Counting on your Fingertips – An Exploration and Analysis of actions in the Rich Touch space

Rama Vennelakanti, Anbumani Subramanian, Sriganesh Madhvanath
Hewlett-Packard Laboratories
24 Salarpuria Arena, Hosur Main Road
Bangalore 560030, India
{rama.vennelakanti, anbumani, srig}@hp.com

Sriram Subramanian
Department of Computer Science
University of Bristol
Woodland Road, BS8 1UB UK
sriram@cs.bris.ac.uk

ABSTRACT

Although multi-touch technology and horizontal interactive surfaces have been around for a decade now, there is limited understanding of how users use the *Rich Touch* space and multiple fingers to manipulate objects on a table. In this paper, we describe the findings and insights from an observational study on how users manipulate photographs on a physical table surface. Through a detailed video analysis based on images captured from four distinct cameras we investigate the various actions users perform, and various aspects of these actions, such as the number of fingers, the space of action, and handedness. Our investigation shows that user interactions can be described in terms of a small set of actions, and there are insightful ways in which hands are used, and number of finger used to carry out these actions. These insights may in turn be used to inform the design of future interactive surfaces, and improve the accuracy of interpreting these actions.

Author Keywords

Touch, Multi-touch, Table-top, Interactions

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

1. INTRODUCTION

Multi-touch interfaces are now readily available in many devices. Mobile devices such as Apples iPhone, home computers such as the HP TouchSmart PC, and horizontal interactive surfaces such as Microsoft Surface, are all examples. The term “multi-touch” was originally devised to describe sensors and interactions that support multi-point interaction i.e., wherein users can simultaneously manipulate multiple points on the surface. Although sensor technologies which support multiple points of interaction, as used in MERL DiamondTouch [17] and Stantum MDK [18] are available, the focus of “multi-touch” interaction has been largely limited to two-finger interactions. For example, iPhone users use two fingers to zoom in and out of a map. On a Microsoft Surface, users can simultaneously rotate, translate, and scale a window through two-finger manipulations. The use of these two-finger touch interactions has become so much the norm that many so-called “multi-touch” sensors in use today – including the ones used in the iPhone, and HP TouchSmart – detect only up to two simultaneous touches.

A common criticism of multi-touch solutions in general has been that their utility is limited to enabling users to rotate and translate better [13]. Recognizing this criticism, researchers have been exploring richer ways in which systems can track users’ fingers on an interactive table, such as tracking the point of contact to determine the force exerted by the finger, or tracking movement in hover space. The driving premise behind these efforts is that these richer forms of input can lead to more creative and intuitive forms of interaction with multiple fingers.

However, there has been limited investigation into how users typically use their fingers and the whole hand to interact with objects on a table. Without this knowledge it is difficult to systematically and consistently design multi-touch interactions that effectively leverage multi-point tracking facilities of a system.

In this paper, we present the results of a detailed study that looked at how people use their hands to interact with physical objects such as photographs on a hard wood table surface, performing common tasks such as searching through, organizing and sorting photographs. In particular, we studied the *Rich Touch* space (described in the next section) of the interactions, with a view to discovering and understanding patterns of multi-touch. The main contributions of this paper are:

- Analysis of interactions with physical objects – photographs – on a normal table surface - in the Rich Touch space, leading to the definition of a set of actions used by either hand
- Analysis of Rich Touch in the execution of actions, with respect to their frequency, handedness, and number of fingers used

In the next section, we describe the concept of Rich Touch, followed by a review of some of the prior work related to this paper. We then describe our experimental study, followed by the results of video analysis and discussions. In the final section, we present some conclusions and directions for future research.

2. RICH TOUCH

Let us consider a scenario involving a person browsing through a set of hardcopy photographs laid out on the table in front of him/her. A typical interaction may be described as follows:

‘She ***moves*** *the photographs,* ***pulls*** *one towards herself while at the same time* ***rotating*** *it to its correct orientation, then* ***picks it up*** *to examine it more closely,* ***puts it back*** *on the table at a different location and* ***points*** *to something in it.’*

When she ***moves***, ***pulls,*** *or* ***rotates*** a photograph**,** finger and/or hand contact is maintained with the photograph and with the surface of the table. When she ***picks up*** a photograph, the photograph is moved from the ***touch space*** (the table surface) into the ***hover space*** (the space above the table), but contact with the object of interaction is maintained. When she ***points*** to something in a photograph on the table, her finger may touch the surface or hover it.

Although physical contact may be absent, the interaction continues. Thus there are instances where in contact with the object of interaction remains continuous, while the object itself undergoes transitions between touch space and hover space. At other times, the object remains in the touch space while the hand and fingers undergo transitions. An important aspect of touch which is evident when the person tries to move an entire stack of photographs, or turn multiple pages in a book – is *pressure*. We can therefore describe the *Rich Touch* space as the continuous interaction space encompassing the distinct spaces of *hover* and *touch* (wherein the latter includes *pressure*), represented in Figure 1. However, our work in this paper focuses on the spaces Touch and Hover.

A second important aspect of the Rich Touch space is the *number of fingers* used to perform an action. For example when rotating a photograph, the user might use four fingers whereas when pushing a picture, the user might just use one finger.

Another aspect of the touch space is handedness. Some actions are more likely to be performed with the dominant hand while some actions are equally likely done with the non-dominant hand. This division of labour may be asymmetric or symmetric depending on the type of action. Thus *division of labour* between the two hands is the third important aspect of the Rich Touch space.

