Your Phone or Mine? Fusing Body, Touch and Device Sensing for Multi-User Device-Display Interaction

Mahsan Rofouei[1,2], Andrew D. Wilson[1], A.J. Bernheim Brush[1], Stewart Tansley[1]

[1] Microsoft Research
One Microsoft Way
Redmond, WA, USA
{awilson, ajbrush, stansley}@microsoft.com

[2] University of California, Los Angeles (UCLA)
Computer Science Department
Los Angeles, CA, USA
mahsan@cs.ucla.edu

ABSTRACT
Determining who is interacting with a multi-user interactive touch display is challenging. We describe a technique for associating multi-touch interactions to individual users and their accelerometer-equipped mobile devices. Real-time device accelerometer data and depth camera-based body tracking are compared to associate each phone with a particular user, while body tracking and touch contacts positions are compared to associate a touch contact with a specific user. It is then possible to associate touch contacts with devices, allowing for more seamless device-display multi-user interactions. We detail the technique and present a user study to validate and demonstrate a content exchange application using this approach.

Author Keywords
Sensor fusion; depth camera; multi-touch; accelerometers.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous;

General Terms
Design; Experimentation; Human Factors.

INTRODUCTION
Increasingly, interactive multi-touch displays are large enough to support multiple simultaneous users, either working individually or collaboratively. In many multi-user scenarios it may be desirable to accurately associate touch input with particular users. Identifying which user is touching an interactive surface in a collaborative setting enables personalization, access control, and score-keeping. Such capabilities could enable useful shared interactive displays for walk-up use in conference rooms and office hallways, for example.

Other interesting applications for interactive displays incorporate smaller devices such as mobile phones. Such devices are personal and private, and so complement a larger display. Researchers have explored collaborative group search using phones [3], downloading content from digital displays [1], and using personal devices to access private content while interacting with public displays [8,9].

We present ShakeID, a technique for associating a specific user's touch contacts on an interactive display to a mobile device held by the user. It exploits the combination of the phone's on-board sensors and touch screen sensing to perform this association.

Previous work has explored several different methods for uniquely identifying multiple users sharing an interactive display. Dohse et al. [5] combines touch detection with hand tracking using a camera mounted above a tabletop and tracks users based on skin color segmentation after establishing identities. The DiamondTouch table [6] identifies four unique users by capacitively coupling touch inputs through users to receivers in the environment. The territory-based approach [7] divides the surface into multiple territories, each assigned to a single user. The Medusa table [12] uses 138 proximity sensors to map touch points to specific users. On mobile phones, PhoneTouch [4] detects "bump" events from the phone accelerometer and touch is used to perform user identification. Hutama et al. [2] also employs a contact-based approach which uses tilt correlation from smartphones to identify touch interactions.

ShakeID differs from these techniques in that it only requires the user to hold the phone while touching the display, rather than bringing the phone in physical contact with the display. Additionally, our system associates touches with mobile devices without requiring additional special hardware (e.g. pen). By using Kinect tracking data, our method also has the potential to provide additional capabilities, e.g. determining which hand is holding the device, and detecting when users move away from the screen and performing session log-outs automatically.

Assuming each user is holding a smartphone or other portable device that can sense its own movement, ShakeID matches the motion sensed by the device to motion observed by a Microsoft Kinect camera pointed at the users standing in front of the touch display (see Figure 1). By comparing the motion of each phone in the scene with the motion of each user, the system can associate each phone to a specific user's hand. Next, by performing a coordinate transform from the 2D space of the display to the 3D

CHI'12, May 5–10, 2012, Austin, Texas, USA.

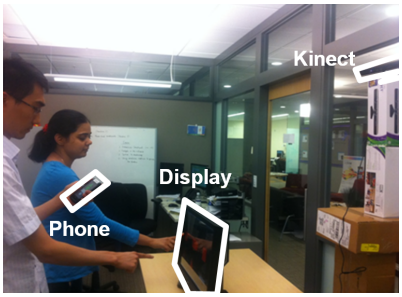

Figure 1. System includes Kinect camera, multi-touch display and 2 accelerometer-equipped phones (one visible).

camera space, the touches on the display are associated to users. Touches are thus associated to users and users are associated with devices they hold. For example, if two users touch a display simultaneously in different locations to grab content, ShakeID can associate each touch to a specific user and transfer the correct content to each user's personal device.

