Hand Occlusion on a Multi-Touch Tabletop

Daniel Vogel, Géry Casiez

HAL Id: hal-00670516
https://inria.hal.science/hal-00670516

Hand Occlusion on a Multi-Touch Tabletop

Daniel Vogel
Cheriton School of Computer Science
University of Waterloo, CANADA
dvogel@uwaterloo.ca

Géry Casiez
LIFL & INRIA Lille
University of Lille, FRANCE
gery.casiez@lifl.fr

ABSTRACT

We examine the shape of hand and forearm occlusion on a multi-touch table for different touch contact types and tasks. Individuals have characteristic occlusion shapes, but with commonalities across tasks, postures, and handedness. Based on this, we create templates for designers to justify occlusion-related decisions and we propose geometric models capturing the shape of occlusion. A model using diffused illumination captures performed well when augmented with a forearm rectangle, as did a modified circle and rectangle model with ellipse "fingers" suitable when only X-Y contact positions are available. Finally, we describe the corpus of detailed multi-touch input data we generated which is available to the community.

Author Keywords

occlusion, multi-touch, hand, tabletop, tablet, finger

ACM Classification

H5.2. Information interfaces and presentation: User Interfaces - Input devices and strategies.

INTRODUCTION

Operating a computer by directly touching the display surface has many benefits, and in tabletop computing, multi-touch is arguably the most natural form of input. However, with any form of *direct* input, where the input device and the output display are coincident, the hand and arm cover — or *occlude* — part of the display. This can be a problem, because compared to manipulating objects on a real tabletop, a tabletop computer is dynamic and can display relevant information, sequential widgets, and system messages in occluded areas. Researchers are aware of occlusion: they suggest it impedes performance [9,21,22] and use it to motivate the design of interaction techniques [7,13,15,20,24]. Yet, there has not been a systematic study of hand occlusion with multi-touch tabletops.

Vogel et al. [25] developed a methodology to study direct pen occlusion by capturing the actual shape of occlusion from a person's point-of-view. We adapt their methods for video capture, augmented reality marker tracking, and image processing to a multi-touch tabletop (Figure 1). Compared to pen input, examining multi-touch occlusion is

more challenging due to the wide vocabulary of touch contact types using different combinations of fingers and postures across different hands. We studied 18 conditions covering typical combinations of 9 different contact types with 3 tasks: tapping, dragging, and transforming. By examining the mean occlusion shapes, we find that individuals use consistent hand postures, and although there are differences between individuals, there is enough commonality to use overall mean shapes to inform interface design. For this purpose, we create calibrated occlusion template shapes to guide designers with interface layouts which reduce occlusion.

We conducted our experiment on a diffused illumination (DI) table top. The raw infrared (IR) image blob near the contact points should be a good estimate of the occluded area, and we propose a geometric model of occlusion combining the IR blob with a "forearm" rectangle. However, input technologies like capacitance only sense X-Y contact positions, so we also created an alternative model. We add ellipses for extended fingers to Vogel et al.'s [25] pen occlusion model without increasing the number of parameters. In a test fit of these models to occlusion silhouettes, the DI model achieves a F1 score of 0.80, while the multi-touch geometric model slightly outperforms it at 0.82. This suggests that real-time prediction of the occluded area, even with only X-Y contact positions, is possible — enabling occlusion-aware interaction techniques [23] on any multi-touch table regardless of hardware technology.

In the course of this project, we generated a large corpus of images synchronized with logged multi-touch data for common tasks. We feel that this is also a contribution, and make it available for related research.
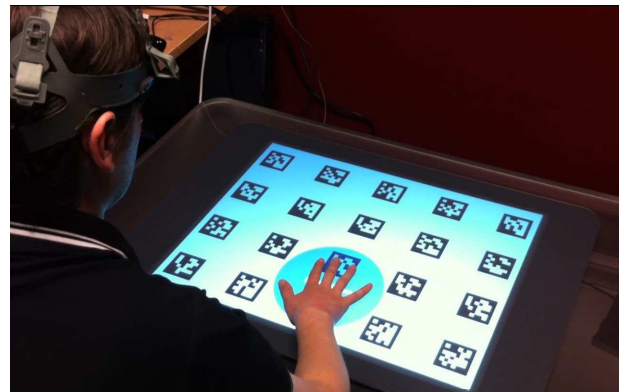
Figure 1. Experiment set-up: the occluded area is captured by a head-mounted camera and rectified using fiducial tracking.

RELATED WORK

Vogel and Balakrishnan [22] list hand occlusion as one of five direct pen input problems. They observed hidden status messages, missed previews, inefficient movements, and occlusion contortion. Pen input work by Hancock and Booth reached similar conclusions [9], but in practice it is difficult to strictly control for occlusion without resorting to different input paradigms like indirect pointing. With touch screens, finger occlusion has long been known to be problematic [21]. Recent work has argued that finger occlusion is not directly responsible for errors [12], but it is undeniable that one cannot see what is beneath their finger. With larger multi-touch tables and tablets, this is compounded as multiple fingers, hands, and forearms cover more of the display. So it makes sense that reducing occlusion is an important aspect of direct input interface design. Vogel et al. [25] provide a summary of occlusion and direct pen input; we focus on direct touch input here.

Effect of Occlusion

Several touch techniques address occlusion directly. For example, expanding feedback beyond occluding fingers [20], shifting a copy of the hidden area out from under the finger [24], and creating methods to manipulate objects remotely to avoid occluding them [27]. More radical solutions like moving touch to the back [28] or underside [29] also work, but they reduce the directness of touch.

Other work uses occlusion as a motivation for the spatial layout of interface designs. For example, FingerGlass [13] and ShadowGuides [7] use spatial offsets to reduce the chance of occlusion, but it is unclear how the exact offset was determined, or if it is optimal. Eden [15] describes multiple design decisions to reduce occlusion and argues for occlusion-awareness in multi-touch applications.

