# Separability of Spatial Manipulations in Multi-touch Interfaces

Miguel A. Nacenta[1], Patrick Baudisch[2], Hrvoje Benko[2] and Andy Wilson[2].

[1]University of Saskatchewan [2]Microsoft Research

## ABSTRACT

Multi-touch interfaces allow users to translate, rotate, and scale digital objects in a single interaction. However, this freedom represents a problem when users intend to perform only a subset of manipulations. A user trying to scale an object in a print layout program, for example, might find that the object was also slightly translated and rotated, interfering with what was already carefully laid out earlier.

We implemented and tested interaction techniques that allow users to select a subset of manipulations. Magnitude Filtering eliminates transformations (e.g., rotation) that are small in magnitude. Gesture Matching attempts to classify the user's input into a subset of manipulation gestures. Handles adopts a conventional single-touch handles approach for touch input. Our empirical study showed that these techniques significantly reduce errors in layout, while the Handles technique was slowest. A variation of the Gesture Matching technique presented the best combination of speed and control, and was favored by participants.

**KEYWORDS:** Tabletops, separability, multi-touch interaction.

**INDEX TERMS:** H5.2.User Interfaces: Input devices and strategies

## 1 INTRODUCTION

Multi-touch interfaces allow users to apply multiple spatial transformations to a virtual object with a single combined gesture. Using two fingers, for example, users can translate, rotate, and scale a photograph simultaneously. The increase in interaction bandwidth afforded by multi-touch has two main potential advantages: it could improve the speed of complex manipulations because operations need not be applied sequentially, and it is often referred to as "natural" because it resembles how we manipulate objects in the physical world.

Unfortunately, multi-touch gestures also make it difficult to perform only one (or just a subset) of the available operations at a time. For example, it becomes hard to only scale and translate an object (without rotating it) because the object will also react to small variations of the angle between the contact points.

Figure 1 shows a scenario in which a designer is laying out a poster for a chess competition. By enlarging and moving one of the pawns of the original figure (Figure 1A), the designer intends to create a sense of depth (Figure 1B). However, when performed on a multi-touch interface, this might result in the pawn not only being translated and scaled, but also rotated (Figure 1C). In many cases, this will interfere with what was carefully laid out earlier and stand out as a flaw, because humans are highly sensitive to variations in rotation and scale [19, 25].

We investigated four strategies to allow users to constrain multi-finger interaction to any subset of translation, rotation, and scaling manipulations while preserving, as much as possible, the freehand nature of the interactions: Handles, Magnitude Filtering, and two variants of Gesture Matching (Frame-to-Frame and First-Touch). The Handles technique allows users to restrict manipulations explicitly. It offers one handle for each possible manipulation style; users select a manipulation by picking the corresponding handle. Magnitude Filtering acts upon touch input only when the resulting rotation, scaling, and translations exceed a minimum amplitude; small movements are filtered out. The Gesture Matching techniques help users avoid undesired manipulations by guessing which kind of manipulation the user is performing (e.g. rotation+translation, translation only).



Figure 1. Poster design scenario: A) initial state B) desired result (translated & scaled) C) likely result (translated, scaled & rotated).

A user study shows that Handles, First-Touch Gesture Matching and Magnitude Filtering reduce the number of unwanted side-effect manipulations by up to 90%, although Handles does so at the expense of increased manipulation time. In addition to technique comparisons, we also performed a movement analysis of unconstrained gestures that allowed us to characterize multi-touch interaction in terms of expected error, simultaneity and order of the different manipulations.

Our findings indicate that the First-Touch Gesture Matching technique and the Magnitude Filtering techniques are well-suited for bringing the benefits of multi-touch interfaces to layout tasks that would otherwise be difficult due to the extended control freedom. Our movement characterization also provides groundwork that can inform the design and configuration of future techniques.

## 2 RELATED WORK IN MULTI-TOUCH

Recent advances in input technology have resulted in a broad range of multi-touch devices, i.e., devices that can track more than one contact point (usually a finger) simultaneously. Compared to single-point input, Multi-touch offers additional degrees of freedom. These have often been mapped to spatial manipulations such as rotation and scaling [12, 17].

Two main motivations drive Multi-touch research. First, since is resembles the way humans manipulate physical objects, Multi-touch can lead to more "natural" interactions. Second, Multi-touch is expected to be more efficient, because it allows users to manipulate multiple degrees of freedom simultaneously [11].

Multi-touch spatial manipulation shares some problems with touch screens including the lack of stability on release [17] and the fat finger/occlusion problem [22]. In this paper, however, we focus on issues related to the added number of contacts.

Although bimanual interaction (e.g. [11]) can be considered multi-touch, multi-touch does not necessarily require the use of multiple hands. In this paper, we focus on a common mode of multi-touch interaction: manipulation of a single object using multiple fingers of the same hand. Users commonly manipulate objects with a single hand, especially if they are located in an area that is hard to reach with the other hand, if the object is too small for two hands, or if the other hand is used for something else, e.g.

nacenta@cs.usask.ca,{baudisch, benko, awilson}@microsoft.com
[1]Computer Science Department, University of Saskatchewan, 110 Science Place, S7N 5C9, Canada.
[2]One Microsoft Way, Redmond, WA 98052-6399, US.

if it is manipulating a different object, is gesticulating, or is maintaining a posture. See [15, 16] for a detailed discussion on the differences between same-hand and bi-manual interaction.

## 2.1 Characteristics of multi-dimensional manipulation

Many studies have looked at the characteristics of input in systems that allow for more than the standard two degrees of freedom. Among others, Zhai and Milgram [26], Masliah and Milgram [14], and Wang et. al. [23] analyze 3D docking tasks. Van Rhijn and Mulder investigate slicing of 3D anatomical data [21]. Mason and Bryden observe rotation tasks of real objects [15]. Latulipe et al. found performance benefits for two mice when performing bi-manual alignment tasks [11]. Buxton and Myers examine bi-manual translation and scaling [6], and Balakrishnan and Hinckley observe a continuous bi-manual tracking task [2].

