mobile devices has rendered mobile notifications ubiquitous, researchers are only slowly beginning to understand how these technologies affect everyday social interactions. In particular, the negative social influence of mobile interruptions remains unexplored from a methodological perspective. We contribute a mixed-method evaluation procedure for assessing the disruptive impact of mobile interruptions in conversation. The approach combines quantitative eye tracking, qualitative analysis, and a simulated conversation environment to enable fast assessment of disruptiveness. It is intended to be used as a part of an iterative interaction design process. We contribute a mixed-method evaluation approach for assessing the disruptiveness of mobile interaction techniques. We postulate an approach that uses a conversation task for two participants. We employ eye tracking as a quantitative measure and combine it with qualitative evaluation based on semi-structured final interviews. In traditional coding techniques for face-to-face conversations used in past work [2, 3], hour-long encoding of video and audio material is needed. In contrast, our work offers an alternative approach that highlights how the eye-tracking and interviews can offer complementary yet different insights. Our method is interestingly different from past approaches as it is designed to offer quick answers that result in actionable insight during a design process. This work is published at CHI'18 [8].

4.1 Choice of Participants and Stimulus

As we endeavored to design an approach that would allow for rapid evaluation, we considered flexibility in terms of participant choice as a key feature. Participants are often hard to recruit, especially for studies early in the design process, when rapid feedback is needed. Our approach can be used with a wide variety of participants as it uses a generic conversation stimulus. Participants are grouped in pairs randomly. Further, the experimental design allows for large within-pair variability in eye tracking metrics and thus it does not put any restrictions on the participants' gender, age, race, native language etc. Previous work showed that gaze fixations are an effective way to evaluate two systems [1]. Moreover Okamoto et al. [10] showed eye movements are affected by conversations. Thus, when using eye tracking in conversations, this should be taken into account in the design. As a result measuring the time of eye contract or time spent looking for the conversation partner is not efficient as a disruption measurement. Therefore we propose time spent on the disruption as the measurement. Therefore, a between-groups design for the experiments is necessary as sequence effects are likely to appear when a conversation is prolonged.

We use a generic discussion task in our approach where we use a task designed by language experts specifically for conversations between strangers with a particular focus on paying attention to the other party in the discussions. Consequently, we use the discussion task from the University of Cambridge's English for Speakers of Other Languages Certificate in Advanced English Speaking Test. The task provides a stimulus that is both manageable for advanced non-native speakers and engaging enough for native speakers. The task also includes a shorter introductory segment that can be used as an icebreaker for the discussion. An additional advantage of using a speaking exam task is the fact that analogous tasks exist and can be easily found for other languages. Again, using such a task necessitates a between-groups design, as there is no possibility to assure that different discussion topics are equally stimulating to a given randomly assigned pair of participants.

4.2 Study Plan

The final study procedure in our approach is shown in Figure 5. At the beginning of each study session, the facilitators conduct individual semi-structured interviews (Entry Interview, see Figure 5). The interview serves as a means of collecting demographic data on the participants. Further, it introduces the conversation task. The purpose of the study is not revealed to the participants until the end of the study. Informing the participants about the focus on mobile disruptiveness may cause potential bias due to increased awareness to interruptions. Thus only one participant will be introduced to the additional disruptive task. After the instruction and training phace, the participants are introduced to each other and start wearing the eye trackers. The facilitator then presents the experimental task. First, the participants run through a warm-up phase to get confident with their conversation afterward also the disruption is taking place to observe the participants reactions and potential behavior change. After the discussion is concluded (we recommend a time of 10 minutes, based on language examination experience), individual debriefings (Final Interview, see Figure 5) are performed. All interviews are audio recorded. As a safety precaution, we also recommend video recording the conversation. After the study, eye tracking data is analyzed using inferential significance testing and simplified qualitative analysis with affinity diagramming is performed on the interview data.

4.3 Summery

We contribute a new mixed-method approach to measuring the disruptiveness of technology. Our new approach uses eye tracking and semi-structured interviews in a generic conversation task to offer rapid, actionable insights for designing interaction techniques that may be used in conversations. Designing techniques beyond existing ones to study the influence of different interaction mechanisms on conversational engagement remains an important challenge. We are eager to see how future designs will explore the design space that we merely begin to understand.

