Multitouch and Gesture: A Literature Review of Multitouch and Gesture

Liwen Xu
University of Toronto
Bahen Centre, 40 St. George Street, Room 4242,
Toronto, Ontario M5S 2E4
xuliwenx@cs.toronto.edu
+1 (647) 989-8663

ABSTRACT
Touchscreens are becoming more and more prevalent, we are using them almost everywhere, including tablets, mobile phones, PC displays, ATM machines and so on. Well-designed multitouch interaction is required to achieve better user experience, so it has gained great attention from the HCI researchers. This paper firstly introduces early researches on this area and then focuses on current researches of multitouch and gesture. It discussed three big problems about it, Virtual Manipulation vs. Physical Manipulation, Gesture Sets and Recognition and Planning When Manipulating Virtual Objects and analyzed their advantages and limitations. Finally based on the limitations of the literatures reviewed, future research directions are given. Though the literature reviewed cannot achieve perfect solution for multitouch interaction, they all provide a source of inspiration for other researchers.

ACM Classification Keywords
Multitouch, gestures, user Interfaces, tangible user interfaces, virtual interaction

INTRODUCTION
The multi-touch technology has been brought to the mainstream market since the announcements of Apple’s iPhone and Microsoft’s Surface in 2007, and became a popular topic in the field of human computer interaction thereafter [1]. However the multi-touch technology was invented at least 25 years earlier than it being world widely used. It is not surprising though, if we look at the history of mouse, which takes thirty years from the laboratory to the industry. The emergence of touch screen technology can be traced back to as early as 1960s. In 1967, E.A. Johnson at R.R.E (Royal Radar Establishment) presented a novel I/O device for computer called “Touch Displays” as an effort to resolve the inefficiency in man-machine communications with traditional keyboards in large data-processing systems [2]. The device is capable of taking instructions from operator’s finger touch directly. It’s also worth pointing out that “Touch Displays” uses capacitive sensing which is basically the same technology that modern screens/tablets use. Later in 1972, the PLATO IV system terminal has included a touch panel for educational purposes. Students can touch anywhere on the panel to answer questions. Behind the touch panel is a 16*16 infrared grid such that when user touches the panel user’s finger will block the infrared beam and therefore the system can detect a touch [3]. The multitouch interface came out shortly after the invention of touch screen technology. Bill Buxton, one of the pioneers in multitouch field, pointed out that the Flexible Machine Interface proposed by Nimish Mehta in his Master’s Thesis at University of Toronto in 1982 was the first multitouch system he is aware of [1]. One year later, Nakatani and Rohrlich from Bell Labs came up with the concept of “Soft machines”, which utilize computer graphics to simulate physical controls like buttons and keyboard covered by a touch screen for human to operate [4]. They also claimed that they have built the system in the lab but didn’t mention many details about the implementation. At about the same time Buxton and his team presented a prototype multi-touch tablet that is capable of detecting multiple simultaneous points of contact by using a capacitive grid and area subdivision to detect points of contact [5].

BACKGROUND LITERATURE REVIEW
Flexibility
Although touch screen technology has made the interactions between human and machine closer than ever, the capability of detecting only a single point of contact is not enough for more complex operations due to its lack of degree of freedom. A single-touch screen with no pressure sensors can only give a binary state of a position on the screen, namely whether it’s touched or not. With such limitation it’s even difficult to detect user drawing a line since the system can hardly tell when user starts drawing without being proper signaled [6]. As a result, adding pressure sensor will increase the degree of freedom to some extent. Indeed researchers have investigated the possibility of using force and torque sensors to enhance the interactive [7]. But obviously the more direct way is to add the support of detecting multiple points of contact. The touch screen allows user to interact with machine without an extra layer of mechanics, therefore users are free to use up to ten
fingers to operate on the touch screen. In fact, simultaneous touches are considered as a necessity in some contexts just like playing a piano [6]. For example, a keyboard shortcut usually involves the operation of holding one key while pressing another key. To simulate the keyboard and such shortcuts on a touch screen, it is required that the touch screen is capable of sensing more than one point of contact [8]. Another example is to operate a set of slide potentiometers, the operator need to control each slide with a finger at the same time [6].