In the construction of future digital tables and in the design of multi-touch interactions it is necessary that we try and assimilate the spaces within which the interactions happen in the physical world into the interactions in the digital world. The motivation and the focus of our study arose from the need to understand the extent of multi-touch required for building a multi-touch system or designing multi-touch interactions.

From these descriptions, it is evident that interactions in the physical world and with physical objects, assume the Rich Touch space as a continuum. However, in the context of digital objects and digital surfaces, it continues to be treated as composed of discrete layers, with the transition information being lost. To our knowledge, the study of these spaces in an integrated manner is relatively new and under-explored.

As a first step towards studying this Rich Touch space, we focus on the actions performed by participants in the context of a common task such as photo browsing and organisation – in terms of actions and their handedness, finger count and space of occurrence. These are described in later sections of the paper. We start with a brief look at relevant published literature dealing with different aspects of the Rich Touch space.

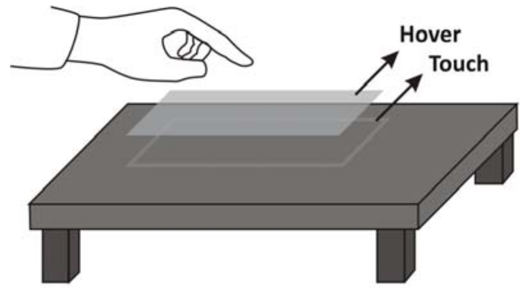


Figure 1. Rich Touch space, as defined in our work encompassing the distinct spaces of Touch (and pressure) and Hover

3. RELATED LITERATURE

Aspects of Rich Touch have been explored in various forms by researchers in the past. In BumpTop [1], the authors explore use of pen pressure and force to create a fluid and intuitive pen based desktop pile management system. Grossman et al. [10] present Hover Widgets which are localized interface widgets that use pen movements in the hove space of the tablet display. Subramanian et al. [24] present a tabletop interaction technique that uses the hover space to create multiple interaction layers.

Cao et al. [4] argue that when interacting with physical objects the shape and size of the contact regions plays an important role in determining the actions that are possible to perform with the objects. Wu and Balakrishnan [26] present a suite of multi-touch and whole hand gestures for a prototype room furniture layout application.

Rotate ‘N Translate (RnT) [13] is a tabletop interaction technique that combines rotation and translation into a single fluid gesture similar to how users pass a paper around a table. The virtual object is partitioned into two regions by a circle around the object. Within the circular area, the user can drag the artefact around the workspace (translate only) and outside the centre, the object rotates around the point of contact as it translates in the direction of movement. Various works [1, 3, 4, 8, 10, 12 –14, 20, 26] have examined different intricacies of table-top interactions.

With the ready availability of multi-touch and multi-user interactive displays researchers [4, 5, 20, 25] have looked at ways in which people organize their tabletop workspace and collaborate with each other [12, 20]. Researchers have also looked at applications of multi-touch interfaces to help people manipulate digital media by mimicking physical interactions to different extents [1, 2, 24].

Terrenghi et al. [24] investigated differences in manipulation of objects on a physical and digital surface paying particular attention to the physical items and their digital counter parts. The comparison was based on performing similar tasks in both digital tabletops and with physical objects. They found although bimanual interaction was common in physical interaction most of the digital tasks involved one-handed interaction. They also observe that despite the difference in spatial affordances, there are fundamental elements to both tasks that are common. They also identify limitations in digital interactions due to lack of the third dimension in the space.

Figure 2. Setup showing four cameras on tripods around the table used in the study

Figure 3. Sample frame in video from the image capture software

Epps et al. [7] present a preliminary evaluation of hand shape use in tabletop tasks. Users were asked to perform specific tasks such as selecting, moving, etc., on a touch-sensitive tabletop surface. They report on the frequency of use of each finger and the typical range of tasks for which they are used. Being a preliminary analysis, the paper is limited to gross indications of usage patterns.

While there is a whole body of work that examines surface interactions, we believe our work adds another dimension to this field of research. With the growing interest in immersive environments, interaction spaces have become as ubiquitous as displays. Large and 3D displays and immersive environments along with multiple interaction devices in the forms of kiosks, consoles and handhelds provide a fertile space where interactions and interaction spaces will move fluidly between touch (including pressure), hover and hand gesture spaces to enable intuitive interactions. So, treating this space as a continuum becomes crucial to designing immersive interactions. We believe our work on the Rich Touch space adds this dimension to the field of study.

In this work, we take a closer look at Rich Touch and examine the use of touch and hover space, fingers and hands in manipulating objects around a tabletop. We believe such an analysis will help understand the types of actions that need to be detected by the system and provide cues for improving gesture recognition on interactive surfaces.

4. THE OBSERVATIONAL STUDY

As already described, our main concern is the understanding of actions that constitute physical Rich Touch interaction, and their characteristics. Thus the main objective of the current study is to ascertain and study actions for (i) their occurrence and frequency (ii) their space of occurrence, handedness and finger count, with a view to providing insights for the design of multi-touch interactions and multi-touch surfaces.

4.1. Study Design

While designing the study, we evaluated factors and design choices that might influence the findings, such as: physical vs. digital surface, surface inclination: vertical vs. horizontal, sit-down vs. stand-up use, table height, number of participants: single vs. two or more, objects to manipulate – photos, books, shapes. A digital table would be constrained by its technology robustness and ability to track multiple fingers; hence the choice of a physical table in this study. A physical vertical surface would provide no room for manipulating the objects of interaction; as objects would need to be fixed on the vertical surface in some way – with pins or magnets, which would restrict the freedom of movement, we chose a horizontal surface. Since most informal interactions happen around coffee tables unlike work other formal tasks which happen around study or work desks, we decided to conduct the study on a table at the height of a coffee table. Books and shapes would occlude fingers from observation cameras and so we decided to use printed photographs as objects of interaction for this study. We decided to carry out the study first with individual participants and then follow it up by a study with two users.