Other researchers have looked at providing extended manipulations (e.g. rotation) with the standard DOF of single-point input [8, 10, 20]; however, neither this research, nor the work from the previous paragraph has focused on the issue of *separability*.

## 2.2 Manipulation separability

Not all interactions on a multi-touch device are intended to produce all possible manipulations, but isolating one or more manipulations can be difficult. In [17], Moscovich and Hughes observed that "due to physiological constraints on finger motion it is difficult to rotate the fingers while keeping them at a precisely fixed distance [to keep the same scale]". In 3D spatial manipulation a similar effect was observed by Ware [24], who found that 3D docking was harder when participants had control over all dimensions than when one of them was "locked". We identify these as two instances of lack of separability, where separability is defined as the ability to purposefully avoid variation in one or more of the available manipulations (e.g. rotate but not scale).

Our definition is inspired by Jacob and colleagues' integrality and separability concept [9]. They propose that tasks that require multiple manipulations can be integral or separable according to their perceptual structure; for example, translation and scaling are integral because they are perceptually related, whereas translation and color are separable because they rely on separate perceptual mechanisms. Input devices can also be integral or separable (depending on how the different degrees of input are linked), and they show in an experiment that matching the integrality of the input device and the task results in better performance. Our use of the separability concept is different; although the input and perceptual structure of our task and device would be considered integral in their terms, it is sometimes desirable to achieve separation of control of the different manipulations. We believe that separability can be improved through the design of appropriate interaction techniques.

Separability is also related to Zhai and Milgram's concept of Efficiency [26]. The efficiency of a gesture refers to the simultaneity of manipulation of different dimensions. To quantify efficiency they used a formula that compares the Euclidean distance between the starting position and the final state (considering each manipulation as one of the Euclidean dimensions), to the actual trajectory achieved by the user. Note that this efficiency does not necessarily relate to faster interaction; a gesture could be very slow but also efficient if all dimensions are being manipulated simultaneously at the same rate until the final state is reached. Separability implies efficiency because avoiding unwanted manipulation in one or more dimensions makes for a shorter Euclidean trajectory.

## 2.3 Related problems

Separability is related to but different from the problem of alignment. The purpose of alignment is to make dimensions equal to pre-set values (guides, preferred directions) or to the dimensional values of other elements in the interface (alignment, equalization). Solutions for alignment include guides, snapping [3], and the use of alignment tools, such as the alignment stick [18].

The goal of separability, in contrast, is to keep certain dimensions unchanged without such external references. In introducing techniques to interactively control separability in multi-touch interfaces, our hope is to give the user selective control while preserving much of the flavor of existing multi-touch interactions.

## 3 INTERACTION TECHNIQUES FOR SEPARABILITY

To provide manipulation separability in multi-touch interfaces, we have explored a number of techniques of which we selected the following four for evaluation: Handles, Magnitude Filtering, and two variants of Gesture Matching.

The selected techniques assume the use of two or more fingers. We acknowledge that there are other ways to ameliorate the lack of separability: for example, the iPhone interface rarely permits translation, rotation and scaling simultaneously (separability becomes less of an issue when fewer manipulations are possible); using virtual tools such as pins or guides [4] can help lock certain dimensions; and the number of touches (fingers) can be used to determine which manipulations are active [7] (one finger means translation only, two fingers means rotation only, and so on).

Most of these alternative strategies are compatible with our approaches (we do not study the separability of translation-only tasks because it is more meaningful to assign one-touch interactions to translation only). However, these approaches also have important shortcomings: the assignment of number of fingers to operations is somehow arbitrary above the distinction between one finger and more; and using pins and guides for manipulation requires extra steps in the interaction that may slow down the action and complicate the interface.

## 3.1 Handles

Single-touch interfaces typically require explicit mode changes. Usually, the modes take the form of handles, which are special regions on the object that are assigned to a certain manipulation. For example, in PowerPoint, the user may grab the object with the cursor on specific handles that determine whether the figure is rotated (small green handle in Figure 2.A), scaled (handles in the corners) or stretched (handles in the middle of the sides).
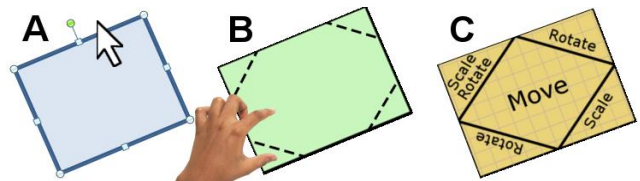


Figure 2. Different kinds of handles: A) Standard cursor handles are too small for touch input; B) Apted et al.'s [1] handle implementation for rotation/scaling, and translation; C) Our own implementation with areas dedicated to specific operation combinations.

We included Handles in our study because it is common in current single-point interfaces and it has been used before in the multi-touch context [1]. Handles increases separability because the operations are explicitly selected by the user at the moment of touch. In order to prevent an object from rotating, users simply avoid touching the "rotate" handle.

We modeled our implementation of the Handles technique after Apted et al.'s design [1] but modified it in order to allow for separate control of the rotate and scale dimensions (Figure 2C). We also added labels so that users could identify the operation associated to each handle. We use text labels on the handles instead of icons in order to avoid the possible ambiguities of icon interpretation during the evaluation (e.g., an icon for scale+rotate could be easily confounded with an icon for rotate only).

Our prototype supports multi-touch interaction. For example, a translation+rotation gesture can be achieved by placing one finger in the move handle (which activates translation) and another in the rotate area (which activates rotation). Two fingers on the move handle will not cause any rotation or scaling, just as any number of fingers in the scale handle will not change the position or orientation of the object.