Efficiency
It is common that our two hands are assigned to separate, continuous work in daily life, therefore it can be assumed that if designed properly, two-hand touch will generally outperforms single touch. Indeed, two experiments were carried out by Buxton and Myers to test the efficiency of bimanual operation [9]. The first experiment asked subjects to do positioning with one hand and scaling with the other hand, and recorded the time engaged in parallel activity. The result showed that subjects were operating in parallel nearly half of the time, which indicated that such bimanual operations are indeed natural for human. The second experiment asked subjects to navigate to and select a specific part in a document. Subjects were divided into two groups: one group used only one hand on a scroll bar while the other group used one hand to control scrolling and the other hand control text jumping in the document. The result showed that the two-handed group out-performed the single-handed group regardless of whether the subjects are novices or experts. This gives strong evidence to support that multi-touch tends to have better performance over single-touch dealing with same objectives.

Precision
The precision problem is a long-existing and inherent problem of the touch screen technology. In early days the resolution of the underlying sensing hardware is the main cause for the precision problem and later on the size of the finger inevitably hinders touch screen from being accurate on targeting and selection. To solve the precision problem, researchers have been exploring different strategies and gestures for precise selection over the twenty years. Before multi-touch technology is available, some research has already been done in finding the best strategy of detecting selection with single point of contact [10] gives a brief introduction of the two commonly used strategies called land-on and first-contact. Land-on strategy detects selection by comparing the touch location and the target location. It is a quite naive approach and therefore lacks of accuracy and tends to have high error rate. Whenever user fails to touch the target at the first try the user has no way to amend the mistake except taking another try. The first-contact strategy mitigates such situation by utilizing continuous feedback about the touch location. The system will now wait until the touch location come across the first target location, and report that target as the subject of the selection. In such way, when user fails to touch the desired target, user still has change to drag his finger to the right target, as a result reduces the touch error rate [10] proposes a third strategy which is called take-off. With such strategy, the cursor is no longer under the finger but above the finger with a fixed amount offset so that user can visually easily decide where the cursor is. Then the selection event happens when user releases his finger. If a target is under the position of the cursor then it is selected. This further reduces the possibility that user select an undesired target. The strategy has been widely accepted as the implementation of selection in the following decade, but new improvement comes along with better screen resolution. Sometimes it is required for user to select small targets whose side length is only a few pixels. In such case the size of the finger becomes the bottleneck for selection resolution. One resolution is the zooming strategy, where one can zoom in the touched part for easier selection of small target. The disadvantage of the zooming strategy is that user will lose the contextual global view that may be important for user’s task. Another possible solution is cursor key that allows user to move cursor pixel by pixel. Such method is robust but indirect. With the availability of multitouch technology, researchers started to consider the bimanual gesture for precise selection. In 2003, Albinsson and Zhai proposed a technique called cross-lever [11]: the position of the cursor is decided by the intersection of two “levers”, user can control the endpoints of the two levers to adjust the position of the intersection point, i.e. the position of the cursor. Such method achieves very low error rates but is time consuming to use. More recently, [12] defines a set of dual finger gestures for precise selection, which strives to give a more intuitive interface.

CURRENT RESEARCH
Virtual Manipulation vs. Physical Manipulation
The manipulation of virtual objects has a central role in interaction with tabletops [13]. To make multitouch gestures easier to learn and memorize, researchers are studying the relationship between the manipulation of the real physical objects and the manipulation of virtual objects displayed on screen. The studies are inspired by the fact that average person can skillfully manipulate a plethora of tools, from hammers to tweezers [14]. Multitouch screen is a media accept physical input (finger touch) to control the visual virtual objects displayed on screen. So far, much work that has been done is focusing on the mid-air interactions, which enable physical movements manipulate virtual objects, such as work with the Kinect [15] and Leap Motion [16]. A common mantra for these new technologies is that they allow us to become more embodied with the digital world and better leverage the natural motor and
body skills of humans for digital information manipulation. [17]. Compared with multitouch screen interactions, the easier part of mid-air interaction is that it is performed in a 3D space (the information of gestures is collected in a 3D space), which matches human’s natural body movements. However, for problem of multi-touch screen interactions, because the input is 2D points of contact on the screen, the degree of freedom of multitouch gesture sets is smaller than the mid-air interaction gestures, which makes it harder to build the relationship between manipulation of physical objects and multi-touch 2D screen manipulation of virtual objects.