Based on these factors and decisions, the present study is designed as follows: A single participant would sit down at a regular coffee table to interact with a set of printed photographs and perform a set of five tasks. The tasks would reflect the common tasks that people do with photos and are classifiable [15] as generate (sort), choose, (search), negotiate (arrange) and execute (share) photographs . We felt that the focus on physical interaction for familiar tasks on a familiar work surface would be most suited for a first study of the Rich Touch space.

Participants: We had four randomly selected volunteers (from our office premises) – two male and two female, in the age group of 25- 30 years for the study. They were all right handed.

Methodology: The participants were invited into the lab and given a brief introduction to the study, and on signing a consent form for participating in the study and video /audio taping of the proceedings, they were asked to fill a brief profiling questionnaire and the Edinburgh Handedness inventory [6]. Each participant was then seated at a wooden-top coffee table – two feet wide, four feet long and one-and-a-half feet high, to perform a set of tasks.

As our objective was to study interaction in the Rich Touch space, participants were instructed to ensure that when they lifted the photos off the table top, they do not lift them above four inches from the surface. Although this constraint placed an unnatural restriction on the user restrictions, it ensured that the photographs along with the participant's hands stayed within the work area (and view of the camera), and that interactions, for the most part, stayed within the Rich Touch space. An observer sat across the table, facing the participant and engaged in conversation with the participants regarding the theme, frequency and occasions of photo-taking and how and with whom these pictures were shared. This was done with a view of making the participants less conscious of the cameras, and to gain insight into what people did with their photo collections. We also ensured that the observer did not intrude into the personal space [9] of the participant while performing the tasks. We therefore had the observer sit at a distance of at least one-and-a-half feet from the work table, in a lean back posture, with her hands off the table. This ensured that the participant had complete use of the work space.

Tasks: The tasks in this study pertained to sorting, searching, arranging and sharing photographs, and took approximately 25 to 30 minutes to perform in total. The specific tasks each participant performed were: from a pile of about 35 photographs (people pictures, event pictures, events, landscapes, flora and fauna): (i) find photographs that they find interesting, (ii) search and find photograph 'X', (iii) sort into similar piles, (iv) pick a few to arrange in a story sequence, and (v) gather photos into a pile.

4.2. Capture Setup

All the user interactions were captured on video. The capture setup involved four standard web cameras; connected through USB ports on a standard PC (see Figure 2). We used two Logitech cameras (ClickSmart 510), one Logitech QuickCam and one Philips (SPC 900NC) camera. Of these four cameras, one camera was positioned to provide an overview of the entire work surface, from above and two cameras were placed at the two far corners of the table. The fourth camera was placed adjacent to the edge of the surface, overlapping the view of another camera – to assist in better disambiguation of touch-hover actions occurring on the table surface.

We developed a software application using the OpenCV library [18] which enabled us to simultaneously capture the four different views (images), from four different cameras in real-time. The image frames from the four different cameras were captured sequentially, switching from one camera to the other in order. As the entire image capture process is extremely fast (order of few milliseconds), the differences in capture times across cameras are very small and so may be ignored.

Each camera was set to capture images of size 320x240 pixels at 30 frames per second. Individual image frames from the four cameras were composed in a tile of size 640x480 pixels (using the first 310x230 pixels from each camera image, ignoring the last 10 pixels on both row and column extremes), by placing the frames side-by-side in a 2x2 arrangement (see Figure 3). The system time was rendered as an overlay on the top-left corner of the composed frame. In this tiled frame, the top-left image shows the view of the entire scene (table), the top-right image shows a closer view of the table and the bottom two images show the views from cameras placed at the table corners. While the top-right and bottom-left images look similar, due to the positioning of two cameras in the same direction, the bottom-left camera was set view lower, to enable examination of the space of action – touch or hover; the top-right camera provided a view of the number of fingers in action.

4.3. Data Annotation

A total of about 80 minutes of video data was captured from all (four) the participants. This amounted to 28836 frames of relevance (discounting frames where the user's hands were not involved in the task or were out of the camera view). The four-camera view provided sufficient information for annotation – the overview camera provided a global view with information on the specific action being performed, the object being manipulated and the general location of the hand; information from the three table-level cameras was used to determine the spatial position of hand, the handedness, and the finger count while performing an action.

The video capture of one participant (about one-fourth of the available video) was analysed and annotated to develop the final set of annotation parameters. The action performed by the participant and the context of the action were both considered to define a stable set of 12 *actions*. For example, the context of the participant's intended action led to the definition of distinct actions which appear similar, if only the extent and trajectory of movement were to be considered. For example 'Rest' is similar to 'Hold' and 'Move' is similar to 'Pull' and 'Push'. The necessary context information was provided by the overview camera.

Thereafter, annotation involved analyzing the sequence of video frames and recording information on (i) space where the action happened – hover or touch, (ii) identification of the hand – right or left, (iii) number of fingers used – one, two, three, four, five fingers, and (iv) textual description of the physical action. The video frames were annotated manually by two independent annotators after 30 minute training each to use the software.

