See Me, See You: A Lightweight Method for Discriminating User Touches on Tabletop Displays

Hong Zhang[1], Xing-Dong Yang[2], Barrett Ens[1], Hai-Ning Liang[1], Pierre Boulanger[2], Pourang Irani[1]

[1]Dept. of Computer Science
University of Manitoba, Winnipeg,
MB, R3T 2N2, Canada
{umzha245, bens, haining, irani}@cs.umanitoba.ca

[2]Dept. of Computing Science
University of Alberta, Edmonton,
AB, T6G 2E8 Canada
{xingdong, pierreb}@cs.ualberta.ca

ABSTRACT

Tabletop systems provide a versatile space for collaboration, yet, in many cases, are limited by the inability to differentiate the interactions of simultaneous users. We present See Me, See You, a lightweight approach for discriminating user touches on a vision-based tabletop. We contribute a valuable characterization of finger orientation distributions of tabletop users. We exploit this biometric trait with a machine learning approach to allow the system to predict the correct position of users as they touch the surface. We achieve accuracies as high as 98% in simple situations and above 92% in more challenging conditions, such as two-handed tasks. We show high acceptance from users, who can self-correct prediction errors without significant costs. See Me, See You is a viable solution for providing simple yet effective support for multi-user application features on tabletops.

Author Keywords

Multi-user application, tabletop interaction, touch discrimination, position aware system.

ACM Classification Keywords

H.5.2 [Information interfaces and presentation]: User Interfaces - Graphical user interfaces.

General Terms

Design, experimentation.

INTRODUCTION

Multi-touch tabletop systems provide a shared environment for users to work together on interactive tasks [13, 14, 16, 21]. The most easily constructed and commonly used tabletops rely on vision-based touch detection. Unfortunately, these common systems cannot discriminate the touches of one user from another. We refer to these systems as being *touch-indiscriminate*. This restriction severely limits the possibilities for multi-user tabletop applications. In a game application, for example, responsibility falls on the individual for moving the correct pieces or taking their turn at the right time. Awkward solutions must be found for discriminating touches in a painting program, such as defining explicit user territories [19, 21], or requiring gestures to delineate every input [15].

Because tabletop systems are inherently collaborative, solutions have been explored to make them *touch-discriminate*. This feature enables application designers to support interactions that would not be otherwise possible (Figure 1). A common solution is to use an identifying device, held or worn by the user, as a proxy for the actual owner of a touch point [5, 18]. Another approach is to rely on users' biometric traits, such as their fingerprints [8]. Unfortunately, none of these existing systems are compatible with common vision-based tabletops without extensive modification or the use of peripheral accessories.

Figure 1. Users interacting with a drawing application with one shared color palette using our touch-discriminate technique: See Me, See You. This form of collaborative work would not be possible without maintaining distinct user states.

We introduce See Me, See You, a *lightweight* method for supporting touch-discrimination on vision-based tabletop systems. We use the orientation of a touching finger, information that can be acquired on common tabletops [22], to associate a touch with a user's position. To assess *if* and *how well* this feature supports user touch discrimination, we ran a series of studies involving tasks of various difficulties and user configurations on a minimally modified vision-based system. The results are encouraging; we find that finger orientations (FO) originating from distinct user positions around a tabletop have minimal overlap, even when two users are standing in close proximity. Tested across a variety of tasks and contexts, our results reveal accuracy rates as high as 97.5%. Our outcomes suggest that FO, albeit easy to acquire on existing systems, can be effective for tasks relying on multi-user state information.

Our contributions in this paper include: (1) a method for associating finger touches with user positions; (2) detailed profiles of FO distribution across various positions around a table; (3) a FO detection algorithm; (4) a corrective feature used to reaffirm a user's position, called the Position Aware Cursor; and (5) evidence that our method can yield highly accurate results in a variety of contexts.

DESIGN CRITERIA FOR LIGHTWEIGHT SYSTEMS

Despite the necessity in some contexts for discriminating user touches, very few systems have made this feature easily accessible. Ideally, designers should be able to quickly implement or test new prototypical concepts or novel application features that rely on multi-user state information. In this vein, we provide a list of qualities that are desirable in a lightweight, touch discriminate, multi-user tabletop:

- *Minimal device constraints*: the system should not require users to hold or wear an external device;
- *Accurate*: the system should be accurate enough to not overburden or distract users from their primary tasks;
- *Scalable*: the system should be versatile enough to handle various configurations such as multiple simultaneous users, users standing side-by-side, and uniform accuracy coverage across different regions;
- *Low cost*: building the tabletop should be achievable at an affordable cost with commonly available technology;
- *Computationally non-prohibitive*: the system should work in real-time and not suffer from excessive lag.

To facilitate the engineering of a lightweight system, we restrict our expectations with some additional caveats:

- *Limited input features*: users benefitting from a lightweight system may be willing to forgo certain types of multi-touch use, such as using the full palm to interact with objects. This would allow them to make the best use of the device's touch discriminating features;
- *Implicit trust:* the system should be designed for users who intentionally *want* touch discrimination. A lightweight system need not prevent identity deception as this would add layers of complication to normal use;
- *User adaptation*: although a lightweight system should not require long training periods, some knowledge about how the system operates can contribute to improved usage and a better user experience.

RELATED WORK

Computer vision-based tabletops are popular because their underlying technology is widely available and inexpensive [3, 7]. The two most common vision-based techniques, frustrated total internal reflection (FTIR) [7] and diffused illumination (DI) [20], recognize touches as blobs of light. Supporting touch discrimination a vision-based tabletop requires an ancillary approach.