Understanding Occlusion

When making such occlusion-motivated design decisions, there is an implied reference to the shape of a "typically occluded" area. However, this is typically determined in an ad hoc manner. With direct pen input, empirically based occlusion-aware layout decisions have been used, such as Hancock and Booth's [9] context menu placement by radial selection time and Brandl et al.'s [4] paper-based experiment for pie menu orientation.

Although encouraging, in neither of these cases is the shape of occlusion analyzed directly as in Vogel et al. [25]. In a controlled experiment, they capture images of the occluded area using a head-mounted camera. These *occlusion silhouette* images are used to visualize mean shapes and develop a simple geometric model. This work led directly to Vogel & Balakrishnan's [23] design for an individually configurable, real-time occlusion model to realize occlusion-aware interfaces and interaction techniques.

Our methodology is based closely on this work, but we introduce new experiment tasks and refined geometric models tailored for multi-touch input. We also contribute other methodology refinements such as placing fiducial tracking markers in the display and introducing more meaningful descriptive statistics for occlusion shapes.

Understanding Multi-touch Postures

It is impossible to study multi-touch occlusion without considering the shape of the hand. Past work has looked at what postures people use for various types of multi-touch interactions in controlled experiments [14], elicitation studies [17,30], and in the field [11]. This has provided insights such as: people use different fingers for the same contact type [14]; people use any number of fingers for operations like dragging [17,30]; and people use different open- and closed-hand postures for single finger contacts [11]. In our experimental tasks and posture conditions, we balance these "almost anything goes" results with the reality of research and commercial systems which already map the type of contact [27] or number of finger contacts to specific tasks (e.g. selecting vs. scrolling vs. paging [2]).

Rather than looking at what postures are used, we examine the literal posture shape from a person's point-of-view. While our focus is different, our methodology and the corpus of data we generated can be applied to investigations of other posture characteristics like the studies above.

EXPERIMENT

The goal of our experiment is to study the shape of the occluded area for canonical multi-touch table postures and interaction tasks. We focus on a multi-touch table because smaller tablet form factors use fewer gestures due to their size and capacitive sensing makes some postures impractical. Limitations are discussed in our conclusion.

We adapt the methodology introduced by Vogel et al. [25]. As participants perform common multi-touch gestures, we record a video of their hands using a head-mounted camera. Then we extract key frames and isolate occlusion silhouettes to create a rectified view of the hands from their vantage point. We are not interested in performance time.

Participants

24 people (16 male, 8 female) with a mean age of 30 (SD 6.6) participated. 3 participants were left-handed (2 male, 1 female). 7 participants said they had experience with a multi-touch table and 21 said they had used a multi-touch phone or tablet. We recorded the height of all participants and found a mean of 176.3 cm (SD 9.4). This measurement is to suggest a reasonable sampling of person size, not to search for correlations between anatomical size and occlusion shape since Vogel et al. found this unlikely.

Apparatus

We conducted the experiment on a Microsoft Surface 1.0 multi-touch table. It has a 610×458 mm display with a display resolution of 1024×768 px (1.679 px per mm). The table-like case is 560 mm high with a 690×1080 mm top, approximating a small coffee table. Participants sat in a fixed chair centred along one of the long sides and we asked them to refrain from leaning from side-to-side. We did not observe anyone having difficulty reaching distant target locations. A fixed body position is a necessary

experimental control, but our relative results could be transformed to a tracked body location [1].

The small head-mounted video camera for the participant's point-of-view recorded the experiment at 960×720 px resolution and 15 FPS. It was positioned as closely as possible to the centre of the eyes, without interfering with the participants' line of sight. Since the camera is mounted above the centre of the eyes, it does not capture the *exact* point-of-view. Vogel et al. estimated that rectified occlusion silhouettes would be shifted down by 6 mm on average. Since we have a greater distance from the eye to the hand, our error would be reduced further.

To enable us to track and rectify the Surface display in the camera image, we displayed a 4×5 grid of 59 mm fiducial markers. We could not put the markers on the bezel due to the size of the display and the field-of-view of the camera. We were initially hesitant to show this pattern under our experimental stimulus, but found that participants effectively ignored it within the first few practice trials.

The experiment code is in C# using the Surface SDK. During task activity, the position, ellipse size, and ellipse orientation of all touch contacts were logged at more than 60 Hz along with 15 FPS of 768×576 px (1.26 px/mm) raw IR captures. In addition to the head-mounted camera, we recorded everything with a stationary camera above the Surface, but did not use this in the present analysis.

Gesture Set
Compared to Vogel et al.'s study where the pen is the only type of contact, we needed to select a representative set of multi-touch gestures: contact postures paired with interaction movements. We consulted multi-touch device SDKs and user guides (nicely summarized in [31]) and related research [3,5,7,11,16,17,30]. To keep the study reasonable and more ecologically valuable, we selected only gestures which use a single posture (e.g. no opening palms [3]) and simple movements along single paths (e.g. no L-shapes or X-shapes [17]). We focus on single-handed gestures since many two-handed gestures may be factored into two separate gestures for the purpose of shape analysis.

We identified three main types of interaction movements: tapping, dragging, and object transformation. A fourth choice would have been "flicks," but these resemble a short, high speed drag. To avoid redundancy between dragging and transforming, we restrict transformations to simultaneous rotation and scale only.

We identified eight common types of postures: using 1, 2, 3, 4, or 5 digits ('digit' includes 'fingers' and 'thumb'); a flat palm, the side of the fist, and the side of the hand. All 8 of these postures are paired with tapping, but only the first 6 with dragging since dragging with the fist or the side of the hand are less common. We paired transforming with 2 digit and 5 digit poses only since a 2 digit posture captures the common pinch gesture and the 5 digit posture is also common for transformations [11]. Like gestures, many two-handed postures can be factored into two one-handed postures (e.g. a non-dominant flat palm setting the mode for a 1 digit dominant hand drag [27]). We only include a two-handed transform gesture since both hands work together.