## 3.2 Magnitude Filtering

The *Magnitude Filtering* technique filters each manipulation of a multi-touch gesture transformation (rotation, scale, translation) such that values below a certain threshold magnitude produce no effect. For example, we may interpose a function between input and output such that the object will only rotate if the rotation indicated by the contact points exceeds 30° (with respect to the original orientation) (Figure 3). Separability is achieved because users can make the desired manipulations large (over the threshold), while small manipulations are ignored. This technique works regardless of where the object is touched because the rotation is calculated using the angle between the line formed by the initial points, and the line formed by the current points, regardless of their position in the object.
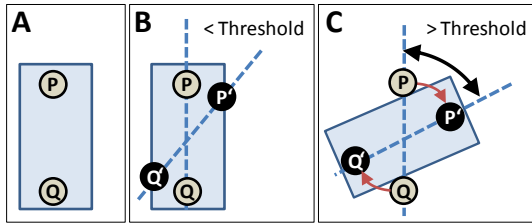


Figure 3. Rotation filtering A) initial touch points P and Q on an object B) the touch points rotate to the P' and Q' positions, but the angle is not yet above the rotation threshold C) further rotation is above the threshold and the object rotates to the angle indicated by the current touch points.

This technique was inspired by snapping techniques that enlarge the motor space over the desired snap locations (e.g. screen limits, or pre-selected values of the *x* or *y* coordinates – guides) (e.g., [3, 5, 13]). Magnitude Filtering differs from snapping techniques in the following two aspects. First, Magnitude Filtering "snaps" objects only to the object's initial state, making it easier for an object to maintain its initial rotation, initial scale, or initial position or to return to it. Snapping techniques, in contrast, generally snap to pre-selected values or to other elements in the environment. Second, we introduce a *catch up zone* where the transformations are amplified to allow a continuous transition between the *snap zone* (where variations of the input do not affect the output) and the *unconstrained zone* (where the output corresponds exactly with the input). This makes all target positions obtainable (a concept introduced by snap-and-go [3]) and allows dragged objects to catch up with the finger dragging it (unlike snap-and-go), thereby preventing excessive separation of finger and object. Figure 4B explains the concept by comparing it to snapping (Figure 4A).
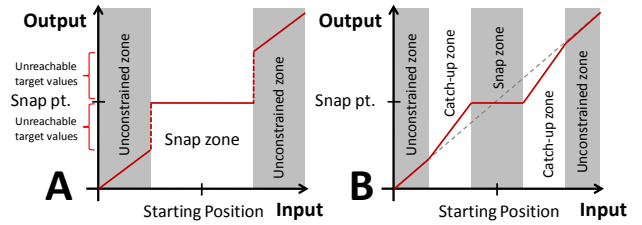


Figure 4. Filtering functions. A) Snap B) Snap with buffer zone.

Since the algorithm avoids abrupt transitions between zones, interactions using Magnitude Filtering feel smooth. Initially, objects offer some resistance to change in each dimension. When the gesture becomes large enough in one of the dimensions (e.g. scale) the object starts changing fast until the user's fingers have caught up with the initial contact point. Further expansion proceeds as with a regular unconstrained manipulation.

## 3.3 Gesture Matching

The *Gesture Matching* techniques explore the idea that when users desire a pure rotation (for example) they may strive to provide an input gesture that itself is a pure rotation. These techniques are based on a battery of different models that try to explain the combined motion of all the touch points on the object. Each model tries to minimize the mean root square difference between the actual motion and a motion generated with the manipulation subset of each model. There are models for simple gestures (translation) and compound gestures (rotation+translation, scale+translation, and rotation+scale+translation). The technique selects the simplest manipulation mode that still explains the actual motion reasonably well. For example, if we find that the translation-only model approximates the motion well enough, we will not engage the more complex rotation+translation model.

Given a set of starting and ending positions of touch points, each model generates two outputs: the error of the best fit of the model to the data (i.e., how distant are the predicted points from the actual points), and the magnitudes of rotation, scale and translation that minimize that error.

The error outputs of each model are then collected by a decision algorithm that chooses which model to apply (see Figure 5): errors are normalized using a sigmoid function, compared to the error of the active model, and subtracted a configurable parameter. The system changes to a new state when, after subtraction, the error of the corresponding model is still lower than the error of the current model. Naturally, the models with the most parameters will generate the least error, and therefore the configurable parameters have to penalize more the more complex models (this process is analogous to *regularization* in the machine learning field). Parameters for each transition can be configured individually.
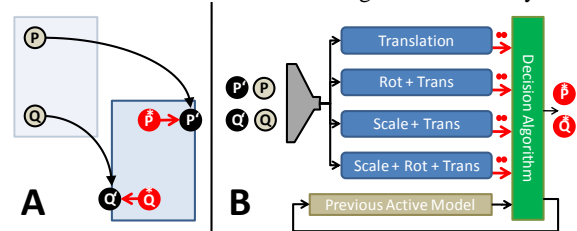


Figure 5. Diagram of the Gesture Matching Technique. A) Notation: P, Q (previous touch points); P', Q' (current touch points); P*, Q* (estimated current touch points – translation model), red arrows (estimation errors). B) Schematic of the technique implementation.

The reader might notice that the list of models in Figure 5 does not include rotate-only or scale-only models. The reason is that multi-touch gestures that we might consider pure rotation or scaling are usually a combination of rotation+translation and scaling+translation respectively. For explanation, consider the rotation gestures depicted in Figure 6; all can be considered strictly rotations, but each uses a different implicit rotation center. Instead of trying to deduce which center was meant for rotation or arbitrarily deciding on one, we decided to merge the rotation and translation models into one, i.e., rotation and translation explain a rotation gesture around *any* center. The reasoning is analogous for scale.
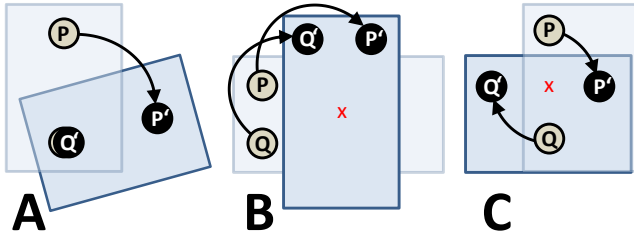


Figure 6. Rotation movements around different rotation centers: A) rotation center is in Q (the Q contact point does not move) B) rotation center is in the center of the object C) rotation center is in the mass center of the contacts.