Annotation was carried out independently to avoid any bias, with each video annotated separately by each annotator. The annotation software allowed the annotator to get a frame by frame view of the recorded video. Annotators marked the actions of both hands in each video frame. When both hands were used, the actions were independently annotated, again with the number of fingers used in each hand. Frames where the hands were off the work table and not performing any task relevant activity were ignored (e.g., a participant's hands resting on their knee as they conversed with the observer). A sample of the annotation used in our work is shown in Table 1.

Table1. A sample of annotated data (fields include user ID, frame number in video, left-hand, action in left-hand, number of left hand fingers used, right-hand, action in right-hand, number of right hand fingers used
user_01;179;left;;;;right;Pick Up;5;Hover;
user_01;180;left;;;;right;Pick Up;5;Hover;
user_01;182;left;;;;right;Pick Up;5;Hover;
user_01;183;left;Hold;5;Hover;right;;;;
user_01;184;left;Hold;5;Hover;right;Hold;5;Hover;
user_01;186;left;Hold;5;Hover;right;Hold;5;Hover;
user_01;187;left;Hold;5;Hover;right;Hold;5;Hover;
user_01;188;left;Hold;5;Hover;right;Hold;5;Hover;
user_01;189;left;Hold;5;Hover;right;Hold;5;Hover;
user_01;190;left;Hold;5;Hover;right;Hold;5;Hover;
user_01;191;left;Hold;5;Hover;right;Hold;5;Hover;
user_01;192;left;Hold;5;Hover;right;Hold;5;Hover;
user_01;194;left;Hold;5;Hover;right;;;;
user_01;196;left;Hold;5;Hover;right;Pick Up;3;Hover;

5. FINDINGS

Having annotated the video data, we analysed the annotated data along several dimensions – space, handedness, finger count and actions. We summarize the findings from analysis of our annotation data. These statistics obtained pertain to a specific set of actions that characterize a specific set of tasks. It should also be pointed out that the statistics were computed at the level of individual frames. Therefore, the frequency of a particular event reflects the number of frames – and hence time duration, for which it was observed, rather than the discrete number of instances of that event. This methodology has both advantages and shortcomings (discussed later). As the video data was annotated by two independent annotators, data from each annotator was analysed separately and the average of the two data points was considered for this analysis.

5.1. Actions and Action Groups

Actions may be combined sequentially to achieve a larger goal. A number of such actions were observed from the captured video data during the process of annotation. For example, it was observed that accomplishing a specific task such as searching for a photograph involves many actions such as 'pulling' photographs towards the participant, 'rotating' the ones that are not oriented correctly, 'pushing' one or more photographs away, 'pointing' at something in the photographs, and so on. These actions are at times performed discretely, while at other times they are performed in quick succession and in one fluid movement – a participant will 'rotate' a photograph while 'pulling' it towards him/her. For the purposes of supporting a coarser level of analysis, these actions may be grouped according to the extent of translation (displacement of the object on or over the surface) involved, into 'gross' and 'fine' action groups:

- *Gross action group* - actions performed with speed and not much precision. These include (see Table 2): *Pick up*, *Hold*, *Move*, *Push*, *Pull*, *Gather*, *Place*, *Separate*, and *Rest*
- *Fine action group* - actions that require precision and usually involve very little translation. These include (see Table 3): *Adjust*, *Rotate*, and *Point*

5.2. Gross Action Group

Move – An action performed to translate objects across space to a specific position

Pick up – An action performed to lift objects off the surface and transitions from touch to hover space.

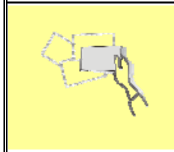
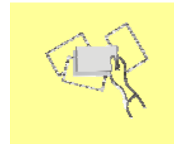
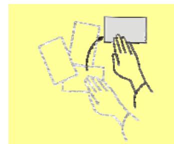
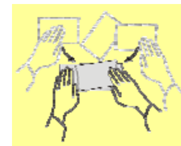
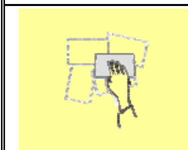
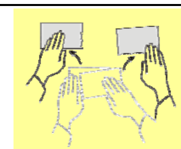
Hold – An action to hold on to an object to either inspect it or as a place holder

Pull – A brief and quick action performed to translate objects across a space towards oneself but not to any particular position

Gather – A brief and quick action performed in short bursts to translate objects to a pile

Place – An action performed to place objects at a specific place involving a transition from hover to touch space

Rest – A state of inaction of the hand, sometimes in the touch space, as a place holder

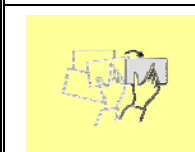
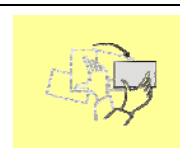
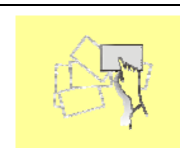
Table 2: Gross action group		
 *Pick-up*	 *Hold*	 *Move*
 *Push*	 *Pull*	 *Gather*
 *Place*	 *Separate*	 *Rest*

5.3. Fine Action Group

Adjust – A precise action performed to position objects in a specific orientation or at a specific location

Rotate – A precise action performed to change the orientation of an object

Point – A precise action performed to draw attention to a specific object or specific content of a specific object

Table 3: Fine action group		
 *Adjust*	 *Rotate*	 *Point*

5.4. Frequency of Actions

As seen in Figure 4, gross actions dominate the actions performed by participants (22,530 frames), accounting for 86% of the all actions as compared to 14% for fine actions. While this disparity may be attributed to the nature of the tasks – search, find, group and arrange – that largely require gross actions, we believe this to be typical of what an average user may do with photographs and other similar objects on a surface, in a walk-up and use scenario. We also looked at the relative frequency of individual actions. From Figure 5, we observe that *Hold*, *Pick up* and *Move* are the most frequently occurring actions.