	1 Digit	2 Digits	3 Digits	4 Digits	5 Digits	Palm	Fist	Side	Two Handed
Tap	25	80	100	120	160	200	100	180	
Drag	25	80	100	120	160	200			
Transform		120			160				140

Table 1. *Postures* and *Tasks* with target diameters (in mm).

Task and Stimuli
The gesture movements define three tasks: *Tap*, *Drag*, and *Transform*. Each task has a main circular target with diameter set according to posture (see Table 1). The smallest diameter is three times the minimum recommended touch target size [12], the largest based on anthropomorphic palm size [18], and intermediate sizes selected to easily accommodate postures. This balances ease-of-selection with location constraints.

In the same spirit of Kin et al.'s experiment [14], the inherent ambiguity of a single circular target, together with the generic term 'digit,' allows participants to use different posture strategies. For example, a 2 digit posture could be a thumb and index finger, or an index and middle finger. Overly suggestive targets or terms like those for teaching specific gestures [3,7] would prevent natural posture strategies, leading to different shapes.

For most postures, our code prevents interaction unless the correct number of contacts are on the target. This worked well for 1 to 5 digits, but had to be relaxed for palm, fist, and side since the Surface detects an irregular number of contacts in these cases. In the spirit of allowing participants to adopt posture strategies, we do not control for the number of contacts with the two-handed condition. We wanted to see if people used two index fingers or some other combination of fingers across hands.

Tap Task. To complete the *Tap* task, the participant touches a circular target using the required posture for 333 ms. This short delay reduces motion blur and increases the tolerance for event log synchronization for the point-of-view frame captures, addressing problems reported by Vogel et al. [25] The centre of the current target is positioned at one of 9 locations spaced evenly on a 3×3 grid. Vogel et al. use a more granular 7×11 grid, but their findings do not suggest this is necessary. We cover a range of extreme positions by centering the grid in the display and spacing columns and rows at 192 mm and 112 mm respectively.

Drag Task. To complete the *Drag* task, the participant uses the required posture to drag a circular target from the centre of the display to one of 8 circular dock locations on the same 3×3 grid. The outer and inner diameters of the ring-shaped dock are set so that the error tolerance for the target

is 30 mm. Like the *Tap* task, the target must be held within the dock using the correct posture for 333 ms.

Transform Task. To complete the *Transform* task, the participant rotates and scales a circular target until a pin aligns with a 30 mm rotation tolerance "key" and the target border fits within a 15 mm outer ring tolerance. This position must also be held for 333 ms to complete the task. All *Transform* tasks are at the center grid position, but with 4 rotation and scale conditions: clock-wise (CW) and counter-clockwise (CCW) 60° rotation and scaling up or down by 45 mm. The initial target angle is -60° for CW and 0° for CCW to minimize key occlusion.
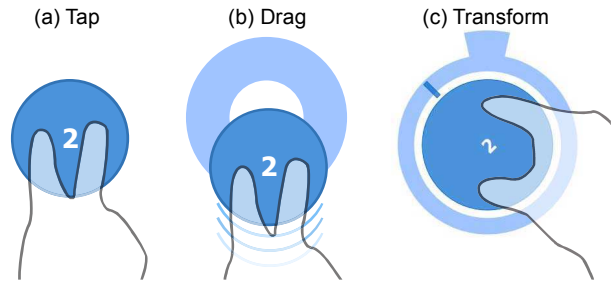

Figure 2. Experiment tasks (using 2 Digits as an example): (a) Tap on circular target; (b) Drag target into ring-shaped dock; (c) Transform target rotate pin to key and scale to fit in ring.

Design

The main experiment had 3 *Blocks*, with each block consisting of 3 *Task Sections*. Each *Task Section* contained all permutations of *Posture* and *Task Condition* for a *Task*, grouped by posture: the *Tap Task* had 8 posture groups, each at 9 grid position *Conditions*; the *Drag* task had 6 *Posture Groups*, each at 8 grid position *Conditions*; and the Transform task had 3 *Posture Groups*, each with 4 rotation and scale *Conditions*. The *Postures* were presented in approximate order of increasing difficulty (the column ordering in Table 1). Within each *Posture Group*, the order of *Conditions* were randomized. All blocks had the same *Task* ordering, but this order was counter-balanced across participants. In summary:

        3 Blocks ×

        8 *Postures* × 9 *Conditions* (*Tap* task section)

    + 6 *Postures* × 8 *Conditions* (*Drag* task section)

    + 3 *Postures* × 4 *Conditions* (*Transform* task section)

    = 540 data points per participant

Before beginning the main experiment blocks, participants completed 26 practice trials: 1 centre *Tap* trial for each *Posture*; 1 *Drag* trial for each *Posture* to a random outer grid position; and all permutations of *Transform* trials. After the main experiment, participants also completed 72 trials with their non-dominant hand covering a subset of *Conditions* for all *Tasks* and *Postures*: 5 *Tap* trials for each *Posture* at all grid positions except corners; 4 *Drag* trials for each *Posture* to all grid positions except corners; and 8 *Transform* trials covering all *Conditions* except the two-handed posture. We elected to use this subset of conditions

to reduce experiment time and fatigue, but provide enough data to show any pronounced differences between hands. The total experiment took less than 1 hour to complete.

IMAGE PROCESSING

To transform the point-of-view video into occlusion silhouettes, we use the same steps as Vogel et al.:

Frame Extraction. After synchronizing the video and the data log using visual time markers, we capture one video frame at the end of all tasks and one frame when the participant first touched the target in *Transform* and *Drag*.

Rectification. We wrote custom software using the ARToolkitPlus augmented reality library [26] to track the fiducial markers. After tracking the image-space marker positions, we use OpenCV to calculate the homography matrix and rectify the image of the hand against the display. The rectified display area is 610×458 px, so 1 px = 1 mm. Our software application enables us to manually track markers when automatic tracking fails (i.e., when less than 4 non-collinear markers are detected due to motion blur, cropping, or hand occlusion). About 5% of the frames had to be partially tracked manually.