As far as we know this is the first attempt to fuse Kinect, mobile device inertial sensing, and multi-touch interactive displays. Our contributions are the ShakeID method for user association and an initial user study that applies the technique to enable sharing and content exchange between phones and a touch display. The study shows that ShakeID is easily learned and requires minimal feedback. In the remainder of the paper we describe how ShakeID works and then present the user study.

SHAKE ID

ShakeID uses a two-step process. The first step associates personal "private" smartphones to users holding them, while the users interact with the single shared "public" display. The second step associates touches on the shared display to users who performed those touches. We assume that the smartphones have been previously paired to the system and focus on identifying the device – there are many existing ways of establishing this pairing [10, 11].

Figure 1 shows the arrangement of the system. We implemented ShakeID using the Microsoft Kinect for Windows SDK to track the hands of multiple users, the Microsoft Surface 2.0 SDK for the multi-touch display and two Windows Phone smartphones. We placed the Kinect sensor within 1.5m distance of a vertical touch display. Physical layout is critical, since the Kinect camera has a limited 0.8-4 meter range. The Kinect SDK provides a skeletal tracking capability which can be used to track the left and right hands of two simultaneous users in the view of the Kinect camera. The Kinect was positioned so that it could capture valid skeleton data and see the users' hands the entire time users interacted with the shared display.

Algorithm

ShakeID relies on the fact that if a person is holding a phone in their hand, the acceleration measured by the phone accelerometer should match the acceleration of the hand holding the phone. ShakeID first associates each phone with a particular user's left or right hand and then similarly associates touch contacts on the touch screen with the multiple users' hands. The combination of these two steps allows touches to be associated with phones and users. Below we describe each of the steps in detail:

Step 1: Associate Phone with Hand

To associate the phone with a user's hand, we continuously correlate phone acceleration for each phone connected to the system with the accelerations of all hands tracked by the Kinect. Data captured from the 3-axis accelerometer in the phone is sent wirelessly to the display system continuously. Meanwhile, a Kalman filter is used to estimate acceleration of hand position over time.

The matching algorithm is as follows: For every phone p, and observed hand h, 3-axis accelerometer data $\mathbf{a}_p$ is compared with 3-axis hand acceleration $\mathbf{a}_h$ over a time window (empirically chosen as one second). However, accelerometer and hand accelerations cannot be directly compared because the orientation of the phone is unknown, and the phone's accelerometers include acceleration due to gravity. We address both problems by searching the space of all possible phone orientations by generating a uniformly distributed set of points on the unit sphere. Given a particular rotation matrix $\mathbf{R}$ representing the orientation of the phone, acceleration due to gravity $\mathbf{g}$ may be subtracted directly from the rotated phone accelerometer values. The hand with the most similar pattern of acceleration is determined to be the holding hand $h^*(p)$:

$$h^*(p), \mathbf{R}^*(\ ) = \arg\min_{h,\mathbf{R}} \sum_t \left\| \mathbf{a}_h(t) - (\mathbf{R}\,\mathbf{a}_p(t) - \mathbf{g}) \right\|^2$$

Note that this search gives the absolute orientation $\mathbf{R}^*(p)$ of the phone as a potentially useful by-product.

Step 2: Associate Touch with Hand

In the second step, the algorithm associates touch contacts on the display to users' hand positions. We first convert the 2D display coordinates of each contact c to Kinect's 3D coordinate system by a linear coordinate transform that is determined in an offline calibration. The hand corresponding to each contact $h^*(c)$ is found by comparing 3D hand positions $\mathbf{x}_h$ to transformed contact coordinates $\mathbf{x}_c$:

$$h^*(c) = \arg\min_h \|\mathbf{x}_h - \mathbf{x}_c\|^2$$

Finally, the system can easily map each touch contact to a phone by noting that the hands found in the first and second steps map to the same skeletal model. Note that this allows, for example, the user to hold the phone in one hand while touching the display with the other.

Limitations

An important limitation of the above process involves the case where the hand holding the phone is stationary. In this case the matching process is likely to match equally well to other stationary hands. However, in practical scenarios,

especially with larger displays, it is less likely that people will be stationary. Furthermore, if users understand that the system uses hand motion to perform user association, they may initiate a simple, natural motion such as shaking, to trigger an association. In any case, the nature of the feedback provided to users when association changes will be important, so users understand and are confident of the associations being inferred by the system.