A common strategy for discriminating touches involves the use of some external device that the system can easily recognize. The DiamondTouch [5], one of the earliest tabletop systems to discriminate user touch, employs this approach. Seats around the table are modified to create a closed circuit between each user and the tabletop upon touch. This allows the system to identify users based on their seat positions. Several more recent methods take advantage of the hardware inherent to vision-based tabletops. Myer and Schmidt's IdWristband [12] and Roth et al.'s IR Ring [18] both emit coded light pulses as identifiers. Marquardt et al. [11] introduce fiduciary-tagged gloves to distinguish between users' hands. However, these techniques do not eliminate the need for external devices.

Several other approaches exist for associating touches with users or their positional proxies, often leveraging unique biometric traits. Holz and Baudisch rely on fingerprints [8] for identification, although accurate detection requires sophisticated sensors. Dohse et al. [6] identify a user's location by tracking the shape and color of users' hands using an overhead camera, which is less prohibitive, but requires peripheral hardware. Other types of equipment can be added to a standard tabletop to detect hand proximity, such as infra-red sensors [1].

Other research is appealing system because it uses only the existing hardware of a common vision-based tabletop. For example, Dang et al. [4] develop heuristics based on the positions and angles between multiple fingers to distinguish left and right hands, while Schmidt et al. [19] explore the contours of users' open palms to identify them for security purposes.

See Me, See You builds on this prior collection of results and, makes use of one predominant biometric, non-invasive trait for discriminating user touches. See Me, See You satisfies all of the design criteria for a lightweight system, demonstrating the effectiveness of FO as a simple yet accessible biometric feature to associate touch with user position.

SEE ME, SEE YOU: A TOUCH-DISCRIMINATE SYSTEM

See Me, See You was conceived to be a quick and easy method for discriminating user touches on common tabletop systems. Our central focus was to follow the aforementioned criteria for lightweight systems to create a method that can be easily re-implemented by others for multiple purposes.

Although the method behind See Me, See You depends on accurate determination of FO, the benefits of the technique are independent from any particular FO detection algorithm. Thus a system using See Me, See You could conceivably be implemented with any available algorithm or technique that can adequately capture a finger's orientation.

Once FO is accurately assessed, we associate user touches with user positions using a machine learning algorithm. We chose this method over a heuristic approach for ease of implementation and robustness due to the ability for such algorithms to generalize given limited training data. Although we chose a support vector machine (SVM) classifier for this purpose, the system may be implemented using any adequate classifier of the developer's choosing.

A Simple Camera-Based Finger Orientation Algorithm

Below we describe our FO algorithm, specifically designed for vision-based tabletop systems.

After implementation and testing, we found Wang et al.'s FO algorithm [22] to be insufficient for our purposes and developed another algorithm. Wang et al.'s algorithm requires users to touch in a motion they call an 'oblique touch', where the fingertip is placed first, followed by a roll to the finger pad. In the following studies, we found that the unintuitive nature of this oblique landing constraint made it unreliable without extensive user training. Roughly 20% of trials resulted in a finger orientation inverted by 180°.

Dang and Andre [4] present an algorithm that extracts FO values from user hand contours. They compare to other naïve algorithms and claim that their algorithm can achieve 94.87% recognition rate when the error tolerance is +\- 10 degree. Although their work is more generalized to support multiple fingers, they do not test their results in real time scenarios. We design a similar but simplified approach.

Our algorithm also relies on hand contours, which can be obtained with a standard DI setup [4], or with FTIR, given the modifications described next. To obtain clear and complete hand contours for our evaluation, we placed an overhead lamp above our FTIR table. To reduce obfuscation caused by the imbedded infrared light array, we introduce a relay into the IR lighting circuit to cycle the lights on and off. In this way we capture a precise hand silhouette image (Figure 2a) for each cycle of the FTIR vision server. We crop this image to 120×120 pixels around the coordinates of touch blobs, large enough to contain a whole hand but not so big as to decrease the system performance. We then detect the direction of the pointing finger from the hand contour image by examining a circular slice around the touch blob (Figure 2c). We chose a circular slice 5 pixels wide with an inner radius of twice the length of the touch blob's major axis. This method works well with a variety of men/women hand sizes. A line from the center of the remaining pixels within this range to the center of the touch blob (red line in Figure 2c) determines the FO angle. We chose FTIR to avoid early touch detection that can occur with DI systems.

Since finger orientations differ among fingers, we chose to restrict our exploration to the index finger. Although our algorithm could be modified to detect the orientation of other fingers, we feel that this restriction is not detrimental as studies have shown that most users extensively use their index finger on tabletops [10].

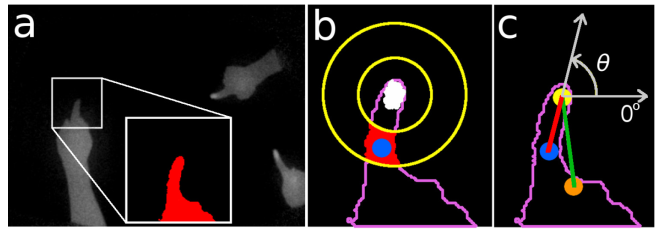

Figure 2. A silhouette of users' hands (a) is cropped and processed to find the contour of a touching hand. The contour is masked to reveal the area between two radii (b) around the FTIR touch blob received from the FTIR server. The finger orientation is given by a line (shown in red) from the touch blob to the center of the remaining area (c). A second line (in green) to the center of the hand contour determines if it is a left or right hand.

Detecting Handedness

In addition to detecting FO, our algorithm can detect the handedness of user touches. When the line for finger orientation is determined, a second line is derived from the hand contour extraction (green line in Figure 2c). In this case, it is from the touch blob to the centroid of all pixels in the extracted hand mass. Assuming that the user is pointing with their index finger, we can determine handedness with relatively high accuracy (~90%) by checking whether this second line lies to the left or right of the first line.

Discriminating User Touches

We classify FO patterns by user position using a multi-class support vector machine. SVM is a machine learning classifier that uses a set of training samples to create a mathematical function, or *model*, that can predict the correct category, or *label,* of a previously uncategorized item. We chose SVM because of its widespread reported success in a variety of problems. We use Chang and Lin's libSVM [2].