Isolation. To isolate the binary occlusion silhouette images for analysis, we use similar image processing steps as Vogel et al. Since the fiducial markers are in the display space, we add a median background subtraction step to remove them. This works reasonably well, but we realized that colouring the fiducial markers saturated blue instead of black would have greatly simplified this task.

RESULTS

Compared to Vogel et al., we had no corrupted participants and very few corrupted images (less than 0.001%). This is largely due to consistent camera settings and room lighting.

Shape of Occluded Area

Since our goal is to study the shape of the occluded area, we begin with an examination of the overall mean occlusion silhouette shapes shown in Figure 8 (second last page). These are created by registering all silhouettes by target grid position (the actual centroid is not robust for palm, side, and fist postures), then finding average pixel values across all participants and conditions for each *Task* and *Posture*. Using the usual experimental assumption that our 24 participants provide a reasonable population sample, the darker areas are more likely to be occluded.

For *Tap* and *Drag*, the darker areas show most *Postures* clearly, suggesting homogeneity across participants and grid positions, but heterogeneity between *Postures*. Differences between *Fist* and *Side* are subtle, but consistent. Medium grey areas suggest different *Posture* strategies. For example, the ghost-like shape of other fingers for *1-Digit* and *2-Digits* suggest a mixture of open and closed hand postures. There is also surprising similarity between *Tap* and *Drag* tasks. For most postures, the differences are unperceivable; the largest differences are with the *Palm*.

For *Transform*, the darker areas are less defined. This is partly due to high positional variance in the different conditions, but also indicates a greater variety of *Posture* strategies. A general thumb and index finger pinch shape can be seen for *2-Digits*, perhaps because most participants reported experience with multi-touch devices where this strategy is standard. For *Two Hands*, the hands can be discerned, but it is unclear what digits are used. The posture for *5 Digits* is particularly heterogeneous. While *2-Digit Transform* has similarities to *2-Digit Tap* and *Drag*, *5-Digit* and *Two Hand* similarities to *Tap* and *Drag* are less so.

Individual Participants
We also examined all mean silhouettes for *Participants* by *Task* and *Posture*. Due to space, our discussion here focuses on *Tap* and *Transform* for 8 participants chosen for diversity, which are reproduced in Figure 9 (last page).

Individual differences in posture strategy, hand shape, and hand size are apparent, but there are common shape characteristics overall. This is most clear in *Tap* where the palm resembles a circle with one or more fingers extended, and the forearm typically projects down to the lower-right. This is more difficult to see with *Transform* where start and end positions are merged together.

Open- and closed-hand strategies are clearly seen for *1-Digit*, *2-Digits* and *3-Digits*. For example, participants 8, 3, 21, and 24 extend their touch fingers from a closed fist, whereas others extend all fingers regardless. Although we observed some variation in digit used [14], each participant generally used the same digit(s) for a given *Posture*. For Palm posture, most participants spread their fingers (participant 24 is an exception) and differences between the fist and side are subtle. With *Two Handed* postures, the non-dominant hand is often positioned lower (e.g. participants 22, 3) and individual hands are similar to *1-Digit Tap* silhouettes.

Across postures and tasks, forearm angle appears consistent. Wrist angle appears consistent in *Tap* (e.g. participant 23 has the most acute angle for *3-Digits)*. With *2-* and *5-Digit Transform*, there is more wrist variability.

Left-Hand and Non-Dominant Hand
We found no pronounced differences for left-handed participants (e.g. participants 22 and 24). We also found no pronounced differences between dominant and non-dominant hand, but the smaller set of data limits this result.

Descriptive Statistics
To help quantify these observations, we devised three statistics (Figure 3): the size of the occluded *area* near the target; the *breadth* of the fingers in the hand posture; and the *angle* of the forearm. Our motivation is to reinforce mean silhouette observations not quantitative tests.

Occluded Area
The mean areas for *Tap* and *Drag* across common *Postures* are very similar at 107 cm^2 (SD 17) and 105 cm^2 (SD 19), supporting our mean shape observations. The mean area for

Transform is somewhat lower at 98 cm^2 (SD 23). Mean areas by *Posture* for *Tap* and *Drag* also support visual observations. The areas of *1-Digit Tap* and *Drag* are both 82 cm^2 (SD 15). The trend continues with *Palm's* large task discrepancy and area, 152 cm^2 (SD 21) for *Tap* and 141 cm^2 (SD 28) for *Drag*. With *Transform*, we observed visual differences for common *Postures* with *Tap* and *Drag*, but quantitatively the difference in area is within 10 cm^2 with measurements of 89 cm^2 (SD 17) for *2-Digits* and 106 cm^2 (SD 18) for *5-Digits*. The mean area for *Two Hands* is 98 cm^2 (SD 34) which is closer to *3-Digit Tap* and *Drag*.

Breadth of Posture
Mean breadth for *Tap* and *Drag* across common *Postures* is similar at 170° (SD 48) and 181° (SD 39), but with higher variance. Differences between the same *Posture* for *Tap* and *Drag* are less than 4°, except *Palm* at 59° which supports our observations. Overall, the spread monotonically increases from 119° to 225° for *1-Digit* to *5-Digits*. For *Transform*, *2 Digit* spread is greater than *2-Digit Tap* and *Drag* reflecting the variability in the pinch gesture.