### 3.3.1 Frame-to-Frame Gesture Matching

Our two variants of the Gesture Matching technique differ in the period of time over which the models are fitted. In *Frame-to-Frame Gesture Matching*, models are fitted each time step using the previous frame and the current frame, and a decision is made each frame as to which manipulation model is used to control the object. Most interactive systems require frame rates of 30Hz or more. If the technique were to select a different model every few milliseconds, the behavior of the object would be very similar to the unconstrained movement (it can interleave many alternating types of short manipulations to achieve any desired final result). To avoid this we added hysteresis to the selection process: the configurable parameters add resistance to leave the current mode.

The Frame-to-Frame Gesture Matching technique is very flexible because it allows for sophisticated behavior configurations; for example, certain transitions (such as rotation+translation to rotation+translation+scale) can be made more difficult, allowing for a certain "feel" of the manipulation. However the technique proved hard to configure because the information contained in frame to frame variation is often not enough to distinguish one type of gesture from another reliably, and so thresholds must be set high. Users must then indicate a change of manipulation mode through a fast gesture that has a strong component of the desired manipulation. For example, an object in rotation mode needs a fast pinch gesture before it will start scaling.

### 3.3.2 First-Touch Gesture Matching

To avoid the problems of the Frame-to-Frame gesture matching, we implemented a variant, called *First-Touch Gesture Matching*, that fits the same models over the duration of the gesture. That is, it uses the touch positions from the first frame of the gesture (first touch) and compares it to the most current data.

The result is a more stable estimation of the gesture (the models have much clearer changes to discriminate the manipulation). In exchange, when any of the thresholds is surpassed, the object jumps to a new position (as fit by the new model). For example, if a gesture starts with a slight scaling movement the object will start scaling. If after a while the touch points start to rotate, hysteresis will keep the scaling mode until it is decided that rotation is a much better match; at that moment the object will return to its

original size and rotate to match the current rotation of the touch points. If the rotation has gone beyond the desired rotation, the user can rotate back a small amount without switching into scale mode (the hysteresis will prevent the mode change unless the touch points separate again significantly).

First-Touch Gesture Matching has the configurability of the Frame-to-Frame version, but it does not require fast gestures to activate different modes. In fact, the behavior of this technique resembles that of Magnitude Filtering. There are several important differences between the First-Touch Gesture Matching and the Magnitude Filtering techniques: First-Touch Gesture Matching will jump to a new position whenever a better fit is found, whereas Magnitude Filtering will never "jump", but change gradually instead. To access the object positions abandoned by a jump, the First-Touch Gesture Matching requires returning towards the initial state, whereas Magnitude Filtering can reach any magnitude in a monotonous movement; and First-Touch Gesture Matching can distinguish between composite manipulations (e.g. rotation and translation), whereas in Magnitude Filtering each manipulation is independent. This last difference is irrelevant for simple configurations, but it can help design a better technique if we know that certain combinations of manipulations are more likely to take place together or in certain orders.

## 4 EMPIRICAL EVALUATION

We designed our evaluation with two goals in mind: to compare the different alternatives that enhance separability in multi-touch interfaces; and to gain insight into the nature of unrestricted motions that could help us design better techniques in the future.

### 4.1 Techniques

We tested the four techniques described in the previous section and a baseline condition that does not constrain rotation, scale or translation; we call this the *Unconstrained* technique.

The Magnitude Filtering technique snap zones were configured as follows: translation 20 pixels (in each direction); rotation $11.25°$ (for each, positive and negative angles); scale 20% (for each, enlargement and reduction). Buffer zones were set to the same size than the corresponding snap zones. The other two techniques (Frame-to-Frame and First-Touch Gesture Matching) were configured through an iterative process that resulted in thresholds similar to those of the Magnitude Filtering configuration (numeric configuration values of the different techniques are not comparable because of the different implementations).

### 4.2 Apparatus

The experiment was run on the commercially available version of Microsoft Surface, which provides a touch input rate of 60Hz. The size of the interactive area is 76.2cm diagonal (30") for a 1024x768px image (4:3 aspect ratio – see Figure 7A).
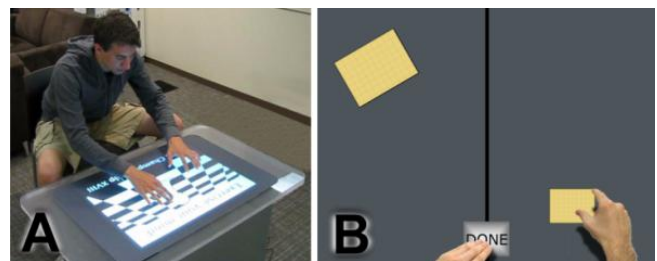


Figure 7. A) Experimental setting. B) Beginning of a trial.

### 4.3 Tasks

For each trial, participants manipulated a rectangular object (initial size 10x7.5cm) located on the right side of the screen until it matched the scale, rotation, and relative position of a reference object displayed on the left side of the screen (Figure 7B). Participants were instructed not to change dimensions that already matched the reference object. Participants pressed a button with their non-dominant hand to end the trial.

Some trials required manipulating only location, only rotation, or only scale. Others required changing two or all three manipulations, resulting in seven types of trials. Each manipulation had two possible values: short and long trajectories for translation (12 and 21cm respectively), small and large rotation (30° and 60° respectively), and small and large enlargement (50% and 100% size increase respectively). There were tasks for all combinations of values (2 rotation, 2 scaling, 2 translation, 4 rotation+scaling, 4 rotation+translation, 4 scaling+translation, and 8 rotation+scaling +translation): 26 different tasks overall.

In order to control noise and make technique comparisons fair, participants were instructed to always grab objects with two fingers. Participants were also encouraged (but not required) to complete each trial with a single gesture, i.e., without releasing and re-acquiring the object.

### 4.4 Participants and Study design

15 participants (8 female, 7 male) of ages between 18 and 59 participated in the study in exchange of gratuity. All participants were right handed except one, who could use either left or right hand as dominant.

The experiment was divided into two parts. The first part consisted of a single block of 78 trials with the Unconstrained condition (3 repetitions of each task, in random order), preceded by training (1 trial of each task – 26 trials). In the second part, the participants performed five blocks like that of the first part, one for each condition (including Unconstrained). The order of the conditions was assigned through a random Latin square and the presentation of each task followed no predictable order.