Training the System

To train the SVM, we collected user input data to create a set of labeled feature vectors (arrays of input values). Our feature vector contains the x-y coordinates of a touch and the corresponding FO angle, θ. For simplicity, we discretized the input space of the tabletop into 64 cells. The label of a feature vector is an integer representing the user's position around the table. Before training, we find the combination of required SVM parameters that give the highest cross-validation score.

Our model is user-independent, meaning that the training set includes data from multiple users and generalizes sufficiently to allow recognition of new users. User-independent systems are generally considered to be more difficult to implement than user-dependant systems, which are trained specifically to recognize one individual user. We find that a small training set from only a few users is sufficient to achieve fairly high cross validation scores (~95%). Although we would like to conduct a thorough investigation to find a minimal training sample set size, we leave it for future work.

Predicting a Touch's Owner

We used the SVM model to discriminate between user touches when the tabletop camera sees a touch point. To trigger a prediction, we construct an unlabeled feature vector for a detected touch, consisting of the *x-y* coordinates of the finger and its orientation, θ. When the feature vector is fed into the SVM, it returns the value of the predicted user position. The system only needs to trigger a prediction once; subsequent finger movement is tracked by the existing computer vision software.

With this approach, a different predictive model is required for each user configuration. However, we can use data collected from a few positions to extrapolate to others, and combine them into various configurations (see Study 2, below). Given the assumption that user pointing profiles are invariant to position, it may be possible to take an alternative approach that generalizes to any possible user position, for example a user standing at a corner. Likewise, the inclusion of multiple fingers from a single hand is likely possible. Extensions to dynamic hand configurations and those involving more than 3 users are left for future work. (See limitations and future work, below.)

EXPLORATORY STUDY: WILL FO WORK?

This exploratory study allowed us to investigate the distribution of 'natural' index finger placements across a tabletop and to contrast the profiles of various standing positions around the table. Our goal was to discover if FO patterns are distinctive enough to be useful as a feature for user touch discrimination. We used the collected data as training samples for an SVM classifier to determine the potential accuracy rate for predicting user positions.

Apparatus and Participants

We use a custom-built FTIR [7] tabletop with dimensions of 66 (length) × 51 (width) × 91 cm (height) (Figure 3). The tabletop uses infrared LED lamps emitting light with a wavelength of 850 nm using a 12 volt power supply and a Vivitek Qumi projector with a 1280 × 800 resolution and a brightness of 300 lumens. The experimental platform uses the TUIO protocol with the Community Core Vision (CCV) tracker [9], and runs on a 1.86 GHz Core 2 Duo PC with Windows XP. To cycle the LEDs for hand contour extraction (see Finger Orientation Algorithm, above), we use a Phidgets 3052 SSR relay board. The table's built-in IR camera captures a 640×480 image at a rate of 60 fps. Due to cycling the camera frames for alternate use by the CCV server and for hand contour analysis, our resulting frame rate is 20 fps. We use the same apparatus for all subsequent studies. Eight right-handed participants (all male) between the ages of 18 and 39 from a local university took part in this study.

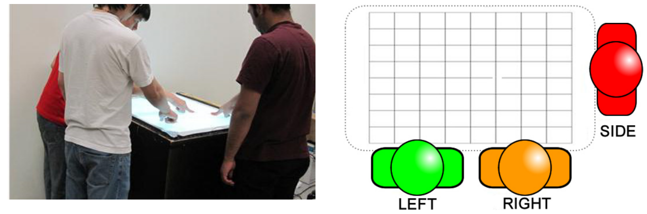

Figure 3. Left: Participants in an experiment around our custom-built FTIR tabletop; Right: dimensions of our system and the three positions for which we collected data to train our prototype system.

Procedure

We collected finger orientation data for various user positions from one participant at a time. The tabletop was divided into an 8×8 grid, with each cell measuring 9.1×6.2 cms. Our only instructions to the participants were to select targets, when they appeared, with their right hand index finger. The targets measured 3.4 × 2.9 cm and were placed at the center of a randomly selected grid cell. In the background we ran our FO algorithm and stored each orientation. We did not provide any additional visual or other type of feedback.

Participants selected a target in each cell, over two repetitions of all cells, while standing in each of three positions around the tabletop, LEFT, RIGHT, or SIDE (Figure 3). We only collected data from these three positions, as all other major positions around the tabletop could be extrapolated from these (discussed in experiment 2). We collected data from 8 participants × 3 positions × 64 target locations × 2 repetitions = 3072 trials. Each complete set of trials took approximately 45 minutes to complete.

Results and Discussion

Figure 4 shows the range of FO values for each cell in the grid. Each triangle represents the full range of finger orientations collected for the corresponding cell. The long midline depicts the mean value and the short line perpendicular to the midline shows one standard deviation from the mean. Following are some notable observations:

1 - Finger Orientation Ranges: Surprisingly, over 80% of all cells exhibit very narrow standard deviations. In about 90% (58/64 cells) of cases, the mean angles fall approximately in the middle of the detected angle range. Cells in front of the user tend to have narrower ranges than those that are off to either side.

2 - Range overlap: The ranges exhibit very little overlap. The LEFT and RIGHT (green and yellow, respectively) positions are nearly shoulder-to-shoulder, likely a worst case scenario. Despite this very close proximity, finger orientation ranges are distinct in over 95% of cells for side-by-side positions and in all cells for orthogonal positions (i.e. SIDE vs. LEFT or RIGHT). The standard deviations of the ranges do not overlap in any situation.

3 - Zones: Overlap between ranges appears to be greater in regions of the table that are either further away from pairs of users. Thus for objects directly in front of a user, their finger orientation is more distinct than in shared territories further away. We consider this factor in our evaluation.