Forearm Angle
Arm angle is remarkably consistent for common *Postures* between *Tap* and *Drag*. The overall means are 58° (SD 16) and 57° (SD 16) respectively, with individual variations less than 3° (centre grid positions removed since all *Drag* tasks begin at this position). We can compare *Transform* and *Tap* at the centre location only. Here the values diverge, with means of 45° (SD 21) and 58° (SD 9) respectively, likely due to the variability of the *Transform Task*. As the hand reaches target locations, the forearm angle changes according to kinematics. While arm angle varies according to contact location, these angles are remarkably consistent between *Drag* and *Tap* and across participants.
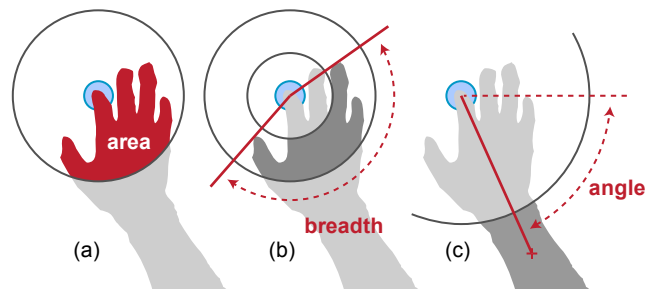

Figure 3. Descriptive statistics for hand occlusion: (a) occluded *area* in cm^2 within a 100 mm radius of the target center; (b) angular hand posture *breadth* within a 50 to 100 mm ring; (c) arm *angle* from horizontal, calculated from the target center to the centre of mass beyond a 150 mm circle. Segmentation radii were selected by trial and error: for *breadth* and *angle*, they crudely segment silhouettes into hand and forearm slices; for *area*, this gives a measure of localized occlusion (an admittedly arbitrary size, but more generalisable than Vogel et al.'s occlusion area measure as the ratio of a 12" display).

Discussion
Our results yield these main findings for occlusion shape:

- Postures have distinct shapes, but with common elements like a hand blob and protruding forearm.

- Tapping and dragging have very similar shapes, but transforming shapes are different.

- Different people may use different posture strategies, but each individual adopts a consistent strategy making posture shapes within an individual similar.

- Left- and right-handed people, and dominant and non-dominant hands, have no pronounced differences.

- Forearm angle is consistent between tapping and dragging, transform is more divergent.

These results are of theoretical interest to researchers and provide causal evidence for the potential impact of occlusion during interaction. But, how can it be used? Next, we show how this occlusion shape information is made accessible to tabletop interface designers, and describe a simple model to capture the essence of the occluded area setting the stage for multi-touch occlusion-aware interfaces.

DESIGN IMPLICATIONS AND RESOURCES
Without empirical guidelines, designers rely on rules-of-thumb or ad hoc observation to reduce occlusion in interface layouts. This is difficult to justify and may be sub-optimal as non-occluded or occluded locations are missed.

Occlusion-awareness Templates for Designers
Using the overall mean silhouettes, we created design-time "occlusion-awareness" templates for designers (Figure 4). Each template is a dimension-calibrated image showing areas which may be occluded relative to the expected contact centroid. Two bands of occlusion severity are illustrated, calculated from pixel density thresholds in the associated mean image: the possibly occluded area (>10%) and the often occluded area (>50%). These templates are available for download[1] as a layer-separated PDF and can be imported into common design applications like Adobe Photoshop or Illustrator. After scaling the template to match the real world units of an interface design, the designer can use it as an overlay to make occlusion-aware layout decisions. We created templates for all postures across t and *Transform* task (since drag is very similar to tap).
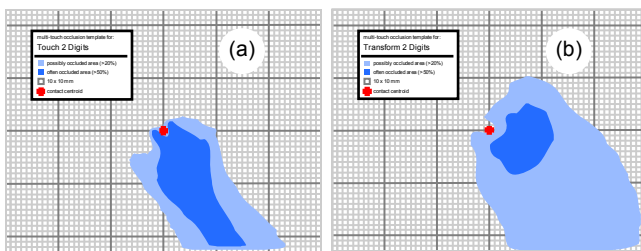

Figure 4. Occlusion-awareness template examples for: (a) 2 Digit Tap; (b) 2 Digit Transform. Light blue is the possibly occluded area (>10%), dark blue is the often occluded area (>50%). The red cross is the expected contact centroid.

All templates are generated as single right-handed shapes, but can be easily flipped for left-handed designs. Although people prefer using their dominant hand for single-handed multi-touch interactions [14], this is not guaranteed. We did not see pronounced differences between dominant and non-

dominant silhouettes in the subset of tested conditions, so designers should be able to mirror and union the right-handed template when designing for either hand.

Example Applications
Figure 5 illustrates two examples of how designers can use these templates to guide occlusion related decisions:

- When specifying the position of an information bubble opened with a two finger tap, a designer can use the 2 digit tap template to position the bubble at an offset and angle least likely to be occluded, but also minimizing distance for ideal Gestalt association (Figure 5a).

- A design for a small multi-touch rotary dial can use the 2 digit transform template to position labels on the circumference of the dial to minimize occlusion (Figure 5b). While labels should avoid the "often occluded area," the designer can utilize the "possibly occluded" band to place lower priority labels. The template shape enables further refinements, such placing highest priority labels at '1', since they are least likely to be occluded.
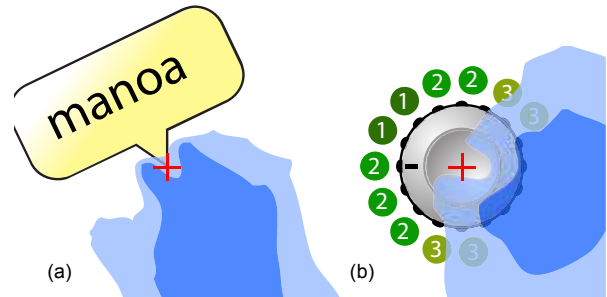

Figure 5. Example applications of design-time templates: (a) the best place to put an information bubble activated with 2 digits; (b) the best option indent locations for a rotary dial operated with 2 digits: 1 is best, 2 are good, 3 are acceptable.

Other scenarios include contextual menu placement [9], pie menus [4], and visualizations [27]. The designers of FingerGlass [13] and ShadowGuides [7] could use these templates to justify their choice for spatial offsets and there are multiple opportunities to refine layouts to minimize occlusion in a complex application like Eden [15].

Although helpful, these are design-time decisions using mean shapes. Knowledge of the currently occluded area at any given moment would be even better.

