Enhancing Multi-user Interaction with Multi-touch Tabletop Displays using Hand Tracking

K.C. Dohse [1], Thomas Dohse [2], Jeremiah D. Still [1], Derrick J. Parkhurst [13]
Human Computer Interaction Program [1], Computer Science Department [2], Psychology Department
Iowa State University
{ kcd | dohse | derrick | jeremiah } @ iastate.edu

Abstract

A rear-projection multi-touch tabletop display was augmented with hand tracking utilizing computer vision techniques. While both touch detection and hand tracking can be independently useful for achieving interaction with tabletop displays, these techniques are not reliable when multiple users in close proximity simultaneously interact with the display. To solve this problem, we combine touch detection and hand tracking techniques in order to allow multiple users to simultaneously interact with the display without interference. Our hope is that by considering activities occurring on and above a tabletop display, multiuser interaction will become more natural and useful, which should ultimately support collaborative work.

1. Introduction

Large displays are useful for information visualization when multiple people must jointly use the information to work together and accomplish a single goal. The social interactions that result from using a shared display can be highly valuable [1]. However, these displays can fail to allow multiple users to simultaneously interact with the information.

Tabletop interfaces can provide a large shared display while simultaneously accepting natural and direct interaction from multiple users through touch detection. For example, multi-touch surfaces that utilize the phenomenon of frustrated total internal reflection (FTIR) have received widespread attention recently [2]. FTIR detection techniques allow the system to track a large number of simultaneous touch points with very high spatial and temporal frequency. The FTIR has several advantages over other multi-touch detection technologies, such as being both low cost and scalable [3]. Other computer vision based tracking systems have a limited ability to detect touches versus near touches which is an important element of interacting with the table surface.

FTIR tracking alone has two shortcomings compared to other methods of touch tracking. Each touch in FTIR appears as an independent event. Although inferences based on distance between touch points can be leveraged to guess which touches are part of the same event, each touch ultimately remains a standalone piece of data. As the number of users and complexity of their actions increases, so does the probability of incorrectly grouping touch points with a single user. The other issue is that the system is inherently susceptible to problems with spurious IR noise (e.g. poor lighting conditions or flash photography).

To solve lighting and touch differentiation problems, we augmented a FTIR tabletop display with an overhead camera. Using the camera, hands on or over the table can be tracked using skin color segmentation techniques. With hand coordinates available, touch points can be assigned the ownership necessary to support multiple users and correctly identify events comprised of multiple touches. This technique works well even when gestures are made by multiple users in close proximity because it does not need to differentiate touches based on closeness. The fusion of hand and touch point locations also increases the robustness of touch sensing in the presence of unwanted IR light because of the redundancy of the point's location.

Additionally, tracking hands allows users to generate interactions without touching the surface, but rather making movements above the table. This creates a hybridization of the two interaction techniques that is still being explored.

Figure 1: Three users working together using a rear-projection multi-touch tabletop display augmented with hand tracking using an overhead camera.

2. Related Work

Techniques and technologies used for interaction detection on tabletop displays are rapidly maturing, but many researchers are still seeking better methods for capturing natural interactions made by multiple users within the context of real world applications. There are a number of approaches to tracking user interactions with a tabletop display. One successful method is to use a surface material that is laden with sensors, such as the commercially available DiamondTouch system. This system uses a technique where a circuit is capacitively closed when the user touches the table [4]. Interfaces like this one use front projection due to the opaque surface needed for the sensors. Other systems such as the metaDESK also use sensors, but integrate them in physical objects that can be manipulated [5].

Another common approach is to use video cameras to track interactions. For example, the HoloWall uses a semi-opaque diffuser that allows infrared (IR) light projected from behind the screen to reflect off of objects at a certain distance from the surface [6]. The TouchLight interface uses two IR cameras to determine when contact with the screen has occurred [7]. Other projection based systems, like the ViCat use overhead cameras to track hand gestures [8]. This table does not use physical touches on the surface, but uses an overhead camera to track hand gestures in order to interact with the display.

Work is also being done to improve the nature of multi-touch interaction. These areas of research are equally vital to the field as designing new systems to support multi-touch. Such as designing cooperative gestures to facilitate teamwork [9]. Other research has aimed to build a framework for designing and evaluating multi-touch interactions [10].