It is worth noting that the definition of actions is based on the intent of the user (which was arrived at by looking at the end result of the action), and not on the space in which it happens. Hence actions may occur in either the touch or hover space, and may sometimes transition from touch to hover spaces or vice versa. They may be performed with either hand.

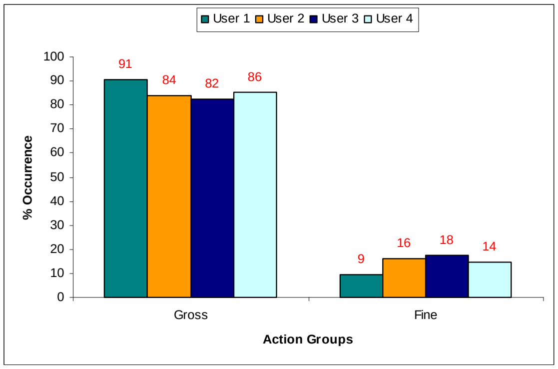


Figure 4 Frequency of occurrence of action.

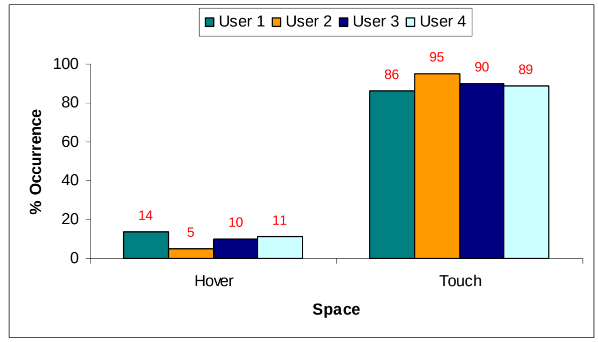


Figure 6: Action space of all actions.

5.5. Action Space

From Figure 6, we observe that across users, about 90% of all the actions occur in the touch space compared to only 10% for the hover space. This may be attributed to the affordances of the objects being interacted with and the interaction surface. The touch space is rich in interactions and a detailed analysis of this space reveals interesting insights that can inform the design of multi-touch interactions and that of future multi-touch tables.

Performing the same analysis at the granularity of individual actions (Figure 7) reveals that while most actions occur in the touch space, *Move* occurs in the hover space 34% of the time, and Hold 14% of the time.

5.6. Handedness of Actions

We looked at handedness, at the granularity of action groups. From Figure 8, we observe that while gross actions may be performed almost equally by the right (51%) or left hand (48%), fine actions are more likely to be performed with the dominant right hand (66%).

Analyzing further on individual actions, as seen in Figure 9, reveals that *Hold* (74%) is a common left handed action, as is also *Rest* (75%). These are gross actions, whereas the fine actions *Adjust*, *Rotate* and *Point* are more often right handed than left handed.

5.7. Finger Prominence for Actions

Figure 10 shows the finger count for each action. This data presents insights into how many fingers people use naturally for different actions. For example, when performed with the dominant (right) hand, *Point* is prominently a one finger action while the other actions could all use five fingers. However when performed with the non-dominant hand, it is most likely to be performed using all four or five fingers.

6. DISCUSSION

6.1. Actions

The important conclusion from our study was the fact that most of the relevant user interactions at the table can be described as compositions of a small set of actions. While these actions were derived for a specific set of tasks in the specific context of photo organization, we believe that they can be generalized to manipulating physical objects with similar affordances on a horizontal surface, and inform the actions supported by an interactive surface.

As seen is Figure 7, most actions span the touch and hover spaces, demonstrating the continuous nature of the Rich Touch space. However only *Move* and *Hold*, surprisingly do not occur in the hover space frequently; the other actions are largely confined to the touch space.

This suggests that for manipulating objects on digital tables, the ability to detect touch is largely sufficient for most tasks, whereas the hover space is needed primarily to move or hold objects without interfering with other objects on the touch surface. Interactive surfaces could perhaps have a mechanism for temporarily "suspending" objects, and moving them while in