These findings stem from participants using only their right hand. A mixture of both left and right hands would inevitably show more variability. However, since our FO algorithm can also detect handedness, we can first identify the handedness of a touch and then use the correct (left or right) profile to determine user position. We later asked the same participants back to collect their left hand profiles for further investigations, discussed in experiment 3.

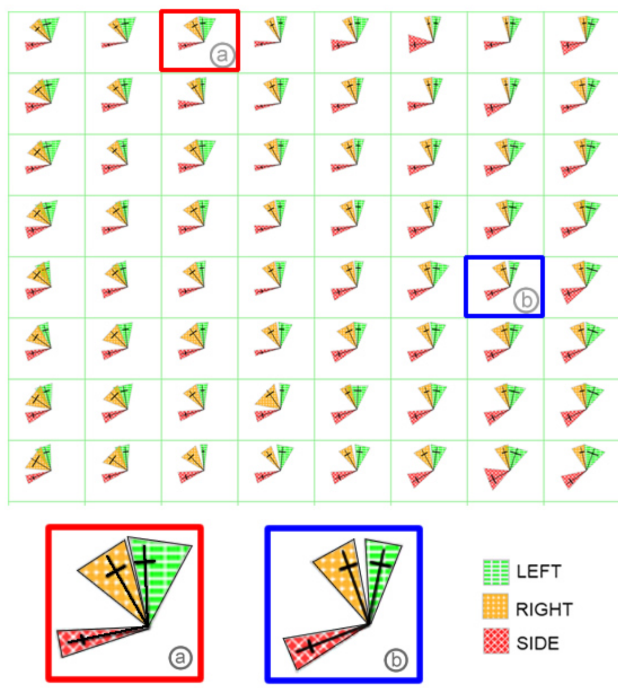

Figure 4. FO ranges across tabletop, with mean and standard deviation (Green: LEFT; Yellow: RIGHT; Red: SIDE). Two example cells are enlarged to show the distinct ranges.

Our observations led to the following hypotheses:

H1: Because of differences in range overlaps, See Me, See You will report higher accuracies for configurations where users stand in opposite or orthogonal positions than when standing adjacent to one another;

H2: Although training data were collected for targets at the center of each cell, the classifier will generalize across the entire cell, keeping accuracy high for unrestricted target positions;

H3: Since data were collected for a selection task, other tasks (such as rotating or scaling) that require users to place their fingers along different orientations will not be as accurate;

H4: Increased overlapping in cells that are more distant from a pair of users will lead to lower accuracy rates in those regions.

STUDY 2: SEE ME, SEE YOU'S ACCURACY

This study examined the accuracy of See Me, See You with a tapping task, common on tabletops for triggering a command or object selection. We wanted to test the robustness of our system with multiple users in a variety of possible configurations.

Participants and Procedure

Eight groups of 3 participants, between the ages of 20 to 35, participated in the study. Five of the 24 participants were female and all were right-handed. None had prior experience using a tabletop or participated in our first study.

The task was identical to the pointing task used in the exploratory study, with two exceptions. First, target positions were not restricted to the center of a grid cell, and second, the task was performed in groups, arranged in 1 of 4 predetermined standing configurations. Participants were instructed only to select their own target (specified by color) with their right hand index finger.

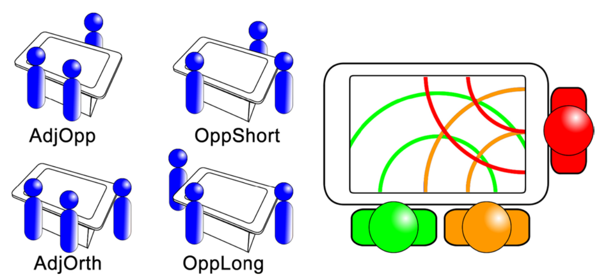

Figure 5. Left: Four configurations of standing positions relative to the table including side-by-side, opposite and adjacent users. Right: Targets were placed in 3 zones outlined by the same color as a user. The zones demarcated areas based on the distance to the right hand of each user.

Design

The experiment employed a 4 × 3 factorial design. The independent variables were *Configuration* and *Zone*:

Configuration: We chose a diversity of configurations that might appear in realistic situations. These include adjacent (side-by-side), opposite (across the long and short dimensions of the table), and orthogonal placements. The 4 configurations are labeled AdjOpp, AdjOrth, OppLong and OppShort (Figure 5 left).

Zone: The findings from the exploratory study showed a greater degree of overlap for regions that are far away from a pair of users. Therefore, we also tested our algorithm's accuracy based on the location of targets relative to each user's position. We defined 3 zones based on the distance to the user's right shoulder. The 3 zones are near (0-25 cm), middle (26-45 cm) and far (45 cm to the end of table) (Figure 5 right).

We presented an equal number of trials in each zone and for each participant. The *Configurations* were counter-balanced to reduce any learning effect. For each trial, a target was placed in a randomly chosen zone. There were a total of 12 targets per user in each configuration. The design can be summarized as 4 *Configurations* × 3 *Zones* × 12 *Trials* × 8 *Groups of 3 users* = 3456 trials in total.

Results and Discussion

The recorded data were analyzed using a repeated measures ANOVA test. The results, summarized in Figure 6, revealed an average accuracy of 97.9% across all the tested conditions. We found no significant effect of *Configuration* ($F_{3,21}$ = 0.858, p = 0.48) or *Zone* ($F_{2,14}$ = 3.47, p = 0.65), thus rejecting **H1** and **H4**. In **H4**, we hypothesized that See Me, See You's prediction accuracy would decrease in far-away regions, which showed more overlap between finger orientations. The results show that this is not the case. Likewise, **H1** can be rejected, since results were not significantly affected by user placement.