Dividing the experiment into two blocks corresponds to the goals of analyzing unconstrained gestures and comparing the proposed techniques. To analyze Unconstrained gestures without any bias from our proposed techniques we ran it first; we replicated the Unconstrained block in the second part to avoid biasing against Unconstrained due to possible learning effects. Instead, our design may have biased this condition positively.

At the end of the experiment the subjects filled in a questionnaire about their technique preferences.

### 4.5 Results: Technique comparisons

To compare how well techniques achieved separability we performed statistical analyses on the final rotation and scale errors of trials that did not require rotation or scaling respectively. We did not perform analysis on translation-only tasks because of a similar problem to that discussed in section 3.3 (it is unclear what a translation-only motion is when there is rotation and scaling involved) and because the translation-only case is less relevant (see discussion at the beginning of section 3). The analysis does not include the data from the first part of the experiment (one block of unconstrained trials). Data from one participant in the Frame-to-Frame condition was lost due to an error, and therefore the participant's data is removed from the repeated measures analyses. All significant differences reported in the post-hoc analysis are signficant at the 0.05 level after applying Bonferroni's correction.

Two-way repeated measures ANOVAs with technique and task as factors showed a strong main effect of technique on rotation errors ($F_{4,52} = 38.0$, $p < 0.001$, $\eta^2 = 0.74$) and on scale errors ($F_{4,52} = 19.6$, $p < 0.001$, $\eta^2 = 0.60$). The rotation data does not meet the sphericity assumption, but the corrected test (Greenhouse-Geisser) shows the same results.

The post-hoc analyses show that Frame-to-Frame Gesture Matching has the largest rotation error average ($\mu = 2.6°$), significantly larger than the rest. Unconstrained gestures followed ($\mu = 1.3°$), also significantly larger than the other three. The three remaining techniques had smaller errors, but not significantly different from each other ($\mu_{Handles} = 0.4°$, $\mu_{First-Touch\_G.M.} = 0.3°$, $\mu_{Magnitude\_Filtering} = 0.2°$).

Results for the scale errors follow similar lines, but the most error-prone were now Unconstrained ($\mu = 7\%$) and Frame-to-Frame Gesture Matching ($\mu = 4\%$), not significantly different from each other. Both techniques were statistically different from the rest ($\mu_{Magnitude\_Filtering} = 2\%$, $\mu_{First-Touch\_G.M.} = 2\%$, $\mu_{Handles} = 1\%$) except for the comparison between the two Gesture Matching techniques. These results are summarized in Figure 8.
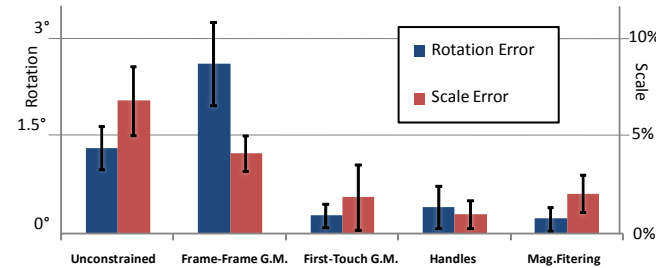


Figure 8. Rotation and Scale errors for no-rotation and no-scale tasks. Error bars represent 95% confidence intervals.

We also tested the proportion of trials with any rotation and scale errors (for tasks that did not require either rotation or scale changes). The results follow the trend of the previous analysis: Unconstrained resulted in the maximum percentage of trials with error (98%), followed by Frame-to-Frame Gesture Matching (65%), Handles (21%), Magnitude Filtering (8%) and First-Touch Gesture Matching (6%). Scale errors shuffle the pattern except for Unconstrained, that still shows the most trials with error (98%), followed by Magnitude Filtering (42%), Frame-to-Frame Gesture Matching (36%), First-Touch Gesture Matching (23%), and Handles (13%). These results are summarized in Figure 9.
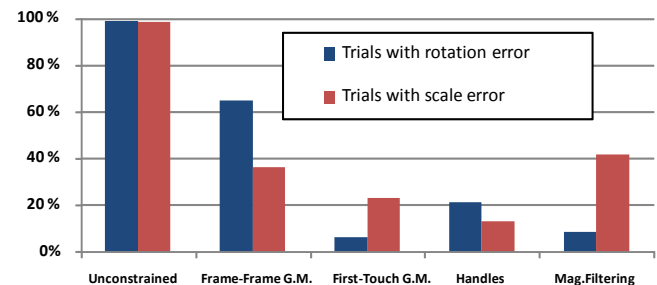


Figure 9. Percentage of trials with any rotation or scale errors for tasks that do not require rotation or scale (respectively).

Two non-parametric Friedman analyses of the percentage data grouped by user and technique indicate a main effect of technique in both rotation and scale percentage of trials with errors ($\chi^2_{rot}(4) = 44.3$, $p < 0.001$, $\chi^2_{scale}(4) = 43.0$).

We also performed a repeated-measures ANOVA on log-transformed task completion times with the same two factors (task and technique) to find out which techniques were faster. Time was measured from the moment that the user touched the object for the

first time and transformed logarithmically (as is usual for linear analysis of temporal data). All means presented henceforth are back-transformed from the logarithmic domain. The analysis shows a strong main effect of technique ($F_{4,52} = 8.3$, $p < 0.001$, $\eta^2 = 0.39$). The post-hoc analyses show that Magnitude Filtering ($\mu = 2,622ms$), Unconstrained ($\mu = 2,629ms$), and First-Touch Gesture Matching (FT) ($\mu = 2,779ms$) were fastest (and statistically indistinguishable from each other); while Frame-to-Frame Gesture Matching ($\mu = 3,482ms$), and Handles ($\mu = 3,482ms$) were significantly slower than the other three.