GEOMETRIC MODEL OF OCCLUSION SHAPE
In this section, we develop and test different ways to model the occluded area suitable for high fidelity technologies like DI and more limited hardware like capacitance. Our aim is to show that applying a single model to a wide range of postures is possible, and establish an upper bound on potential performance. This sets the stage for a configurable real time model of occlusion as future work.

Diffused Illumination (DI) Capture Model
Unlike a pen tablet, a DI tabletop captures an image of the hand and forearm near the surface. Without a switchable diffuser [10] or calibrated overhead camera, this is the best

available method for capturing actual hand shape. With minor rotation and offset transformations relative to the contact centroid, this should match a portion of the occluded area and forms a key part of our first geometric model. The problem is that the whole forearm is not usually captured due to its height above the diffuser, so we add a rectangle with a constant offset of 100 mm from the same centroid (Figure 6a). This DI model has five parameters: a distance and angle to describe the offset of the DI image, an angle for rotation of the DI image, and a rotation angle and width for the rectangle.
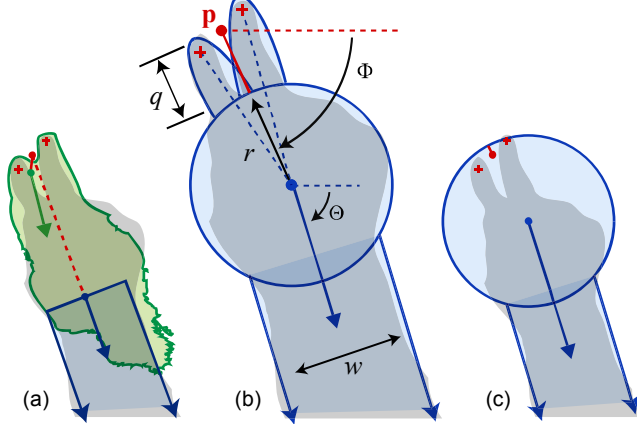

Figure 6. Three occlusion shape models: (a) DI and rectangle; (b) multi-touch circle and rectangle; (c) Vogel et al.

Multi-Touch Circle and Rectangle Model

Typically, FTIR only provides the shape and size of the contacts [8], and capacitive only X-Y contact positions. To cover a wider range of devices, we extended Vogel at al.'s circle and rectangle model (Figure 6c) to multi-touch postures. Our model has exactly the same five parameters (Figure 6b): distance q and angle Φ describe the offset of circle from point **p**, the centroid of the *actual* finger contact points; r is the circle radius; Θ is angle of the rectangle; and w is the rectangle width. To represent extended fingers, we add an ellipse for each contact and position it relative to the circle. Specifically, the major axis is aligned with the vector from the contact point to the circle centre and its length is such that the minor axis forms a chord on the circle. We set the minor axis to 15 mm and scale the major axis such that the tip extends 10 mm beyond the contact. Since these are constants, no additional model parameters are introduced for the ellipses. When faced with modeling extended finger postures, the ellipses should increase accuracy compared to only Vogel et al.'s circle and rectangle (Figure 6c). A limitation is that we only have actual X-Y contacts, so a single finger contact with an open hand (like participant 8 in Figure 9) would only have a single contact. In this case, the model can increase r and decrease q to remove the ellipse and cover the entire hand with the circle.

Testing Models by Fitting to Captured Silhouettes

To test and compare the models, we use the same approach as Vogel et al. [25]: we "fit" the model to each silhouette as accurately as possible and use precision-recall plots and F1

scores to compare fidelity. Note that we are not *learning* model parameters, but rather estimating an upper bound for model capability. The reader can consult the prior art for methodology details noting these changes: we use a single fitting stage; we only used pattern search; we use the posture contact centroid; and our objective function simply maximizes the F2 score. F2 favours recall over precision so more of the occluded area is covered creating more false positives, but fewer false negatives — a quality Vogel et al. argue is desirable. To remain consistent with past work, we compare fidelity with the equally weighted F1 score. Since the models are one-handed, we remove two-handed transform cases. Fitting each model to the 9209 test cases took more than 12 hours with a 2.66 GHz quad processor.

Mean F1 scores for the DI model and our multi-touch circle and rectangle model are comparable: 0.801 (SD 0.078) and 0.808 (SD 0.064) respectively. We also tested the "fingerless" Vogel et al. model, which has a very respectable F1 of 0.785 (SD 0.066). Since our model is based closely on it, a similar score is expected, but it is encouraging to see the finger ellipses improve fidelity without additional parameters.

Our geometric model is primarily useful for non-DI devices where only individual finger contacts are sensed, not postures like palm, fist, and side. Thus, comparing mean F1 scores using only 1 to 5 digit contacts is more relevant. In this test, the DI model achieves a similar F1 score of 0.802 (SD 0.074) and a precision-recall plot illustrates a precision bias (Figure 7a). Our multi-touch circle and rectangle model improves with 0.819 (SD 0.055) and the plot suggests very high recall and good precision (Figure 7b). The Vogel et al. model also improves to 0.797 (SD 0.057).
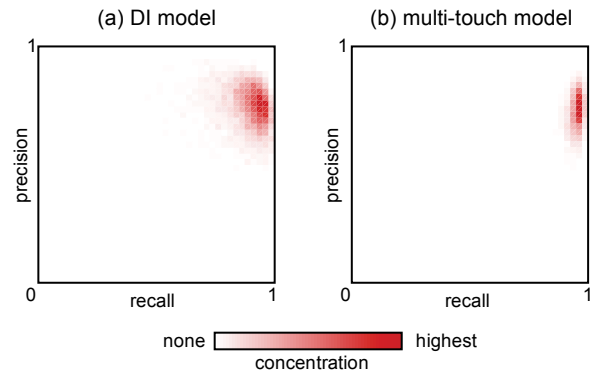

Figure 7. Precision-recall concentration plots: (a) DI shadow; (b) multi-touch circle and rectangle. Points in upper right indicate better performance.