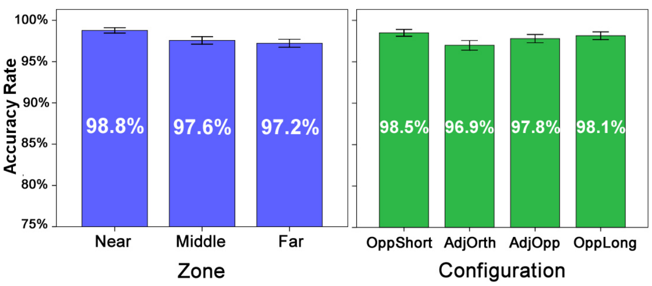

Figure 6. System accuracy based on zones (left) and configurations (right). Error bars represent 1 s.e. Scale starts at 75%, to show differences.

By *Configuration*, the accuracy rates were 98.5% (s.e. 0.4%) for OppShort, 96.9% (s.e. 0.9%) for AdjOrth, 97.8% (s.e. 1%) for AdjOpp and 98.1% (s.e. 0.9%) for OppLong. Notice that the accuracy of AdjOrth was slightly, but not significantly, lower than the others (Figure 5, 2nd from left). This was because there were more overlaps in finger orientations when participants were standing in this configuration. Although this study investigated situations with only 3 users, we believe our system is extensible, since we have tested the closest and most difficult configurations. Further testing with more users and other table sizes will be required to verify this conjecture.

Inspection of Errors

In inspecting the errors from the current study, we observed that many were caused by a failure of our finger orientation algorithm. Because our prototype uses overhead lighting to produce hand contours, some group situations can result in problematic overlapping shadows, for example, when a user's finger is occluded by a neighbor's arm. Figure 7 shows two such situations in which hand contour extraction failed. We expect that See Me, See You's accuracy can be increased with future FO detection methods or on systems that natively provide hand shadows (such as PixelSense [17]).

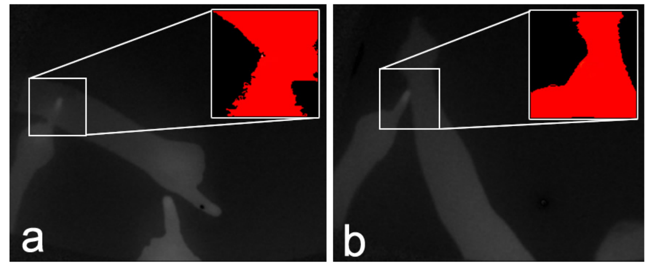

Figure 7. When a tap occurs (a) inside or (b) nearby the shadow of the other user's arm or hand, the algorithm failed to detect the correct FO.

STUDY 3: STEPPING UP COMPLEXITY

The previous study showed that See Me, See You is highly accurate across multiple user positions and when the targets are placed across the display, but only demonstrated this for the case of selecting objects. Real-world applications often involve more complex tasks. For instance, a user may want to rotate and scale a picture or draw on the table. These tasks may involve using both hands or may lead users to touch the table in a different orientation. See Me, See You relies solely on users' touch orientation. Prediction errors can result with users' changing their touching behavior, whether intentional or subconscious. However, we hypothesized that this issue could be resolved by educating users about how the system works so that they can adapt themselves to the system. We further hypothesized that such adaptation is effortless and welcomed by the users.

Participants and Procedure

We recruited 9 groups of 3 participants, each between the ages of 20 and 35, for this study. All 27 participants were right handed and 2 were female. Five had participated in Study 2, but none participated in the initial study.

We tested our system using three tasks involving the manipulation of a 7.8 × 9.8 cm object:

1. *Rotation with right hand (RR)*: Rotating an object is likely to produce some finger orientations (on land down) that do not coincide with what we used for training our algorithm. In this task, participants were restricted to using their right hand only.
2. *Rotation with either hand (RE)*: This is the same task as the one above except that participants were allowed to use either hand to rotate the object.
3. *Scaling (S)*: This task requires participants to use both of their index fingers to tap on a rectangular object, and drag in opposite directions. This task would further test the limits of our trained system as well as the accuracy of our handedness detection algorithm.

In task 1, hand prediction is unnecessary and thus all inputs were passed to the correct FO model, allowing us to evaluate our handedness detection algorithm again. The two-handed tasks test the system under more realistic conditions. In these tasks, all inputs were first evaluated for handedness and then passed to the appropriate model for user touch discrimination.

Design

The experiment consisted of 2 phases. The 1st phase imitated a walk-up-and-use scenario, where participants performed the 3 tasks without any knowledge of how the system works. The 2nd phase started with a short orientation session (about 5 minutes long), where participants were informed about how the system works. In this phase only, participants received feedback during the 3 tasks about whether the system correctly recognized them. A colored arrow was shown, along with a smiley face for correct predictions, or a sad face for incorrect ones. Participants were given practice trials until they understood the meaning of the feedback and had learned to avoid situations that commonly caused recognition failure, such as shadow occlusion or extreme FO angles. They were not allowed to correct their FO, even if an error occurred; participants were not instructed in either phase about how and where to place their index finger in a target.

Participants were asked to stand in the AdjOrth configuration, which produced the lowest accuracy in Study 2. In each trial, 3 targets, color-coded by user, were placed simultaneously in random positions. A small offset distance was used to ensure that targets did not overlap with each other or appear too close to the edge of the table.

The experiment employed a 3 × 2 within-subject factorial design. The independent variables were *Task* (RR, RE, and S); and *Feedback* (feedback or non-feedback). *Task* was partially counter balanced, however the non-feedback phase was always presented first. We allowed short breaks between tasks and phases. Participants filled out a questionnaire upon completion.

Results and Discussion

For all the 3 tasks, the recognition of user position was made based on the initial touch of an object. For the scaling task, we used the FO from whichever hand touched the object first. The resulting data were analyzed using Repeated-Measures ANOVA and Bonferroni corrections for pair-wise comparisons.