Another limitation arises when people are standing too close for the Kinect to correctly compute skeleton data. This situation is less likely for large displays suitable for multiple users.

Because the Kinect SDK currently only provides active skeletal tracking with joint data for up to two people, the current implementation is limited to two simultaneous users. Our approach should work if more users can be tracked simultaneously, although accuracy could decrease.

USER STUDY

To evaluate ShakeID, we implemented a system for content sharing between phones and devices using ShakeID. We conducted a user study with 7 pairs of participants (14 adults: 8 male and 6 female). Each study consisted of two people performing a set of content sharing tasks. Participants were members of our organization with no prior knowledge of ShakeID. Each received a $10 incentive.

Based on our experience piloting the study, we added visual feedback indicating the position of phones at all times (see Figure 2). We also informed participants that phone movement is used to associate devices with people and if they experienced any problems (e.g. not making enough movement for the system to identify, etc.), moving the phone in their hand should improve accuracy (e.g. "shake to associate").

During each session the participants completed the following tasks:

Training: Standing alone, each participant copied 20 shapes from their phone to the matching empty spot on the display (*PhoneToDisplay*) and then copied 20 shapes from the display to their phone (*DisplayToPhone*). To copy a shape from phone to display, the participant tapped the shape on the phone. The shape may then be pasted to the

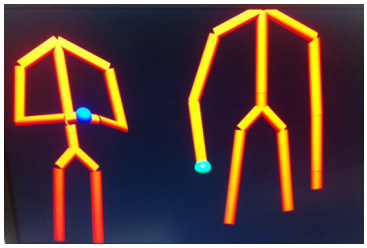

Figure 2. Feedback window. Blue (left) and green (right) spheres represent hand-held phones.

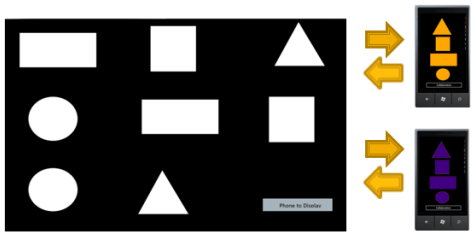

Figure 3. User study content exchange

display by tapping the desired location on the display. To copy from display to phone this process is reversed.

To clearly indicate when the system fails to associate the correct device with the user, each phone is associated with a particular color. As Figure 3 shows, white shapes on the shared display are colored by copying shapes from the phone to the display: white shapes on the display copied to the phone become the phone's color. We randomly selected which phone we gave to each user and the system determined which one the user was holding.

Parallel Use: Participants worked side-by-side, representing scenarios in which individuals work in parallel around a shared display. Each participant conducted 20 *PhoneToDisplay* copies and 20 *DisplayToPhone* copies. To simulate movement that might happen during collaboration in real-life scenarios, participants were asked to switch sides halfway through the task.

Collaborative Use: To simulate collaborative actions where participants share and discuss content through a shared display, participants copied shapes between them (e.g., P1 and P2 switched shapes by P1 copying a shape to the display and P2 copying it from the display to their device and vice versa). They repeated this for 12 shapes, with each participant originating the sharing 6 times. As before participants switched sides halfway through the task.

During the study we recorded several parameters including instances when the association was incorrect (e.g., the wrong color was received while copying) and the length of time between a tap on the display and phone. We measured tap delays to show overall system response time and to capture the use of different interaction techniques, e.g. using both hands for instantaneous copy and paste.

Results

During the Parallel and Collaborative tasks, 94% and 92% accuracies were observed, respectively. These errors primarily occurred when the hand holding the phone hand moved out of the field of the view of the Kinect and so accurate position data for the hand was not available. If an incorrect association was made, participants often shook or waved the phone to re-associate. Average delay between two taps among all users was 0.80 seconds. Ten participants used one hand to tap the two devices, while 4 participants used two hands simultaneously resulting in almost instantaneous actions of 0.16 seconds on average.

After the study, we surveyed participants' mental and physical demand using the NASA TLX questions. On a 7-point scale with 1 = "very low" and 7= "very high" participants reported low mental demand (median 2), low physical demand (median 2), and that ShakeID was easy to learn (median 2) and use (median 2).