5 Conclusion and Future Work

In this work we investigated in three aspects in using finger orientation as an additional input dimension: first the recognition, second ergonomic constraints, and lastly we contribute a method to evaluate new interactions in social settings. In detail, we presented two approaches to estimate the finger orientation, using a depth camera or an off-the-shelf touchscreen. We showed that both approaches outperform the existing approaches. We further, presented that due to ergonomic restrictions, not all finger orientations are suitable for touchscreen input. Finally, we laid the foundation to investigate the social implications of finger orientation by presenting a new evaluation method for interactions techniques in face-to-face conversations.

5.1 Social Implications of Finger Orientation

In a next step, we want to investigate possible social implications of using finger orientation as an input. Here we will use the new mixed-method approach. We plan to compare finger orientation input and regular touch input using a set of common interactions in face-to-face conversations, such as researching a near by location or responding to an incoming instant message.

5.2 Discoverability of Finger Orientation

Lastly, we want to investigate ways to communicate the new interaction technique to the user. While a number of use cases have been proposed in previous work, the discoverability of using the finger orientation as an input is still not answered. We see hosting focus groups as one way to investigate possible representations in the graphical user interface and implications for designers. While these results tend to be vague outcome another option is to conduct expert interviews with a possible focused outcome. In both cases, we would like to study the proposes solutions in a follow-up study, in which users are able to test these new ways of communicating finger orientation input.

References

1. Bertel, S., Dressel, T., Kohlberg, T., and von Jan, V. Spatial Knowledge Acquired from Pedestrian Urban Navigation Systems. In Proc. of MobileHCI '17.

2. Brown, B., McGregor, M., and McMillan, D. Searchable Objects: Search in Everyday Conversation. In Proc. of CSCW '15.

3. Goodwin, C. Restarts, Pauses, and the Achievement of a State of Mutual Gaze at Turn-Beginning. *Sociological Inquiry* 50, 3-4 (1980).

4. Holz, C. and Baudisch, P. The Generalized Perceived Input Point Model and How to Double Touch Accuracy by Extracting Fingerprints. In Proc. of CHI '10.

5. Kratz, S., Chiu, P., and Back, M. PointPose: Finger Pose Estimation for Touch Input on Mobile Devices Using a Depth Sensor. In Proc. of ITS '13.

6. Mayer, S., Gad, P., Wolf, K., Woźniak, P. W., and Henze, N. Understanding the Ergonomic Constraints in Designing for Touch Surfaces. In Proc. of MobileHCI '17.

7. Mayer, S., Le, H. V., and Henze, N. Estimating the Finger Orientation on Capacitive Touchscreens Using Convolutional Neural Networks. In Proc. of ISS '17.

8. Mayer, S., Lischke, L., Woniak, P. W., and Henze, N. Evaluating the Disruptiveness of Mobile Interactions: A Mixed-Method Approach. In Proc. of CHI'18.

9. Mayer, S., Mayer, M., and Henze, N. Feasibility Analysis of Detecting the Finger Orientation with Depth Cameras. In Proc. of MobileHCI'17.

10. Okamoto, D. G., Rashotte, L. S., and Smith-Lovin, L. Measuring interruption: Syntactic and contextual methods of coding conversation. *Social Psychology Quarterly* (2002).

11. Rogers, S., Williamson, J., Stewart, C., and Murray-Smith, R. AnglePose: Robust, Precise Capacitive Touch Tracking via 3D Orientation Estimation. In Proc. of CHI '11.

12. Takeoka, Y., Miyaki, T., and Rekimoto, J. Z-touch: An Infrastructure for 3D Gesture Interaction in the Proximity of Tabletop Surfaces. In Proc. of ITS '10.

13. Wang, F., Cao, X., Ren, X., and Irani, P. Detecting and Leveraging Finger Orientation for Interaction with Direct-touch Surfaces. In Proc. of UIST '09.

14. Xiao, R., Schwarz, J., and Harrison, C. Estimating 3D Finger Angle on Commodity Touchscreens. In Proc. of ITS '15.