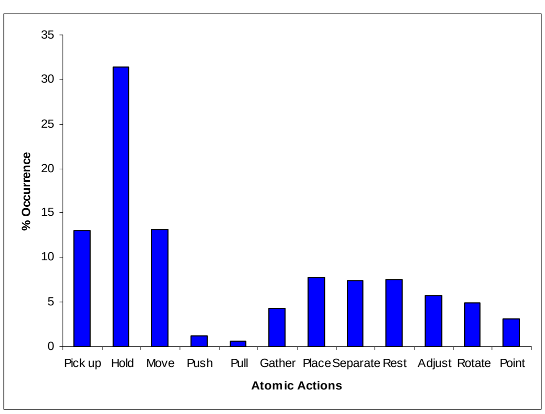


Figure 5. Frequency of occurrence of actions

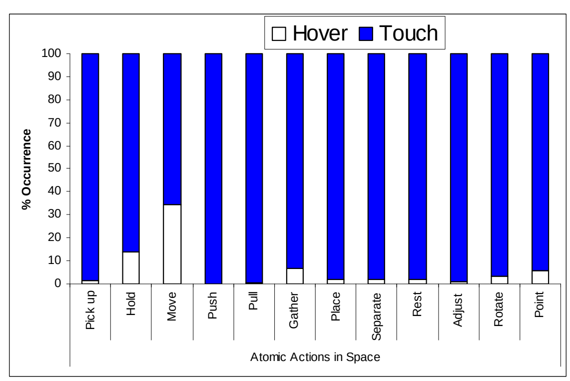


Figure 7: Action space of actions

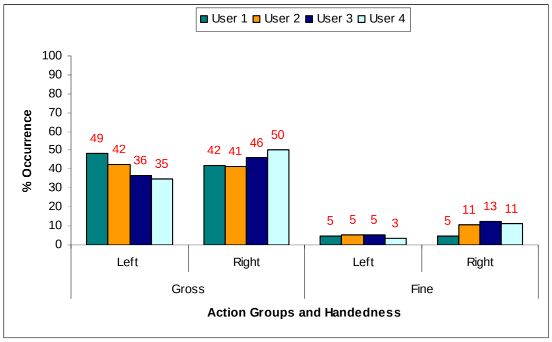


Figure 8: Handedness of action groups

suspension.

Gross actions involving large displacement dominate the actions performed by participants. This is going to be true for interactive surfaces as well. While a lot of the focus on multi-touch interaction has been on fine actions such as zooming-in and rotating objects, efficient ways of performing gross actions perhaps need more attention.

6.2. Finger prominence

There are definite patterns in the number of fingers used for different actions. For instance *Point* with the dominant hand was done mostly with a single finger, while the *Rotate* action always used three or more fingers.

From the perspective of designing interactions, this informs the choice of most natural "multi-touch" gestures for specific actions. For instance, rotating a photo is perhaps best performed using multiple fingers of each hand. Contrary to this, many interfaces today expect people to use exactly two fingers for this purpose.

6.3. Handedness

Actions may be performed using either hand. For gross actions involving large displacement, the non-dominant hand is likely to be used as much as the dominant one. For fine actions, the dominant hand is greatly preferred. This has been shown by previous studies as well, and is further substantiated by our study. If the interactive surface could reliably detect the hand that is being used (using hand geometry, for example) a different set of interactions could be enabled for each hand and placed within easy reach of that hand. There are also differences in

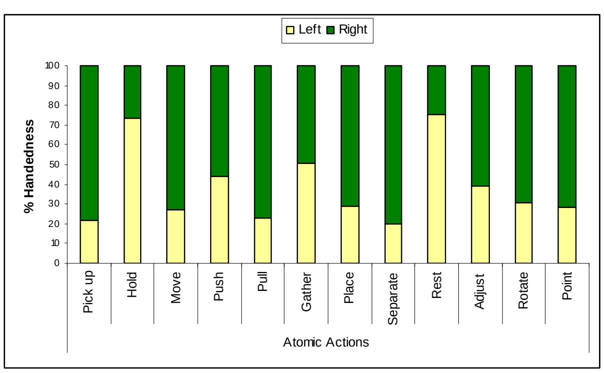


Figure 9: Handedness of actions

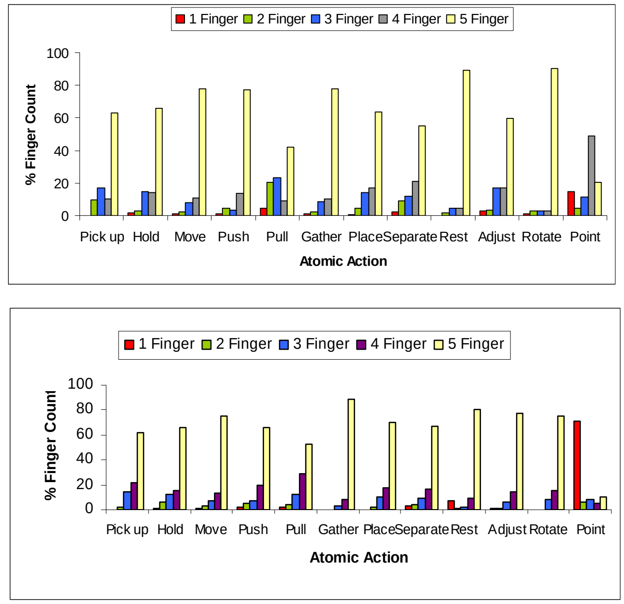


Figure 10. Finger count frequencies for actions using left hand (top) and right hand (bottom)

these patterns between hands – for instance, *Point* with the non-dominant hand often uses three or more fingers.

Many studies on bimanual interaction have shown that the two hands are used in concert through an asymmetric division of labour [3, 11]. In order to explore how this division of labour happens in tabletop tasks, we studied the probability of joint occurrence of actions performed using the left and right hands, and the results for the touch space are depicted pictorially in Figure 11. We observe that combinations along the matrix diagonal, representing symmetric use of the two hands, e.g. *Gather*, *Separate*, *Adjust*, *Rotate* and *Push*, are more frequent.

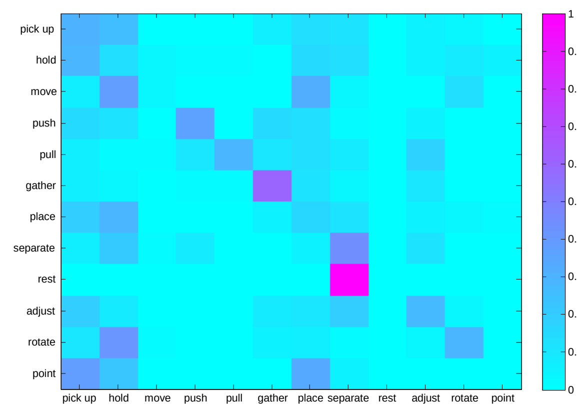


Figure 11. Probability matrix of two-handed in touch space. Rows represent actions of left-hand and columns represent actions of the right-hand

In other instances, the actions performed by the two hands are different but complementary.