Alzayat et al.’s studies showed that the psychophysical effect of interaction with virtual objects is at least different from interaction with physical objects. One phenomenon that was used in the experiments for measurement is figural after effects (FAE). Gibson firstly discovered this phenomenon [18, 19, 20]. After that, several experiments were conducted by Kohler and Dinnerstein to observe this effects [21]. In those experiments, participants were blindfolded and inspected the width of a sample cardboard piece by holding it, and then they were asked to tell the width of several different cardboard pieces with different widths. The experiments showed that participants overestimated the width of the narrower cardboard pieces (comparing to the sample cardboard) and underestimated the width of wider cardboard pieces. However, the control group who did not touch the sample cardboard provided more precise estimation. Alzayat et al. conducted two experiments, one was for reproducing the figural after effects, and the other was for testing whether this phenomenon took place while virtual objects were used in experiments. The results of the experiments showed that FAE was a reliable phenomenon, which appeared in both experiments in physical condition. But the phenomenon was very hard to observe in virtual condition. Therefore based on this result, Alzayat et al. drew a conclusion that tangible interfaces paradigm and the multi-touch interfaces paradigm are different. This fundamental difference may cause some higher-level differences.

Though Alzayat et al. have proved that tangible interaction is not perceptually equivalent to multi-touch interaction. The relationship between them is still very strong. With same common objects, people usually share same knowledge about it. While interface designers have to pay attention to the difference between tangible interaction and multi-touch interaction, they can also leverage people’s common knowledge about objects in design.

Harrison supposed that touch gestures design be inspired by the manipulation of physical tools from the real world [14]. Since, for commonly used tools, people usually have grasp of the way to use it, so when the similar tools are displayed on the screen, without any instructions, people may firstly try gestures close to the gestures which are used to manipulate the physical tools. This is a more instinctive behavior obtained from their knowledge. In addition, because these gestures are related to the people’s direct grasp of the tools, they are much easier to be memorized. Harrison selected seven tools to form a refined toolset: whiteboard, eraser, marker, tape measure, rubber eraser, digital camera, computer mouse, and magnifying glass[14]. These tools are from 72 tools brainstormed by four graduate students [14]. The experiments recorded the participants’ gestures used to manipulate both physical and virtual objects and found that people hold objects in a relatively consistent manner. The result showed that participants could discover the way to manipulate virtual mouse, camera, and whiteboard eraser tool easily. However, they still needed a sample grasp to get the correct operation for manipulation of marker, tape measure, magnifying glass, and rubber eraser [14]. Harrison proved that for some simple and common functions, such as mouse function, people could discover what it does and how to use it through trial and error [14]. From Harrison’s work, we can see leveraging the familiarity with physical tool is a useful way to improve the touch-screen interactions.

There are still many questions about the relationship between tangible interaction and multi-touch interaction remaining to be answered. This is a pretty new field and the researches that have been done are just first step into this field.

**Gesture Sets and Recognition**

Besides those basic gesture sets, to achieve better user experience of multitouch interactions, researchers are trying to design many different gesture sets. They have already proposed several touch gestures sets and solved the corresponding recognition problems.

One of the most important problems in gesture sets and recognition problem is the contact shapes. The easiest approach is to ignore the actual shape of the contact areas, simply treat them as points [22]. There are many techniques designed for it. For example, BumpTop [23] is one of the gestural techniques that only deals with point of contacts.

However, if we could detect the gesture beyond finger counting, the touchscreen would get more information and provide more friendly user interface. Rekimoto introduced SmartSkin to recognize multiple hand positions and shapes and calculates distance between the hand and the surface (screen) [24]. ShapeTouch explored the interactions that directly utilize the contact shapes on interactive surfaces to manipulate the virtual objects [25].

To enrich the information gathered during the multitouch interactions, Murugappan et al. defined the concept extended multitouch interaction [26]. The sensing was achieved by mounting a depth camera above a horizontal
Surface. That enables the extended multitouch interaction to detect multiple touch points on and above a surface, recover finger, wrist and hand postures, and distinguish between users interacting with the surface [26].

Extended multitouch interaction is a very powerful technique, and makes the system very “smart”. Because it can recover finger, wrist and hand postures, which increases the degree-of-freedom of gestures, this technique brings great potential for multitouch interaction design. In addition, it is able to distinguish different users, which is an amazing contribution to multi-touch interaction. This feature enables interaction of multiple users at the same time. The limitations of extended multitouch interaction are also obvious. Because the sensor is the depth camera, occlusion is a big problem that will affect its performance. If foreign objects obstruct the hands in the depth images, this system will fail to detect the hand and its postures. Also, the depth camera makes this system unportable, which makes this technique hard to be applied to prevalent portable devices (e.g. tablets and mobile phones). All those devices require sensors to be embedded in the devices themselves. Anyway, this technique is very useful for those location fixed surfaces.