From pilot studies we observed differences between techniques in the time users took to start interacting with the object after each trial started. To test these differences we performed an ANOVA on the time to start gesture. The result shows a strong main effect of technique ($F_{4,52} = 49.3$, $p < 0.001$, $\eta^2 = 0.79$). Post-hoc analysis show that it took significantly longer to start with Handles ($\mu = 1,404ms$) than with any of the other techniques ($\mu_{Frame-to-Frame\_G.M.} = 775ms$, $\mu_{First-Touch\_G.M.} = 708ms$, $\mu_{Unconstrained} = 692ms$, $\mu_{Magnitude\_Filtering} = 686ms$). Results are summarized in Figure 10.
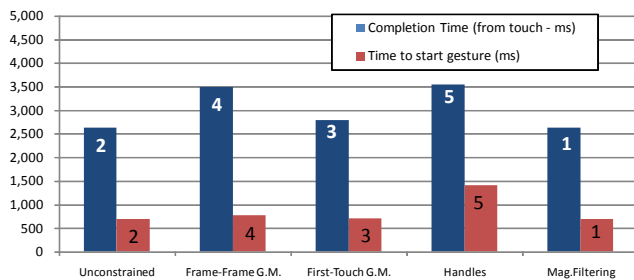


Figure 10. Completion and time to start trials. The sum of the two columns is the total trial time. Units are milliseconds. Numbers in the bar indicate order from 1 (fastest) to 5 (slowest).

## 4.6 Results: Subjective

Across participants Magnitude Filtering and First-Touch Gesture Matching were ranked as the preferred techniques, closely followed by Unconstrained and, at a distance, by Handles and Frame-to-Frame Gesture Matching (see Table 1). Ordered non-parametric statistical contrasts of the technique preferences (pairwise Wilcoxon Signed Ranks) showed statistical differences between Handles and Frame-to-Frame Gesture Matching and the other three. Participants also ranked the techniques in terms of speed and accuracy with very similar results (not reported here).

Table 1. Subjective preferences (# users assigning the rank).

| | 1 best | 2 | 3 | 4 | 5 worst | Mean |
|---|---|---|---|---|---|---|
| Magnitude Filter. | 5 | 7 | 0 | 3 | 0 | 2.07 |
| First Touch G.M. | 6 | 2 | 4 | 3 | 0 | 2.27 |
| Unconstrained | 3 | 5 | 7 | 0 | 0 | 2.27 |
| Handles | 1 | 1 | 4 | 5 | 4 | 3.67 |
| Frame-Fr. G.M. | 0 | 0 | 0 | 4 | 11 | 4.73 |

Many participants disliked the Handles technique because "[I have] to think a little bit more" and "I cannot just automatically instinctively do it [manipulate the object]". These comments refer to the fact that, with Handles, the type of movement must be decided before contact, whereas with the other techniques you can decide as you go. Several participants commented that the Frame-to-Frame Gesture Matching technique was difficult to control; a participant noted that "it has a mind of its own".

## 4.7 Results: Characterization of Unconstrained Gestures

Each participant ran a block of unconstrained trials before they used any other technique. We collected these data to understand the basic characteristics of unconstrained rotation-scale-translation gestures in multi-touch interfaces and to look for general patterns that could help us design the next generation of techniques. This section discusses three analyses: gestural noise, allocation of control and manipulation order.

Each of the signals referred to in the following sub-sections were conditioned using standard human movement signal processing procedures: the signal (variation of a magnitude in time for a given trial) was recorded directly by our software at a typical sampling rate of 60Hz; then it was resampled at 50Hz to correct for sampling period variability, then padded and processed through a four-order low-pass Butterworth filter (cut-off frequency: 8Hz). The signals of manipulations that changed were further differentiated to find the rate at which error was reduced towards the goal state, and filtered at 4Hz.

### 4.7.1 Gestural noise

We analyzed trial data to characterize the expected variability of quantities that were not supposed to change; i.e., we measured the typical scale changes for tasks that do not require scaling, and orientation changes for tasks that do not require rotation.

We found that the maximum orientation error in a trial averaged across all trials that did not require rotation was 5.1° and the average maximum scale error across all trials that did not require scaling was 100% (a doubling in size). Figure 11 shows the overall distribution of *all* orientation and scale points for the trials indicated above. The distribution of the error for the different manipulations can help set appropriate parameters for the techniques.
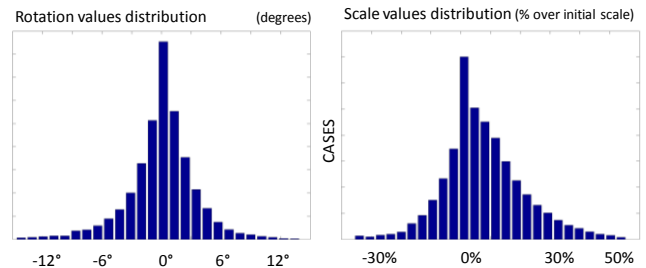


Figure 11. Histograms of the rotation (left) and scale magnitudes (right) for trials not requiring rotation or scaling respectively.

### 4.7.2 Allocation of control

An important characteristic of any multi-dimensional gesture is degree to which objects are manipulated simultaneously in several dimensions [9, 14, 26]. Several metrics of control allocation have been proposed in the input control literature, from which we chose the *m*-metric proposed [14] for being the most comprehensive.

The *m*-metric measures the degree of simultaneity of two or more signals on a continuous scale between 0 and 1, where 0 indicates that the signals never change simultaneously (e.g., they take turns in how they change) and 1 indicates that they are perfectly synchronized (e.g., one signal is an amplified version of the other). In our case we used the *m*-metric to calculate which manipulations are more coordinated with each other. We calculated three *m*-metric coefficients, one for each of the possible manipulation couples: rotation-translation, rotation-scale and scale-translation. Each unconstrained trial generated one measure for each of the combinations. A two-way repeated-measures ANOVA

with task type and manipulation-couple as factors showed a strong significant main effect of manipulation-couple ($F_{2,28} = 48.3$, $p < 0.001$, $\eta^2 = 0.77$). Post-hoc comparisons confirmed that rotation and translation are more simultaneous (average 0.43) than either scale and translation (0.32) or scale and rotation (0.32 – all post-hocs $p < 0.05$, with Bonferroni correction). These results indicate that whereas rotation and translation seem simultaneous, scaling proceeds more independently.