Bi-manual actions that are performed naturally – whether symmetric or asymmetric, should inform the design of interaction for interactive surfaces. Further, the joint distribution of actions performed with the two hands can inform the interpretation of these actions. In other words, knowing what the

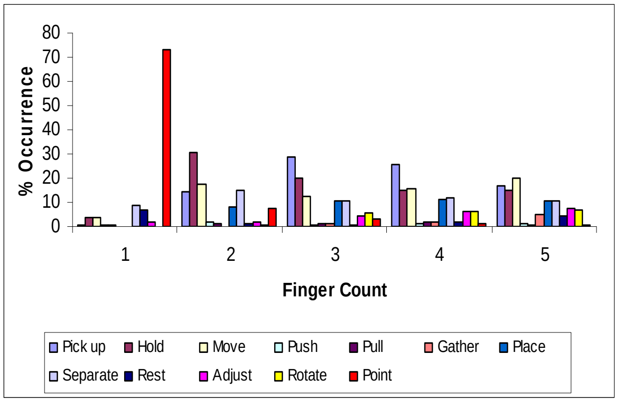


Figure 12. Frequency of actions given finger count

non-dominant hand is doing can improve the interpretation of the other hand, and vice versa.

6.4. Limitations

We acknowledge that our study, and so the findings and their applicability have some evident limitations. Sample size – although the number of users in our study may not match the widely accepted data size, we believe that the insights from our study are valuable enough to stimulate an interesting debate in the research community. Setup – due to image capture constraints, the participants were asked not to lift objects off the table, thus limiting natural interactions to an extent. We could have resolved this with a more sophisticated capture set up. Annotation and analysis methodology - the video data was annotated in such a way that a new annotation was created anytime, either the event itself changed or an attribute of the event (such as the action space, or the finger count) changed. From this we were able to derive annotation at the level of individual video frames. We had a choice of performing statistical analyses at the granularity of frames, or discrete events. We decided on the former, since it gave us granular data on the various states of each event. The statistics associated with actions, or with their combinations, are therefore all to be interpreted as relating to *action frames*, and correlated with the time duration spent on that action. However this also introduces variability in that, different users (or even the same user at different times) may perform the same action over shorter or longer time intervals. Further, actions such as *Hold* or *Rest* that last long, relative to other actions may skew the statistics. These are issues to be examined more closely in subsequent studies

7. CONCLUSIONS

Actions do not have a uniform distribution. The relative frequency of actions can be used to derive prior probabilities for the recognition of these actions. If the interactive surface could detect that an action was occurring in a particular space (e.g. touch or hover), with a particular hand (dominant or non-dominant), or using a particular number of fingers, then the surface may be able to significantly improve the accuracy of interpreting these actions. The specific actions available to the user at any point can also be made a function of these characteristics and conveyed to the user. As a case in point, one can examine the most frequent actions performed using a given number of fingers. This analysis (for the dominant hand) is represented pictorially in Figure 12, and summarized in Table 4.

Table 4. Actions likely for each finger count

Finger count	Likely actions
1	*Point* (73%)
2	*Hold* (30%), *Move* (17%), *Pick up* (14%).
3	*Pick up* (29%), *Hold* (20%), *Move* (13%)
4	Pickup (26%), *Hold* (15%), *Move* (15%).
5	*Move* (20%), *Pick up* (17%), *Hold* (15%).

An intelligent surface can use the number of fingers on or approaching the surface (along with other cues such as the action space and handedness) to either anticipate, or improve the interpretation of the action. Specific actions available to the user at any point can also be made a function of these characteristics and conveyed to the user.

As mentioned earlier, with the growing interest in immersive spaces, touch, hover and distance interactions will come to coexist. Treating these spaces as a continuum is the flavour of things to come for intelligent and efficient design of interfaces and interactions. Being able to pre-empt the user based on the space of action and the other cues can provide more interesting designs in interfaces.

8. SUMMARY

In this paper, we described a study of interaction with physical objects in the Rich Touch space, comprising of touch and hover spaces. This study is a first step towards the systematic study of interactions in the Rich Touch space. In particular, the study focused on photo manipulation tasks on a physical table surface. We found that physical interactions can be decomposed into sequences of *actions*, panning the touch and hover spaces. The interactions are discernable in terms of the space they occupy, handedness, and the number of fingers used. We believe explorations like ours; can inform the design of future interactive surfaces in a number of ways.

9. REFERENCES

1. Agarawala, A. and Balakrishnan R. (2006), Keepin' it real: pushing the desktop metaphor with physics, piles and the pen. *In Proc. SIGCHI Conf. on Human Factors in Computing Systems*, CHI '06, ACM, 1283-1292.
2. Apted, T., Kay, J., and Quigley, A. (2006). Tabletop sharing of digital photographs for the elderly. *In Proc. of the SIGCHI Conference on Human Factors in Computing Systems.* R. Grinter, et al. Eds. CHI '06. ACM, New York, NY, 781-790.
3. Balakrishnan, R., and Hinckley, K. (2000). Symmetric bimanual interaction. *Proc. of CHI 2000 – the ACM Conference on Human Factors in Computing Systems*. p. 33-40.
4. Cao, X., Wilson, A.D., Balakrishnan, R., Hinckley, K., and Hudson, S.E. (2008) Shapetouch: Leveraging contact shape on interactive surfaces. *In IEEE Intl. Workshop on Horizontal Interactive Human-Computer Systems*, IEEE, 129-136.
5. Dietz, P. and Leigh, D. (2001). DiamondTouch: a multi-user touch technology. *In Proc. of the 14th annual ACM symposium on User interface software and technology,* Orlando, Florida, 2001, ACM Press.