3. System

We developed a rear-projection tabletop display with multi-touch sensing capabilities using the principles of frustrated total internal reflection (see Figure 1). A 4’ x 3’ acrylic diffusing sheet is mounted at waist level and projected on from below through a series of mirrors in order to create a rear-projection tabletop display. Infrared light entering the sides of a second acrylic sheet mounted just above the diffusing sheet is totally reflected at the acrylic-air border preventing any light from escaping the surface. When touched, infrared light is able to leave the acrylic creating a spot of light below each point a finger is in contact with the acrylic. Touch points can then be tracked using an infrared camera in the same optical path as the projector. A visible spectrum camera is mounted above the display for hand tracking via skin detection.

3.1 Touch Detection

Each frame captured by the camera has a known base image subtracted from it and is then converted to a binary image by setting each pixel to black or white based on a user defined threshold. The center of non-contiguous regions in the binary image are extracted as touch locations. Since touches are viewed as atomic, largely analogous to a mouse click, the reduction of a circle of a light to a single point makes sense from a design philosophy perspective as well as minimizing memory and network usage. These locations are then compared to the known locations of all other touches to determine if they represent a new touch or the movement of a finger already in contact with the screen. When new points of light appear, there is a short period where the point is tracked but kept in limbo and not reported to the rest of the system. This minimizes the risk of random environmental noise from interfering with an application. Similarly, existing points which are being tracked are kept briefly in limbo after they disappear in order to continue tracking a point even if it is not visible for a few frames. Each touch is assigned an identification number which it keeps until the finger is lifted from the screen. When no fingers are in contact with the screen, the system resets the identification list and the next touch will be given the identification number of zero.

3.2 Hand Detection

The use of skin color segmentation has proved successful in a wide range of applications [11, 12]. Yang and Waibel have presented a successful skin-color algorithm that tracks skin in real time [13, 14]. They are able to achieve skin tracking by dimensional reduction of the available color space. Further, the dimensional reduction is achieved by targeting the clustered normalized skin colors. According to Yang, Weier and Waibel (1997), color differences between people appear within the intensity dimension rather than color [15]. This dimensional reduction to target skin normalized color is an effective method for skin detection. Thus, we adopted their method of color space dimensional reduction to effectively target spatial regions likely to be skin.

Hands are detected by creating a binary image of the pixels which fall within the appropriate RGB (red, green, blue) ranges. Each contour of the binary image is found using the cvFindContours function provided by opencv. Very small contours are assumed to be noise in the image and are discarded with the remaining elements assumed to be the hands and arms of users. Three pieces of data, the "fingers" point, the "table edge" point and an integer value representing the side of the table, are extracted from the contour and passed over the network to be used in applications. This data is acquired by iterating over the points in each contour to find the points with minimum and maximum Y-axis values. For simplicity we will refer to the two primary opposing sides of the table as side one and side two. Users standing on side one will create contours which have very low Y-axis values at one end and middle-ranged Y-axis values at the opposite end of their arm contour. In this case, the low-value is known to be the "table edge" point and the high-value point is known to be the "fingers" point. Conversely, users on side two will create contours with very high Y-axis values at the table edge and middle-ranged Y-axis values at their fingertips. Similarly, the system can easily be extended to recognize users on the remaining two sides by doing comparisons with X-axis values.

Due to the table often operating in low lighting and each application causing substantially different colors and intensities on screen, all the hand detection parameters for color ranges and sizes are adjusted at runtime. A limitation of the system worthy of note is that to facilitate accurate tracking of hands, applications need to maintain a reasonably constant color and intensity across the screen. Map based applications, virtual board games and workspaces for collaborative planning, command and control are not adversely affected by this limitation and comprise the focus of our research.

3.3 Event Generation

The touch screen interface requires a more complex event handling system than a standard mouse and keyboard interface since it must allow applications to treat multiple touches as belonging to a single gesture event. To facilitate robust event recognition from camera tracking data, a publisher-subscriber design pattern was implemented to distribute event generation responsibilities over multiple functions, allowing each function to have a very limited scope. Each subscriber checks if the raw input from the computer vision components of the system meet its specific criteria to generate an event. To allow each application to control the events the system recognizes and to have multiple modes of operation within an application, subscribers are dynamically added and removed at runtime by the client application. Client applications need only handle events. The functions that recognize events can be treated as a black box.

Data from both the overhead skin tracking and infrared touch tracking is handled simultaneously by data listeners, allowing for events to incorporate both data sources into their event detection criteria.