IMAGE CORPUS

Generating the occlusion design-time templates and testing different occlusion models leverages the large corpus of images and metadata we created in our experiment. This includes 16,320 sets of images synchronized with contact positions, sizes, and orientations. Each image set has an occlusion silhouette, raw and rectified versions of a point-of-view frame capture, a DI capture, and an overhead frame

capture. We believe this corpus will be a valuable resource for related research such as palm rejection, finger to hand mapping [6], and identifying users by hand contour [19]. The entire corpus is publically available to download[1].

CONCLUSIONS AND FUTURE WORK

We adapted an established image-based methodology [25] to study the shape of occlusion on a multi-touch tabletop. By examining the shapes of mean occlusion silhouette visualizations and calculating descriptive statistics, we found common characteristics across people, postures, and tasks. Based on this, we created occlusion awareness templates to guide interface layout decisions, and tested different geometric models suitable for high- and low-fidelity multi-touch input technologies. This latter contribution is a necessary step towards a real-time, configurable model to enable multi-touch occlusion-aware techniques such as those created for direct pen input [23].

Using a tabletop allowed us to test a wider assortment of postures and test DI captures as a potential model, but we had to accept potential limitations. Although the Microsoft Surface is popular and has spawned other similarly-sized tables, larger and taller tables could influence body posture and resulting occlusion. More broadly, an obvious question is how well our results generalize to other multi-touch phones, tablets, tables, inclined desks, and vertical walls. We argue that at least for near horizontal cases, the relative relationship and viewing angle of operator to device is similar. The biggest shape change was due to forearm angle when reaching, which does not apply to small devices. However, body postures may contort more drastically when reaching targets at extreme edges of a large surface. A second limiting factor is that our results are relative to body location. For larger surfaces, we assume that the location of people around the table can be determined using sensors [1], or perhaps using the angle or shape of the hand in the spirit of Dang et al. [6].

Finally, we are eager to see how the community might use the large corpus of data we created.

REFERENCES

1. Annett, M., Grossman, T., Wigdor, D., and Fitzmaurice, G. Medusa: a proximity-aware multi-touch tabletop. *Proc. UIST'11*, ACM (2011), 337–346.

2. Apple Computer. iPad 2 User Guide. 2011.

3. Bragdon, A., Uguray, A., Wigdor, D., Anagnostopoulos, S., Zeleznik, R., and Feman, R. Gesture play: motivating online gesture learning with fun, positive reinforcement and physical metaphors. *Proc. ITS'10*, ACM (2010), 39–48.

4. Brandl, P., Leitner, J., Seifried, T., Haller, M., Doray, B., and To, P. Occlusion-aware menu design for digital tabletops. *Proc. CHI'09,* ACM (2009), 3223-3228.

5. Cao, X., Wilson, A., Balakrishnan, R., Hinckley, K., and Hudson, S.E. ShapeTouch: Leveraging Contact Shape on Interactive Surfaces. Proc. *Horizontal Interactive Human Computer Systems*, IEEE (2008).

6. Dang, C.T., Straub, M., and André, E. Hand distinction for multi-touch tabletop interaction. *Proc. ITS '09*, ACM (2009), 101-108.

7. Freeman, D., Benko, H., Morris, M.R., and Wigdor, D. ShadowGuides: visualizations for in-situ learning of multi-touch and whole-hand gestures. *Proc. ITS'09*, ACM (2009), 165–172.

8. Han, J.Y. Low-cost multi-touch sensing through frustrated total internal reflection. *Proc.UIST'05,* ACM (2005), 115-118.

9. Hancock, M.S. and Booth, K.S. Improving menu placement strategies for pen input. *Proc. GI'04*, CIPS (2004), 221-230.

10. Hilliges, O., Izadi, S., Wilson, A.D., Hodges, S., Garcia-Mendoza, A., and Butz, A. Interactions in the air: adding further depth to interactive tabletops. *Proc. UIST'09*, ACM (2009), 139-148.

11. Hinrichs, U. and Carpendale, S. Gestures in the wild: studying multi-touch gesture sequences on interactive tabletop exhibits. *Proc. CHI'11*, ACM (2011), 3023–3032.

12. Holz, C. and Baudisch, P. Understanding touch. *Proc. CHI'11*, ACM (2011), 2501-2510.

13. Käser, D.P., Agrawala, M., and Pauly, M. FingerGlass: efficient multiscale interaction on multitouch screens. *Proc. CHI'11*, ACM (2011), 1601-1610.

14. Kin, K., Agrawala, M., and DeRose, T. Determining the benefits of direct-touch, bimanual, and multifinger input on a multitouch workstation. *Proc. GI'09*, CIPS (2009), 119-124.

15. Kin, K., Miller, T., Bollensdorff, B., DeRose, T., Hartmann, B., and Agrawala, M. Eden: a professional multitouch tool for constructing virtual organic environments. *Proc. CHI'11*, ACM (2011), 1343-1352.

16. Micire, M., Desai, M., Courtemanche, A., Tsui, K.M., and Yanco, H.A. Analysis of natural gestures for controlling robot teams on multi-touch tabletop surfaces. *Proc. ITS'09*, ACM (2009), 41–48.

17. Morris, M.R., Wobbrock, J.O., and Wilson, A.D. Understanding users' preferences for surface gestures. *Proc. GI'10*, CIPS (2010), 261-268.

18. Pheasant, S. and Hastlegrave, C. *Bodyspace: Anthropometry, Ergonomics and the Design of the Work*. CRC, 2006.

19. Schmidt, D., Chong, M.K., and Gellersen, H. HandsDown. *Proc. NordCHI'10*, ACM (2010), 432-441.

20. Shen, C., Ryall, K., Forlines, C., et al. Informing the Design of Direct-Touch Tabletops. *IEEE Comput. Graph. Appl. 26*, 5 (2006), 36-46.