The results revealed an average accuracy of 94.7% across all the tested conditions. ANOVA tests yielded a significant effect of *Feedback* ($F_{1,8} = 5.7$, $p < 0.05$). There was no significant effect of *Task* ($F_{2,16} = 0.74$, $p = 0.49$).

The system had higher accuracy in the feedback condition (96.5%, s.e. 0.4%) than in the non-feedback condition (92.8%, s.e. 1.5%). We found no significant learning effect during the 1st phase, suggesting that this difference was primarily due to the following orientation session. When broken down by task, we find accuracies of 95.8% (s.e. 0.6%), 94.4% (s.e. 1.8%), and 93.7% (s.e. 1.1%) for RR, RE and S, respectively.

Effect of task complexity

Although analysis did not yield a significant effect of task complexity, one-way ANOVA tests showed a significant difference between the 3 tasks in the non-feedback condition ($F_{2,1941} = 4.57$, $p < 0.05$). RR had the highest accuracy (95.1%, s.e. 0.9%), followed by RE (92.6%, s.e. 1%), which was higher than S (90.7%, s.e. 1.1%) (Figure 8 left). Post-hoc analyses showed only a significant difference between RR and S ($p < 0.01$).

Many of the errors in the 1st phase were a result of overlapping shadows that interfered with FO detection (as shown in Figure 7). Task S had the highest number of these errors because 2 hands per user resulted in more overlapping arms. Additionally, in this task, users would often place their hands with the index finger parallel to an object's edge to avoid occlusion. As predicted, the system accuracy decreased with increasing task complexity (between RR and S), confirming **H3**. We assume that the knowledge and feedback reduced this effect in phase 2.

Effect of Feedback

In the feedback condition, system accuracy increased to 96.6% (s.e. 0.7%), 96.3% (s.e. 0.7%), and 96.6% (s.e. 0.7%) for RR, RE, and S, respectively (Figure 8). Pairwise comparisons showed a significant improvement over the non-feedback condition for all the tasks except RR ($p < 0.01$). These results suggest that by understanding the causes and recognizing instances of problems, users were able to adapt and improve their experience.

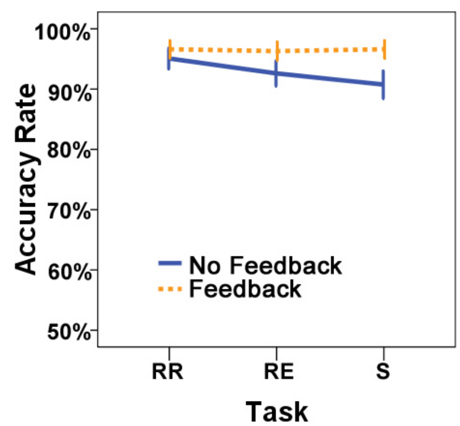

Figure 8. System accuracy shown by task and feedback (graph starts at 50%).

Accuracy of the handedness detection algorithm

For evaluation of handedness detection, we use only trials from the RR task, in which hand use was controlled. In this task, the right hand was correctly determined 93.8% of the time (Figure 9). We feel it is reasonable to expect a similar accuracy for detecting the left hand. Within the set of trials for which handedness was correctly recognized, user positions were also predicted correctly in 95.6% of cases. Interestingly, even when handedness detection failed, user identification remained high at 91.3% (Figure 9).

Subjective preference

The post-experiment questionnaire shows that users welcome See Me, See You as an easy-to-use plug-in for existing tabletop applications. All scores reported below are based on a 5-point Likert scale (5 for highest preference).

The participants gave an average score of 4 in support of user feedback. Of all participants, 85% agreed that the feedback helped them learn from mistakes, and better adapt to the system. When asked "*Did you change your finger direction after knowing how the system worked?*", they responded with an average of 3.1. They reported an average of 1.8 when asked if they felt it was uncomfortable to change their FO. In most cases, however, such a change was not necessary; only 1 user (3.7%) gave a positive score (of 4) when asked if the required number of corrections was excessive. Participants also gave feedback regarding our UI design, with 78% in support of showing the detected FO in addition to the visual feedback of the recognition result. This motivated our design of the Position Aware Cursor, which we describe in the following section.

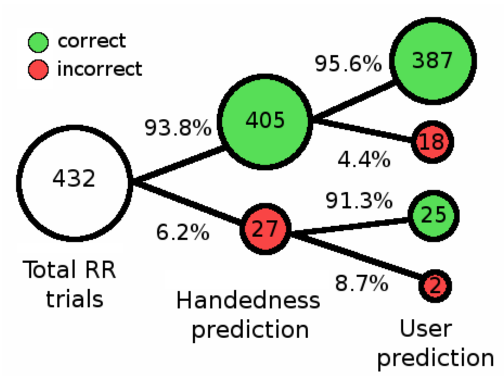

Figure 9. System accuracy for hand and user predictions. User prediction is still high even when handedness prediction fails.

USER MOBILITY AND FLUID ERROR RECOVERY

Our results suggest that the robustness of See Me, See You will allow the design of multi-user features on a common tabletop. We enhanced See Me, See You with two additional features. The first allows users to move around the table and the second allows for a fluid method of correcting prediction errors. Both features are compatible with the lightweight requirements outlined earlier.

Position Avatar

To grant users the flexibility of moving around the table, we associate each user with a *Position Avatar*. Users log in to the system by selecting a Position Avatar icon. Thereafter, the icon indicates their position at the tabletop edge. When a user chooses to changes positions, she can drag the Position Avatar along. In this implementation, the onus is on the user to manually inform the system of their movements. Although a more sophisticated device could automatically track the user with peripheral hardware, we resorted to manual placement to maintain the lightweight nature of See Me, See You.