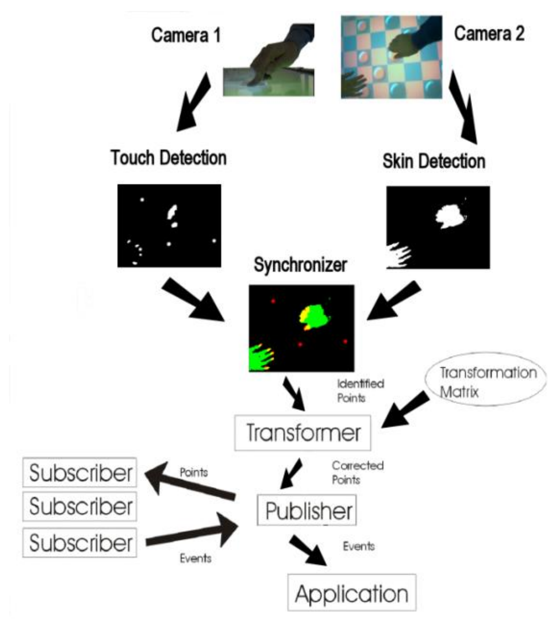

Figure 2: System Dataflow; Red in the sychronization phase indicates erroneous points removed.

4. Results

Tracking hands creates two important functionalities, the first is the ability to associate a touch with a user. Multi-touch interfaces inherently accept interaction from multiple users, however, it has been impossible so far to discriminate between users in the discreet manner of computer vision skin tracking. User identities are maintained by the overhead tracking information and are not dependent upon an individual's skin tone. By having one user standing on each side of the table, each touch can be associated with a corresponding user with very little risk for error. With the added information of which hand created a touch, role specific functions can be given to different users for increased collaborative power. Users should avoid overlapping hands to prevent occlusions in the overhead camera's line of site, which may interfere with tracking. The other main benefit is that because a hand is continuously tracked, multiple touches created by the same hand can be treated as part of the same event. By doing this, touches no longer need to be treated as singular events, but can be given a history and associated with previous touches. Furthermore users can now interact with objects on the table without physical contact. One such usage was to have a checker float in front of the users hand after touching it in a simulation of the checkers board game.

Additionally, hand location data can be used to increase the touch sensitivity without increasing error. The grayscale image is converted to binary by comparing each pixel with a variable threshold value. A lower threshold creates more false positives due to random noise in the image. By tracking hand locations the system automatically ignores erroneous touch data by removing touch locations with large distances from any hand.

To allow the system to simultaneously track multiple hands, touches, and run computationally intensive programs, we adopted a distributed architecture so that each camera is connected to a separate computer, each of which communicates over a standard network connection. This allows the system to be responsive to user actions with minimal delay while maintaining a high frame rate in the application using only average hardware.

Both cameras are low cost webcams running at a 640 x 480 resolution with frame rates of approximately 25 fps for touch tracking and 16 fps for overhead hand tracking under regular usage.

The representation of touches as single points based on their center and hands as finger location points and table edge points offers enough precision for actions such as button presses while minimizing memory and network usage.

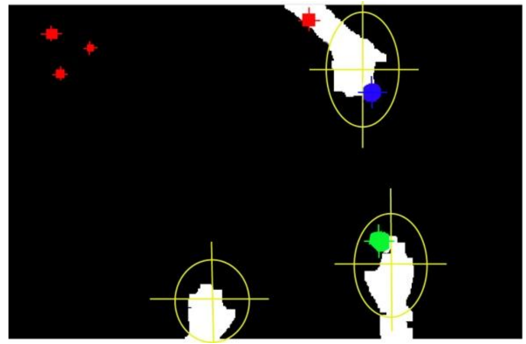

Figure 3: Red reflects touch points determined to be noise; Blue shows user A touch; Green shows user B touch; White reflects detected skin; Yellow circles indicate the acceptable spatial region for a touch to occur

5. Conclusions and Future Work

Augmenting a FTIR multi-touch table with computer vision skin tracking improves collaboration, expands the methods of interaction and increases touch tracking robustness. The overhead camera is able to discreetly track multiple users by establishing identities for each individual when they make their initial actions and tracking them for the remainder of their usage.

Maintaining user identities has two important aspects: preventing users from inadvertently interfering with each other and delegating role specific functions to different users. When two or more people are using a FTIR interface, touches can be confused by the system and unintentional actions can be performed, but this is overcome when the system can differentiate touches between users. In many collaborative projects it is common for people to have different specialties and assignments, efficiency is increased by allowing different users to interact with the same object or region and have different actions performed.

The system is also able to recognize gestures that occur above the table, therefore interactions are not relegated to physical contact with the table. Objects on the table can be picked up and moved in a very natural manner because the object follows the hand that selected it. Furthermore, a whole new assortment of natural gestures can be integrated, reducing the amount of cumbersome gestures a user needs to learn.