6. Edinburgh Handedness Inventory. http://www.cse.yorku.ca/course_archive/2006-07/W/4441/EdinburghInventory.html
7. Epps, J., Lichman, S., and Wu, M. (2006), A study of hand shape use in tabletop gesture interaction, *In Proc. CHI '06 Extended Abstracts on Human Factors in Computing Systems*. CHI '06, ACM, 748-753.
8. Forlines, C., Shen, C., Wigdor, D., and Balakrishnan, R. (2007). Direct-touch vs. mouse input for tabletop displays. *Proc. of CHI 2007 – the ACM Conference on Human Factors in Computing Systems.* p. 647-656.
9. Griffin, E. Proxemic Theory, http://www.afirstlook.com/docs/proxemic.pdf
10. Grossman, T., Hinckley, K., Baudisch, P., Agrawala, M., and Balakrishnan, R. (2006), Hover widgets: using the tracking state to extend the capabilities of pen-operated devices, *In Proc. SIGCHI Conf. on Human Factors in Computing Systems,* CHI '06, ACM, 861-870.
11. Guiard, Y. (1987). Asymmetric Division of Labor in Human Skilled Bimanual Action: The Kinematic Chain as a Model. Slightly edited version of an article originally published *in Journal of Motor Behavior*, vol. 19, pp. 486-517. http://cogprints.org/625/
12. Kruger, R., Carpendale, M.S.T., and Greenberg, S... (2002). Collaborating over Physical and Electronic Tables. In *Poster In Extended Abstracts of Conference on Computer Supported Co-operative Work*. ACM, pages 137-138.
13. Kruger, R., Carpendale, S., Scott, S. D., and Tang, A. (2005). Fluid integration of rotation and translation, *In Proc. SIGCHI Conf. on Human Factors in Computing Systems,* CHI '05, ACM, 601-610.
14. Kruger, R., Carpendale, S., Scott, S.D., and Greenberg, S. (2004). Roles of Orientation in Tabletop Collaboration: Comprehension, Coordination and Communication. *Journal of Computer Supported Collaborative Work*, 13(5-6):501-537.
15. McGrath, J. (1984) “Groups and Human Behavior” (Excerpt), Citation: From “Groups: Interaction and Performance”, pp.12-17, Reprinted in: *Groupware and Computer-Supported Cooperative Work*, pp.113-115.
16. MERL DiamondTouch, http://www.merl.com/areas/DiamondTouch
17. Multi-Touch evaluation and development kit , Stantum, http://www.stantum.com
18. OpenCV library. http://sourceforge.net/projects/opencvlibrary
19. Rekimoto, J. and Saitoh, M. Augmented surfaces: a spatially continuous work space for hybrid computing environments. *In Proc. SIGCHI Conf. on Human Factors in Computing Systems*, CHI ’99, ACM (1999), 378-385.
20. Scott, S. D., Sheelagh, M., Carpendale, T., and Inkpen, K. M. (2004). Territoriality in collaborative tabletop workspaces. *In Proc. of the 2004 ACM Conference on Computer Supported Cooperative Work. CSCW '04.* ACM, New York, NY, 294-303.
21. Scott, S., Carpendale, M.S.T., and. Inkpen, K. (2004). Exploring Casual Tabletop Interactions. *Research report 2004-742-07,* Department of Computer Science, University of Calgary, Calgary, Alberta.
22. Shen, C., Vernier, F. D., Forlines, C., and Ringel, M. (2004). DiamondSpin: an extensible toolkit for around-the-table interaction. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '04. ACM, New York, NY, 167-174.
23. Subramanian, S., Aliakseyeu, D., and Lucero, A. (2006). Multi-layer interaction for digital tables. *In Proc. ACM Symposium on User interface Software and Technology,* UIST 2006, ACM, 269-272.
24. Terrenghi, L., Kirk, D., Sellen, A., and Izadi, S. (2007). Affordances for manipulation of physical versus digital media on interactive surfaces. *In Proc. SIGCHI Conf. on Human Factors in Computing Systems,* CHI '07, ACM, 1157-1166.
25. Wigdor, D., Penn, G., Ryall, K., Esenther, A. and Shen, C. (2007). Living with a Tabletop: Analysis and Observations of Long Term Office Use of a Multi-Touch Table. *In Proc. of the Second Annual IEEE International Workshop on Horizontal Interactive Human-Computer Systems,* 2007. TABLETOP '07. , 2007, 60-67.
26. Wu, M. and Balakrishnan, R. (2003). Multi-finger and whole hand gestural interaction techniques for multi-user tabletop displays. *In Proc. ACM Symposium on User interface Software and Technology,* UIST '03, ACM, 193-202
27. Wu, M., Shen, C., Ryall, K., Forlines, C., and Balakrishnan, R. (2006). Gesture registration, relaxation, and reuse for multi-point direct-touch surfaces. *Proc. of TableTop 2006 - the IEEE International Workshop on Horizontal Interactive Human Computer Systems.* p. 183.190.