### 4.7.3 Order of manipulations

To learn about the temporal distribution of the different manipulations in time we performed an analysis similar to Mason et al.'s [15] and Wang et al.'s [23] temporal analysis. In a first step we normalize the signals in magnitude and time to have equal areas; then we calculate the contiguous area of the signal that contains the time of fastest change and covers the 50% of the total variation towards the goal value. This calculation gives the estimated periods when the signal experienced most of its change.

The results are summarized in Figure 12. Periods of high change occur within the first quarter of the gesture, consistent with the usual movement patterns of targeting and docking tasks. The graph also shows how manipulations start in a typical order: first translation, then rotation and finally scale. The rotation manipulation is contained within the translation manipulation, which is consistent with Wang et al.'s analysis of 3D docking problems (rotation and translation only) [23].
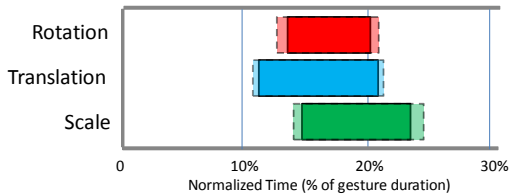


Figure 12. Duration of the periods of maximum activity of each manipulation with respect to the total duration of the gesture. Dotted lines represent confidence intervals.

## 5  DISCUSSION

We divide our discussion in three main themes: how to reduce undesired manipulations; the limitations of our experiment and what we learned about unconstrained motions.

### 5.1  Reducing Undesired Manipulation

The data from our study confirms that the lack of manipulation separability can be a problem for tasks where any error in size or rotation is important: 98% of the unconstrained gestures contain some undesired rotation or scale; and the average error amounts to 2.5° and 7% respectively.

All but the Frame-to-Frame Gesture Matching technique succeeded in reducing both the average error and the proportion of trials with errors. However, there are important differences in how the techniques perform.

### 5.1.1 First-Touch Gesture Matching vs. Magnitude Filtering

These two techniques were rated best by users and were among the most successful in reducing scale and rotation error (together with Handles). The similarity of the results is consistent with the similar behavior of the techniques; however, the main differences between these techniques are not in our empirical data, but in the way they are implemented and the configurability they afford.

Magnitude Filtering is a straight forward technique to implement and configure: each manipulation is filtered separately, and only a couple of parameters can be adjusted (the snap and buffer

zones of the transfer function). In contrast, the Gesture Matching technique requires the setup of parameters for each transition. Our experience proved that the configuration of the Gesture Matching techniques is complex; on the other hand, the large parameter space offers many possibilities. Gesture Matching techniques could be configured differently for different applications so that interaction designers have control not only of the level of noise that is tolerated in certain manipulations, but also in the way that the technique feels: for example more or less "sticky". We also believe that Gesture Matching techniques could take advantage of a deeper knowledge of human spatial manipulation gestures. For example, we could configure it to take into account the order in which manipulations are usually performed: transitions from translation to translation+rotation modes could be made easier than transitions from translation+scale to translation+rotation.

### 5.1.2 Problems with Handles

The Handles technique might seem the obvious choice for preventing unwanted operations because it is explicit and works as most single-point interfaces. However, we discovered that trials took about 50% longer than with other techniques, and that it does not reduce error better than any of the two winners. We speculate that, with handles, the user must think at touch time about the manipulations that the movement will require and also target a smaller region of the object. We also speculate that it is likely that sometimes the initial grip of the object is not accommodating enough to comfortably reach the goal position (anatomical constraints of the hands, see also [16]), requiring changes in touch positions in the middle of the gesture. The high percentage of trials with errors (21% for no-rotation tasks and 13% for no-scale tasks) also points to the difficulty of selecting proper handles in advance (grabbing the wrong handle was scored as an error). We rule out that these results are due to the difficulty in finding the correct handle because the tasks always started with the object in the same position and orientation, and the participants had plenty of opportunities (training) to learn the handles arrangement.

Although our design of handles might be improved through design (e.g. intelligent handles that adapt to the users' position, circular handles), we believe that the problems exposed by our experiment, and some other intrinsic problems of handles (visual clutter, occlusion of content, fat finger issues with small handles) should be carefully taken into account by designers when choosing multi-touch techniques for spatial manipulations.

### 5.1.3 Frame-to-Frame Gesture Matching

Even though we implemented and configured all techniques to the best of our abilities, Frame-to-Frame Gesture Matching showed very little benefit for separability, and was by far the least preferred. We believe that two factors explain the failure: there is very little information about a gesture between two frames, making the technique error-prone and hard to configure; and participants found difficult to perform the required fast gesture.

### 5.2  Limitations of the study

It is important to note several significant limitations to our study. We performed the study on one-handed multi-touch gestures on a horizontal surface. It is difficult to use our results to draw conclusions about the performance of single touch interactions, bimanual interactions or performance on vertical surfaces, and further studies are needed to investigate those comparisons.

Our study was also constrained to tasks with relatively large angle, distance and scale changes. The techniques that we compared do *not* make impossible to perform small adjustments (see details in section 3), but they might hinder these tasks. The trade-

off between the configuration of techniques to improve separability (e.g. thresholds of magnitude filters) and performance with small adjustments deserves further study.

### 5.3 Nature of multi-touch spatial transformations

Although most of the results from our characterization of unconstrained gestures do not help us address the separability problem, we believe that they are useful for understanding the nature of multi-touch movements, and can be useful to configure and inspire future techniques. For example, scale error has a larger variability than we expected, which may explain why Magnitude Filtering performed relatively poorly in the no-scaling trials (the threshold was probably too low in the scale dimension).

We also learned that unconstrained gestures have large variations in all dimensions before they approximate the goal position. This suggests that it is difficult to solve the separability issues with techniques based on thresholds that disappear once they are surpassed (unlike the tested version of Magnitude Filtering) or techniques where it is difficult to go back to the original state of the object. Our analysis also indicates that there might be benefit in considering scaling as a different class of manipulation (it is less simultaneous than the rest), and that different manipulations tend to start at slightly offset times within the gesture. All of these findings offer promising avenues for future advances.