PAC: The Position Aware Cursor

Error recovery is an instrumental feature of a lightweight system, as it may not always guarantee 100% accuracy. Inspired by comments from our participants, we designed the Position Aware Cursor (PAC, Figure 10) to provide users with a fluid and robust solution in cases of wrong predictions. PAC has two elements: (1) A color-coded arrow showing the user's touch orientation, and (2) a set of wedges showing the possible FO ranges available, based on the locations of other users. In this example, the angle and direction of these wedges are based on the data collected in our exploratory study (Figure 4). If an incorrect prediction occurs, the user can re-orient her finger to a new wedge. We envision that such a feature could be disabled when a user becomes acquainted with the technique.

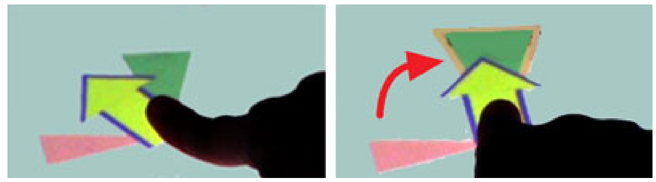

Figure 10. The Position Aware Cursor. Left: a user lands her finger, and the system predicts her location correctly. Right: the user rotates her finger to changes her identity.

SUBJECTIVE IMPRESSIONS OF SEE ME, SEE YOU

In a final informal evaluation we collected subjective user feedback with See Me, See You in two prototype applications: a multi-user paint application and a game.

Three groups of 3 participants (2 females), between the ages of 21 and 30, participated in this evaluation. With the paint application, participants were asked to collaborate and replicate a sample drawing. This required that they each control certain user-specific states such as line thicknesses and color. Each participant completed ⅓ of the drawing. In the multi-user game, participants were asked to quickly find and select two tiles with matching graphical patterns. Tiles could occlude and overlap one another, thus requiring participants to move tiles around the table. Users were given a score based on the number of pairs they matched and the game ended when all tiles were selected. We encouraged participants to use PAC for error corrections.

We note the following observations: (1) Participants finished the tasks relatively quickly, and were not hindered by any system features. (2) In informal interviews, participants indicated that they appreciated the multi-user capabilities of See Me, See You, and mentioned that they preferred them to taking turns to carry out the same tasks. (3) They appreciated that they were not required to wear peripherals or hold a pen for user identification. (4) Two participants mentioned that they used PAC to correct errors. (5) Participants found that PAC helped them understand the method by which the system associated touch with user position. (6) Interestingly, one participant commented that the only concern he had with See Me, See You was the inability to move from one position to another. We then allowed him to try out the Position Avatar, of which he reported satisfaction. (7) Two participants from one group suggested that such a system could be implemented by recognizing their fingerprints. Given the technical challenges and hardware requirements for fingerprint recognition with current technology, See Me, See You is an ideal alternative for distinguishing multiple users' touches.

DISCUSSION

Overall, our results are highly encouraging and confirm the potential of See Me, See You as a viable approach for multi-user capabilities on common vision-based tabletop systems. We highlight some of our primary findings.

Reliability across entire tabletop. Overall, the SVM classifier is robust in our application. Although our training set is collected on only 64 target locations, the system is able to classify interactions across the entire continuous table space (confirming **H2**).

Accuracy across tasks. See Me, See You responds well to untrained finger orientations that result from non-pointing tasks as well as from awkward approaches when user reach around one another during simultaneous interaction. Feedback further improves the prediction accuracy.

Generalizing to users. The system easily generalizes to new users who did not contribute to the training data set. This type of generalization is typically a difficult problem in machine learning, but is possible in our approach because of the distinct ranges of FO values across multiple users.

User configurations. As expected, there is a slight penalty in prediction accuracy for adjacent users sharing a table edge. This is due to adjacent users exhibiting the most amount of overlap in FO. However, the loss was smaller than we expected as we did not find any significant differences in accuracy across different user configurations.

User adaptation. Another interesting observation was the willingness and ability for users to adapt to the system. We found higher success rates when users were told how the system operates. Users were comfortable in altering their finger landing orientation to make the system work even more effectively. Users also reported that they did not feel any additional cognitive or motor effort than when they were not given any system knowledge. Furthermore, groups displayed an eagerness to cooperate, by adjusting their hand position to make room for others and by taking turns when simultaneous selection was impractical, thus exhibiting common courtesy.

Complementarity. See Me, See You could work as either a stand-alone system or one that could be used in conjunction with other methods, as in [1, 3, 6]. For example ceiling mounted cameras can provide some information about users interacting around a tabletop. In areas of high occlusion, where cameras may not properly detect certain actions, the system could resort to using See Me, See You.

Recommendations to Designers

Our exploration of FO profiles highlights some important implications for designers:

- People appear to produce consistent finger orientations, at least within a restricted demographic. FO is easiest to distinguish in selection tasks, but is also reliable in more complex situations. There is also potential for FO in contexts other than user discrimination.
- Locations that are on orthogonal and opposite sides of the table can be distinguished with a very high reliability. One user per side is an ideal configuration, but designers should not deter from using this feature in more crowded conditions.
- The Position Aware Cursor is a fluid and easily implementable feature that can improve the reliability and robustness of a touch-discriminate system.

Limitations and Future Work

Finger orientation is a natural attribute that designers can make use of to discriminate user touches. Improvements to our technique will be necessary for See Me, See You to be used in the wild, however, in exit surveys, most of our participants responded positively when asked if the system is accurate enough for real-world use. Our study also opens up a number of possibilities for future exploration:

User position. See Me, See You does not directly identify users and cannot detect movement. The use of user position as a proxy for the actual user and our Position Avatar provide a good compromise over methods such as overhead cameras or outward-facing infrared range sensors which would limit the lightweight nature of our system.

Position Profiles. We collected FO profiles for specific positions around the table. This may suffice for many applications, however, the fullest potential lies with fewer restrictions. It should be possible to generalize our approach to accommodate untrained profiles, for example a user standing at a corner. However, additional hardware might be required to track a user's position and orientation.