With the surface/screen only, the information it can detect is the contact areas. We can still design many useful gesture sets using the contact areas only. Wigdor et al. presented a new multi-touch gesture technique called Rock & Rails that improved the direct multi-touch interactions with shape-based gestures [22].

The most important characteristic of Rock & Rails is that it utilizes non-dominant hand to mode actions of the dominant hand. In other words, non-dominant hand is the action selector of the dominant hand. This technique is actually commonly used in mouse/keyboard interface. By pressing a keyboard button, people can change the action mode of mouse (e.g. Mac OS, Windows, Photoshop). However, it is not widely used in multi-touch interface, so Rock & Rails is a great try to transfer keyboard/mouse techniques to multi-touch interface. By using Rock & Rails, the screen will receive the input as two areas of contacts, fingertip and hand shape. As discussed, the hand shape is the “action selector”. The fingertip is the “manipulator”.

Rock & Rails interactions are based on three basic shapes of the non-dominant hand, the action selector. The basic three shapes are Rock, Rail and Curved Rail. Placing a closed fist on the touch-screen creates the Rock shape. Placing hand flat upright creates Rail shape. Curved Rail is some hand pose that is somewhere between Rock and Rail. Figure 1 gives samples of the three hand poses.

Table 1. Input/Mode mappings of our three hand shape gestures. The gestures can be performed by either left hand or right hand [22].

Figure 1. Rock & Rails interactions. From left: Rock, Rail, and Curved Rail [22].

Placing hand inside the objects or outside the objects also trigger different action modes. The six situations and the corresponding commands are summarized in Table 1.

One of the advantages of Rock & Rails is reducing occlusions. Occlusion is a big problem of direct touch, which was noted by Potter et al. in 1988 [28]. Rock & Rails solved this problem via Proxies. Users can create a proxy by hand pose, a Rock outside of object, to easily create a proxy object and then link the proxy object to the target object. Any transformation of the proxy object will be applied to the target object. In this way, users can modify the target object by only manipulating the proxy object, which successfully avoid the occlusion. Users can easily create or delete the proxies without affecting any linked objects. One proxy can also link to multiple objects to
Wigdor et al. gathered eight participants to test the Rock & Rails interactions. The results Wigdor et al. reported showed that participants all presented positive attitude towards Rock & Rails.

Rock & Rails interaction makes object manipulation much easier and faster than traditional way. This technique is especially useful for graphic design on multitouch screen. It enables users to do isolated scaling, rotation and translation rapidly, which is desired by graphic designers. One of the best designs of Rock & Rails is the proxy. It perfectly avoids occlusion and improves the accuracy of the object modification. One of limitations of Rock & Rails is that its gesture set is not from natural body actions. Users cannot grasp the gesture mappings immediately based on their everyday experience and need a brief tutorial to start. Also because of this reason, users may forget the gesture mappings later. Another limitation is that Rock & Rails is a two-handed gesture interaction. So, this interaction cannot be used in one-handed devices (e.g. mobile phones and small tablets).

Planning When Manipulating Virtual Objects
Not much research work in psychology has been done on how people manipulate virtual objects displayed on screen. However, a lot of evidence from psychology research work supports that planning before the action influences people’s manipulation of objects in physical world. How people plan their acquisition and grasp to facilitate movement and optimize comfort has been the focus of a large body of work with in the fields of Psychology and Motor Control [13]. This work can be expressed using the concept of orders of planning [27].

Olafsdottir et al. conducted three experiments to test if this was also the case in multitouch interaction [13]. The task for the participants is to move and rotate the object from start location to the target location. See Figure 5 for task scenario.

Wigdor also created an action called “Ruler”, which is very useful but needs more demonstration to the users. Users can use ruler to do 1D translation and rapid alignment. 1D translation is achieved by a Rail hand shape placed next to the object. The design allows the ruler to be at any position and orientation based on the Rail hand shape. If the object is active (selected by the fingertip), the ruler will be snapped to the object. Once a ruler is created, object can only be translated along the ruler in only one dimension. Once a ruler is created on one object’s bound, users can also translate other objects towards the ruler to make them align with the ruler. This action is called “alignment”.