This technology was originally developed for, and is showcased well, within the dynamic environment of military command and control. While planning and reviewing missions, the augmented FTIR multi-touch table can mimic the traditional sandbox interface in virtually every way, while providing additional features impossible to recreate in the real world. The fast paced and multi- faceted nature of live mission command and control can utilize all of the advantages our table offers. For example, while managing a group on unmanned aerial vehicles there are many activities that need to be performed by soldiers with different individual objectives, but all of the soldiers have a common goal and can greatly benefit from the shared space that a table provides. User tracking and role assignment effectively acts as a gesture multiplier and ensures that no soldier can interfere with another's mission goal in this time critical activity.

Future work will consist of adding stereo to hand tracking, automatically calibrating skin color parameters, incorporating tangible objects and usability testing. A second overhead camera will be added so that above table gestures will be tracked in three dimensions by determining the disparity between the two images. To allow more robust skin detection, the system will calibrate the skin color parameters based on the colors being projected on the screen, rather than predefined values. Tangible objects can be integrated because the skin tracking algorithm can be readily adjusted for any color or intensity, thereby making objects placed on the surface distinguishable from projected images. Current interactions have been fairly basic; usability research will be conducted to improve the merging of touch and above table interactions.

Our goal is to continue to advance multi-touch technology by removing as many barriers between the user's intentions and the interfaces input capabilities as possible. We believe the research efforts discussed in this paper remove some existing barriers and provide users with a more intuitive experience than multi-touch or hand gesture interaction alone.

6. Acknowledgements

We thank the Air Force Office of Scientific Research (AFOSR) for supporting this research.

7. References

[1] Sara A. Bly and Scott L. Minneman. Commune: A Shared Drawing Surface. In SIGOIS Bul/etin, Massachusetts, 1990, pp.184-192.

[2] Kasday, L. (Nov. 1984). Touch Position Sensitive Surface. U.S. Patent 4,484,179.

[3] Han, J. Y. (2005). Low-Cost Multi-Touch Sensing through Frustrated Total Internal Reflection. In Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology.

[4] Dietz, P. H., & Leigh, D. L. (2001). DiamondTouch: A Multi-User Touch Technology. In ACM Symposium on User Interface Software and Technology (UIST), 219-226.

[5] Ullmer, B. and Ishii, H. The metaDESK: models and prototypes for tangible user interfaces. Proceedings of the 10th annual ACM symposium on User interface software and technology, ACM Press, Banff, Alberta, Canada, 1997.

[6] Matsushita, N., & Rekimoto, J. (1997). HoloWall: Designing a Finger, Hand, Body and Object Sensitive Wall. in ACM Symposium on User Interface Software and Technology (UIST).

[7] Wilson, A. (2004). Touchlight: An Imaging Touch Screen and Display for Gesture-Based Interaction. In Proceedings of the International Conference on Multimodal Interfaces. p. 69-76.

[8] Chen, F., Close, B., Eades, P., Epps, J., Hutterer, P., Lichman, S., Takatsuka, M., Thomas, B., and Wu,M., “ViCAT: Visualization and Interaction on a Collaborative Access Table”, in Proc. IEEE Workshop on Horizontal Human-Computer Systems (2006), 59-60.

[9] Morris, M., Huang, A., Paepcke, A., and Winograd, T. (2006). Cooperative gestures: Multi-user gestural interactions for co-located groupware. Proceedings of the ACM CHI Conference on Human Factors in Computing Systems. pp. 1201-1210.

[10] Wu, M., Shen, C., Ryall, K., Forlines, C., and Balakrishnan, R. (2006). Gesture registration, relaxation, and reuse for multi-point direct-touch surfaces. Proceedings of IEEE TableTop - the International Workshop on Horizontal Interactive Human Computer Systems. pp 183-190.

[11] Ohta, Y., Kanade, T., & Sakai, T. (1980). Color information for region segmentation. Computer Graphics and Image Processing, 13(3), 222-241.

[12] Swain, M. J., & Ballard, D. H. (1991). Color indexing. International Journal of Computer Vision, 7(1), 11-32.

[13] Yang, J., & Waibel, A. (1995). Tracking human faces in real-time. Technical Report CMU-CS-95-210, CS department, CMU.

[14] Yang, J., & Waibel, a. (1996). A real-time face tracker. Proceedings of the 3th IEEE Workshop on Applications of Computer Vision, Sarasota, Florida, 142-147.

[15] Yang, J., Weier, L., & Waibel, A. (1997). Skin-Color Modeling Adaptation. Technical Report CMU-CS-97-146, CS department, CMU.