### 5.4 Lessons for practitioners

We summarize the contributions of our study in four main statements:
- Separability can be a serious issue for spatial manipulation applications.
- Magnitude Filtering and First-Touch Gesture Matching can help improve separability.
- The Handles technique makes manipulation slower and it has intrinsic problems, although it does help separability.
- Gesture Matching techniques can be difficult to configure, although they offer configurability.

### 6 CONCLUSION AND FUTURE WORK

Layout tasks often require careful control of the ways in which objects are manipulated. Multi-touch interaction can be faster and more natural, but it also presents the problem of separability: it becomes difficult to control one of the dimensions (e.g. orientation) without slightly affecting other (e.g. size). In this research we explored four different techniques to reduce unwanted manipulations for single-hand multi-touch spatial transformations. We found that First-Touch Gesture Matching and Magnitude Filtering improve separability without negatively affecting performance, and that using Handles results in similar gains in separability, but at the cost of extra interaction time.

In the future, we are interested in developing new techniques based on what we have learned. For example, in the case of Gesture Matching, more sophisticated temporal models such as Hidden Markov Models may allow the calculation of transforms from moments other than "first touch". Also interesting are techniques that balance the explicitness of the handles and the implicitness of Gesture Matching; for example, techniques that give subtle cues about what mode it is about to activate, and allow the user to react in consequence. Finally, we are also exploring how to improve separability in situations where many other manipulations are available (e.g. stretching, shearing and perspective transforms).

#### REFERENCES

1. Apted, T., Kay, J., and Quigley, A. 2006. Tabletop sharing of digital photographs for the elderly. *Proc. CHI'06*, 781-790.

2. Balakrishnan, R. and Hinckley, K. 2000. Symmetric bimanual interaction. *Proc. CHI'00,* 33-40.

3. Baudisch, P., Cutrell, E., Hinckley, K., and Eversole, A. 2005. Snap-and-go: helping users align objects without the modality of traditional snapping. *Proc. CHI '05*, 301-310.

4. Beaudoin-Lafon, M. and Lassen, M. 2000. The architecture and implementation of a Post-WIMP Graphical Application. *Proc. UIST'00*.

5. Bier, E. A. and Stone, M. C. 1986. Snap-dragging. *Comput. Graph.* 20, 4 (Aug. 1986, 233-240.

6. Buxton, W. and Myers, B. A. 1986. A Study in Two-Handed Input. *Proc. CHI'86*, 321-326.

7. Hancock, M., Carpendale, S., and Cockburn, A. 2007. Shallow-depth 3D Interaction: Design and evaluation of one-, two- and three-touch techniques. *Proc. CHI'07*, 1147-1156.

8. Hancock, M. S., Vernier, F. Wigdor, D., Carpendale, S., and Shen, C. 2006. Rotation and translation mechanisms for tabletop interaction. *Proc. Tabletop'06*, 79-86,

9. Jacob, R. J., Sibert, L. E., McFarlane, D. C., and Mullen, M. P. 1994. Integrality and separability of input devices. *TOCHI.* 1, 1 (Mar. 1994), 3-26.

10. Kruger, R., Carpendale, S., Scott, S., and Tang, A. 2005. Fluid Integration of rotation and translation. *Proc. CHI*'05, 601-610.

11. Latulipe, C., Kaplan, C. S., and Clarke, C. L. 2005. Bimanual and unimanual image alignment: an evaluation of mouse-based techniques. *Proc. UIST'05*, 123-131.

12. MacKenzie, I. S., Soukoreff, R. W., and Pal, C. 1997. A two-ball mouse affords three degrees of freedom. *Ext. Abs. CHI'97.* 303-304.

13. Mandryk, R. L., Rodgers, M. E., and Inkpen, K. M. 2005. Sticky widgets: pseudo-haptic widget enhancements for multi-monitor displays. *Ext. Abs. CHI'05,* 1621-1624.

14. Masliah, M. R. and Milgram, P. 2000. Measuring the allocation of control in a 6 degree-of-freedom docking experiment. *Proc. CHI'00,* 25-32.

15. Mason, A.H. & Bryden, P.J. 2007. Coordination and concurrency in bimanual rotation tasks when moving away from and toward the body. *Exp. Brain. Res.* 183. Springer. 541-556.

16. Moscovich, T. and Hughes, J. F. 2008. Indirect mappings of multi-touch input using one and two hands. *Proc. of CHI'08*, 1275-1284.

17. Moscovich, T. and Hughes, J. F. 2006. Multi-finger cursor techniques. *Proc. Graphics Interface'06*, 1-7.

18. Raisamo, R. and Räihä, K. 1996. A new direct manipulation technique for aligning objects in drawing programs. *Proc. UIST'96*, 157-164.

19. Schiffman, H.R. 2001. Fundamental Visual Functions and Phenomena. *Sensation and Perception*. Wiley, 89-115.

20. Shen, C., Vernier, F., Forlines, C., and Ringel, M. 2004. Diamond-Spin: an extensible toolkit for around-the-table interaction. *Proc. CHI*'04, 167-174.

21. van Rhijn, A. and Mulder, J. D. 2006. Spatial input device structure and bimanual object manipulation in virtual environments. *Proc. VRST'06.*, 51-60.

22. Vogel, D. and Baudisch, P. 2007. Shift: a technique for operating pen-based interfaces using touch. *Proc. CHI'07*, 657-666.

23. Wang, Y., MacKenzie, C. L., Summers, V. A., and Booth, K. S. 1998. The structure of object transportation and orientation in human-computer interaction. *Proc. CHI'09*, 312-319.

24. Ware, C. Using hand position for virtual object placement. *Vis. Comput.* 6, 5 (Nov. 1990), 245-253.

25. Williams, R. 2008 *The Non-Designer's Design Book, Third Edition*. Peachpit Press.

26. Zhai, S., Milgram, P., 1998. Quantifying coordination in multiple DOF movement and its application to evaluating 6 DOF input devices. *Proc. CHI'98*, 320-327.