Multiple Fingers. We collected profiles for the index finger only. Our system can be extended using existing algorithms (e.g. [3, 4]) to detect multiple fingers from the same hand.

Number of users. Our studies investigated situations with up to three users. We believe that our system is extensible to more users using more advanced FO algorithms.

FO algorithm. Most of our errors stemmed from our finger orientation algorithm. We expect that future systems will have bullet-proof methods for capturing finger orientation. Furthermore, secondary biometrics such as finger pressure could be leveraged to increase the accuracy of our system close to 100%.

Impersonation. PAC is a valuable tool for error recovery, but could assist mischievous users in impersonating others. In most group situations, however, there is nothing to gain by impersonation. Also, social protocols, such as courtesy, or fear of being rejected by the group, might mitigate such issues. Future study outside a lab environment would provide further insight on this matter.

Target smaller devices. Testing with multiple users using a smaller device will be needed to determine how well FO works in platforms other than tabletops. Because smaller devices, such as tablets or smartphones, are more mobile compared to tabletops, a lightweight touch discrimination technique will be highly desired.

CONCLUSIONS

In this paper, we have presented See Me, See You, a simple, yet flexible and accurate, approach to discriminating user touches on tabletops. We have introduced a new technique for capturing finger orientation. We have demonstrated that finger orientation profiles are quite uniform around a tabletop and can be used reliably for identifying user locations. Results from our experiments have indicated that See Me, See You performs accurately in tasks of varying complexity across different configurations of user locations around a tabletop. We have also introduced two enhancement techniques for multi-user applications: Position Avatar and Position Aware Cursor. With these two techniques, users can change locations and perform self-correcting actions in a fluid manner, without interrupting their activity. In conclusion, See Me, See You is a viable lightweight solution for providing simple yet effective support for multi-user application features on tabletop surfaces.

ACKNOWLEDGEMENTS

We thank our study participants. We also thank our lab mates for their valuable feedback and acknowledge NSERC for partially funding of this project.

REFERENCES

1. Annett, M., Grossman, T. and Fitzmaurice, G. Medusa: A proximity-aware multi-touch tabletop. In *Proc. UIST '11* 337-346.
2. Chang, C.-C. and Lin, C.-J. LIBSVM--A Library for Support Vector Machines. http://www.csie.ntu.edu.tw/~cjlin/libsvm.
3. Dang, C.T., Straub, M. and Andre, E. Hand distinction for multi-touch tabletop interaction. In *Proc. ITS'09*, 101-108.
4. Dang, C. T. and Andre, E. Usage and recognition of finger orientation for multi-touch tabletop interaction. In *Proc. INTERACT'11*, 409–426.
5. Dietz, P.H. and Leigh, D. 2001. DiamondTouch: A multi-user touch technology. In *Proc. UIST'01*, 245-252.
6. Dohse, K.C., Dohse, T., Still, J.D. and Parkhurst, D.J. 2008. Enhancing multi-user interaction with multi-touch tabletop displays using hand tracking. In *Proc. ACHI'08*, 297-302.
7. Han, J.Y. Low-cost multi-touch sensing through frustrated total internal reflection. In *Proc. UIST'05*, 115-118.
8. Holz, C. and Baudisch, P. The generalized perceived input point model and how to double touch accuracy by extracting fingerprints. In *Proc. CHI'09*, 581-590.
9. Kaltenbrunner, M., Bovermann, T., Bencina, R. and Costanza, E. TUIO. http://www.tuio.org.
10. Kin, K., Agrawala, M. and DeRose, T. Determining the benefits of direct-touch, bimanual, and multifinger input on a multitouch workstation. In *Proc. GI'09*, 119–124.
11. Marquardt, N., Kiemer, J. and Greenberg, S. What caused that touch? Expressive interaction with a surface through fiduciary-tagged gloves. In *Proc. ITS'10*, 139-142.
12. Meyer, T. and Schmidt, D. IdWristbands: IR-based user identification on multi-touch surfaces. In *Proc. ITS'10*, 277-278.
13. Morris, M.R., Paepcke, A., Winograd, T. and Stamberger, J. TeamTag: Exploring centralized versus replicated controls for co-located tabletop groupware. In *Proc. CHI'06*, 1273-1282.
14. Müller-Tomfelde, C. (ed.) 2010. Tables – Horizontal Interactive Displays. Springer.
15. Partridge, G. and Irani, P. IdenTTop: A flexible platform for exploring identity-enabled surfaces. In *Proc. CHI EA'09*, 4411-4416.
16. Piper, A.M., O'Brien, E., Morris, M.R. and Winograd, T. SIDES: A cooperative tabletop computer game for social skills development. In *Proc. CSCW'06*, 1-10.
17. PixelSense. http://www.samsunglfd.com/product/feature.do?modelCd=SUR40, accessed on January 2012.
18. Roth, V., Schmidt, P. and Güldenring, B. The IR Ring: Authenticating users' touches on a multi-touch display. In *Proc. UIST'10*, 259-262.
19. Schmidt, D., Chong, M. K. and Gellersen, H. IdLenses: Dynamic personal areas on shared surfaces. In *Proc. of ITS'10*, 131-134.
20. Schoning, J., Hook, J., Motamedi, N., Oliver, P., Echtler, F., Brandl, P., Muller, L., Daiber, F., Hilliges, O., Loechtefeld, M., Roth, T., Schmidt, D. and von Zadow, U. Building interactive multi-touch surfaces. *Journal of Graphics, GPU, and Game Tools*, 14, 3 (2010), 35-55.
21. Scott, S.D. Territory-based interaction techniques for tabletop collaboration. In *Proc. UIST'03*, 13-20.
22. Wang, F., Cao, X., Ren, X. and Irani, P. 2009. Detecting and leveraging finger orientation for interaction with direct-touch surfaces. In *Proc. UIST'09*, 23-32