Initially we hoped that ShakeID could successfully associate input to users unaware of how it worked. We conducted pilot studies with a "walk-up" condition where users performed the tasks with no knowledge that the association relied on movement of the phone with 8 people. We found that for some users it worked well since they naturally gestured with their hand holding the phone (5 people), while others (3 people) did not move the hand holding the phone while approaching the system.

FUTURE WORK

We anticipate a number of extensions to our basic approach for future work. While our current approach matches acceleration data directly, it may be beneficial to instead match on features derived from the acceleration. Matching with orientation invariant features such as points of maxima would remove the need to search over all orientations. This search would also be required less frequently if the mobile device sensors allow for the calculation of absolute orientation, such as those that include 3-axis magnetometers and gyros. This will likely improve the reliability of the matching process.

Our present work matches the phone to the users' hands only. But the technique may work when matching to body parts other than hands. For example, it may be possible to successfully associate a device to a user if it is in the user's pants pocket by matching to users' hip joints data from the Kinect SDK.

There may be additional ways to take advantage of the device sensors. For example, the orientation recovered by the matching process, combined with frame to frame sensor updates, can be used to provide fast and accurate hand orientation information that is not provided by the Kinect SDK. Furthermore, low-latency, high frame rate hand position data could be derived by combining the body tracking position data with the (oriented) device acceleration data using a Kalman filter.

Our current implementation does not address situations where a malicious user might imitate another user's motion to gain control of their device. Determining how easy this is in practice and designing mechanisms to prevent this is important. For example, assigning more weight to the skeleton which holds the phone first or most may mitigate this concern.

CONCLUSION

We described user identification method useful in multi-user interactive display settings. This method performs association in two steps of associating phones to users and associating touches to users. ShakeID cross-correlates acceleration data from smartphones that people carry together with hand acceleration captured through Kinect to perform user identification. To validate the accuracy of this approach we conducted a 14 person user study and showed accuracy rates of 92% and higher.

REFERENCES

1. Maunder, A. J., Marsden, G., and Harper, R. SnapAndGrab: accessing and sharing contextual multi-media content using Bluetooth enabled camera phones and large situated displays. In *CHI 2008 Extended Abstracts*, 2319-2324.
2. Hutama, H., Song, P., Fu, C. and Goh, W. B. Distinguishing multiple smart-phone interactions on a multi-touch wall display using tilt correlation. *Proc. CHI 2011*, 3315-3318.
3. Morris, M. R., Fisher, D., and Wigdor, D. Search on surfaces: Exploring the potential of interactive tabletops for collaborative search tasks. *Inf. Processing & Management*, 2009.
4. Schmidt, D., Chehimi, F., Rukzio, E., and Gellersen, H., PhoneTouch: a technique for direct phone interaction on surfaces, *Proc.UIST2010*.
5. Dohse, K.C., Dohse, T., Still, J., and Parkhurst, D. J. Enhancing multi-user interaction with multi-touch tabletop displays using hand tracking. *Proc. Advances in Computer-Human Interaction 2008,* 297-302.
6. Dietz, P.H., and Leigh, D. DiamondTouch: A multi-user touch technology. *Proc. UIST 2001*, 219–226.
7. Scott, S. D., Carpendale, M. S. T., and Inkpen, K. M. Territoriality in collaborative tabletop workspaces. *Proc. CSCW 2004*, 294–303.
8. Rukzio, E., Schmidt, A., and Hussmann, H. An analysis of the usage of mobile phones for personalized interactions with ubiquitous public displays. *Workshop on Ubiquitous Display Environments in conjunction with UbiComp 2004*, Nottingham, UK, 2004.
9. Grimes, A., Tarasewich, P., and Campbell, C. Keeping information private in the mobile environment. Position paper presented at *First Intl. Workshop on Social Implications of Ubiquitous Computing at CHI 2005*.
10. Schöning, J., Rohs, M., Krüger A. Using Mobile Phones to Spontaneously Authenticate and Interact with Multi-Touch Surfaces. *AVI*: *Workshop on designing multi-touch interaction techniques for coupled private and public displays* PPD*, 2008*.
11. Seewoonauth, K., Rukzio, E. , Hardy, R., Holleis, P. Touch & connect and touch & select: interacting with a computer by touching it with a mobile phone, *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services,* September 15-18, 2009.
12. Michelle Annett, Tovi Grossman, Daniel Wigdor, and George Fitzmaurice. 2011. Medusa: a proximity-aware multi-touch tabletop. *Proc. UIST 2011* 337-346.