Isolated uniform scale is achieved by a Rock hand shape and one or more fingertip to uniformly scale the object. See Figure 2. Isolated non-uniform scale is achieved by a Rail hand shape and a fingertip moving perpendicular to the palm of the non-dominant hand, see Figure 3. Isolated rotation is achieved by a Curved Rail hand shape and a fingertip rotating the object, see Figure 4.

All the actions described above are fairly straightforward. Users can quickly grasp the way to do these manipulations.
tools can only leverage the familiarity of part of the physical screen. For those techniques with touchscreen only, the gesture sets are often limited by the contact shapes captured by the screen. For example, because of this limitation, TouchTools can only leverage the familiarity of part of the physical tools [14]. Therefore, future work could include solving this problem. One possible solution is to detect the hovering whole hand postures. Though techniques such as extended multitouch interaction is able to solve this problem, they all have limitations, for instance, extended multitouch interaction requires additional depth camera. So future work could try to solve that by integrating the sensor in the device, which enables the techniques to be applied to mobile devices.

Rock & Rails is a two-handed multitouch interaction technique [22]. So it is more suitable for devices with big screens. For devices with smaller screen, such as iPad, people are usually more comfortable doing one-handed manipulation with the other hand holding the device. As portable devices are prevalent nowadays, research on one-handed multitouch interaction is also a topic of great importance. So a possible future research direction is to propose more one-handed gesture sets.

TouchTools found people could easily discover the mouse, camera, and whiteboard eraser tools [14]. So researchers can continue this work to discover other tools that can make use of TouchTools.

The relationship between physical manipulation and virtual manipulation is still a new field to explore. Though Alzayat has proved that they are not exactly the same, and many evidence showed that leveraging the physical manipulation is a very useful way to improve user experience, many questions are still waiting us to answer. For example, more experiments are needed to eliminate alternative hypotheses except the one proposed by Alzayat [17], which explains the difference between physical condition and virtual condition. So continuing the work in this field is also a good direction for future research.

CONCLUSION
Early researches and current researches on multitouch and gestures are reviewed in this paper. Firstly, brief introduction of multitouch research from the beginning to about five years ago was provided as background literature review. Then this paper spent great effort discussing the amazing research work done in recent years. Though none of the techniques discussed in this paper is perfect for multitouch interaction, they all made great contribution to the multitouch interaction and provided a source of inspiration for other researchers. From the literature review, we can see that this area is new and challenging, much of them still need to be explored.

ACKNOWLEDGMENTS
I would like to thank ACM Digital Library and Google to provide me various amazing papers for my literature review. I also gratefully acknowledge the guidance and teachings of Olivier St-Cyr and Aakar Gupta.

Figure 6. The position of interactive object sex pressed in polar coordinates\((r, \theta)\), where \(r\) is the radial distance and \(\theta\) is the clockwise angle with respect to the vertical axis of the screen. The grip orientation is expressed by the clockwise angle \(\phi\) defined by the thumb and the index [13].

The experiments measured the initial grasp orientation \(\phi_{init}\) when participant started move the object. Figure 6 demonstrates what the grasp orientation is. Another kind of values that has to be measured is \(\phi_{default}\), which is the task-independent grasp orientation, when participants do not have to do any manipulation to the objects.

Experiment 1 showed that the users choose their \(\phi_{init}\) based on the difference between \(\phi_{default}\) values of start object and target object. Experiment 2 showed that users tend to end the task with comfortable hand positions, so the \(\phi_{init}\) is chosen according to this reason. Experiment 3 included both translations and rotations in the tasks and showed that both of the above effects occur in parallel, with planning for rotations having a stronger effect [13].

This study proved that manipulation of virtual objects is also influenced by planning and motor control. This research field is of great significance, because of the fact that the positions on the screen constrain the motions and grasps of users. For example, if the position is very close to the user, the gesture for manipulation is of great freedom. But if the position is very far, user can only just reach it, then the allowed gestures are limited. Hence, knowing planned movement before the user’s action is valuable for improving user experience [13]. With this knowledge, UI designers can make better positioning of objects on the screen.

FUTURE RESEARCH
For those techniques with touchscreen only, the gesture sets are often limited by the contact shapes captured by the screen. For example, because of this limitation, TouchTools can only leverage the familiarity of part of the physical tools [14]. Therefore, future work could include solving
REFERENCES