21. Shneiderman, B. Touch Screens Now Offer Compelling Uses. *IEEE Software 8*, 2 (1991), 93-94, 107.

22. Vogel, D. and Balakrishnan, R. Direct Pen Interaction with a Conventional Graphical User Interface. *Human-Computer Interaction 25*, 4 (2010), 324-388.

23. Vogel, D. and Balakrishnan, R. Occlusion-aware interfaces. *Proc. CHI'10*, ACM (2010), 263-272.

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

---

[1] http://www.nonsequitoria.com/mt_occlusion

25. Vogel, D., Cudmore, M., Casiez, G., Balakrishnan, R., and Keliher, L. Hand Occlusion with Tablet-sized Direct Pen Input. *Proc. CHI'09*, ACM (2009), 557-566.

26. Wagner, D. and Schmalstieg, D. ARToolKitPlus for Pose Tracking on Mobile Devices. *Proc. CVWW'07*, (2007).

27. Wigdor, D., Benko, H., Pella, J., Lombardo, J., and Williams, S. Rock & rails: extending multi-touch interactions with shape gestures to enable precise spatial manipulations. *Proc. CHI'11*, ACM (2011), 1581–1590.

28. Wigdor, D., Forlines, C., Baudisch, P., Barnwell, J., and Shen, C. Lucid touch: a see-through mobile device. *Proc. UIST'07*, ACM (2007), 269-278.

29. Wigdor, D., Leigh, D., Forlines, C., et al. Under the table interaction. *Proc. UIST'06*, ACM (2006), 259-268.

30. Wobbrock, J.O., Morris, M.R., and Wilson, A.D. User-defined gestures for surface computing. *Proc. CHI'09*, ACM (2009), 1083-1092.

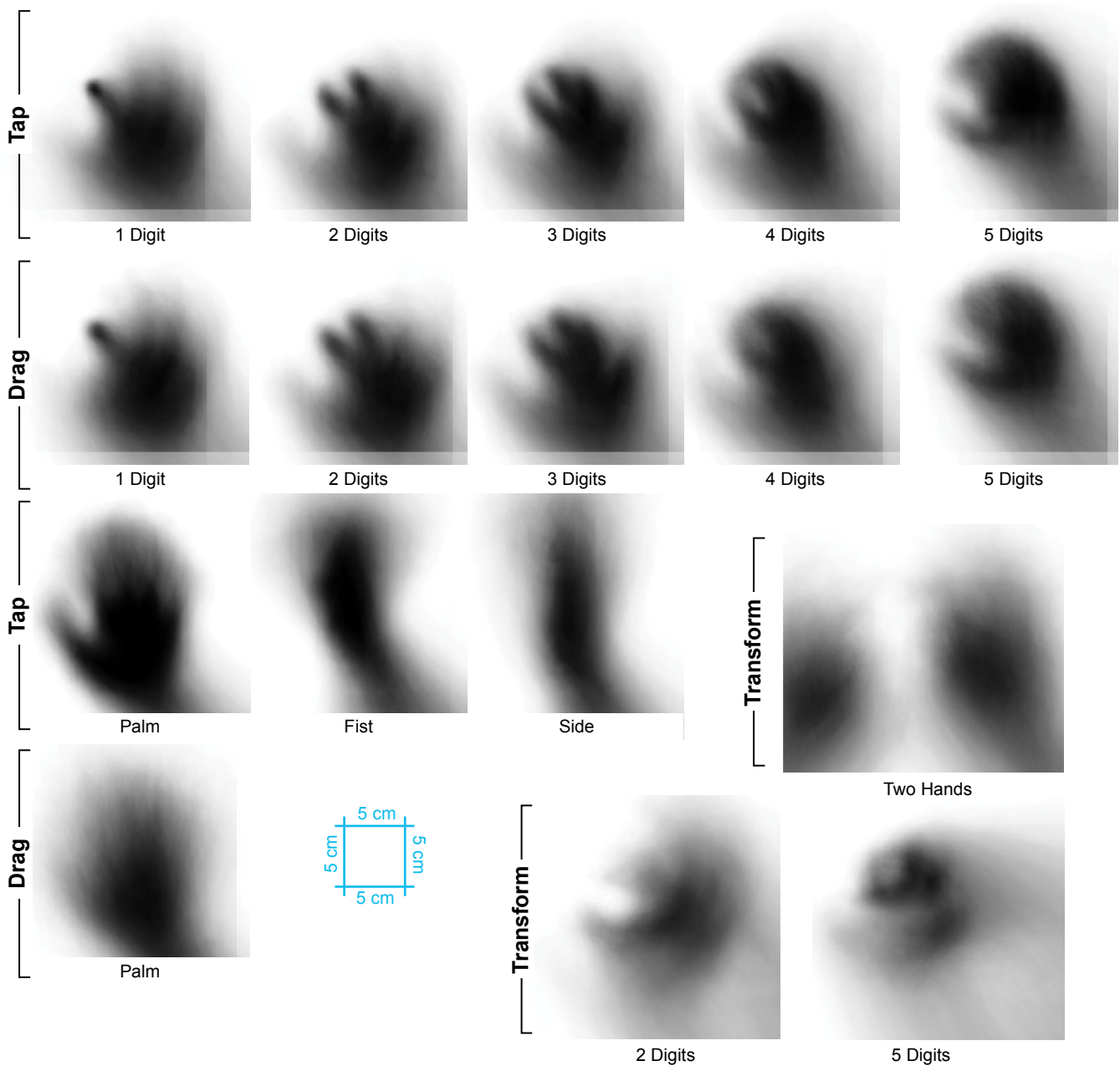31. Wroblewski, L. *Touch Gesture Reference Guide*. 2010.

Figure 8. Mean occlusion silhouettes for *Tap*, *Drag*, and *Transform* by *Posture*. *Tap* and *Drag* silhouettes use end of task capture, Transform silhouettes use beginning and end. Horizontal and vertical hatching is due to rectification and camera cropping.
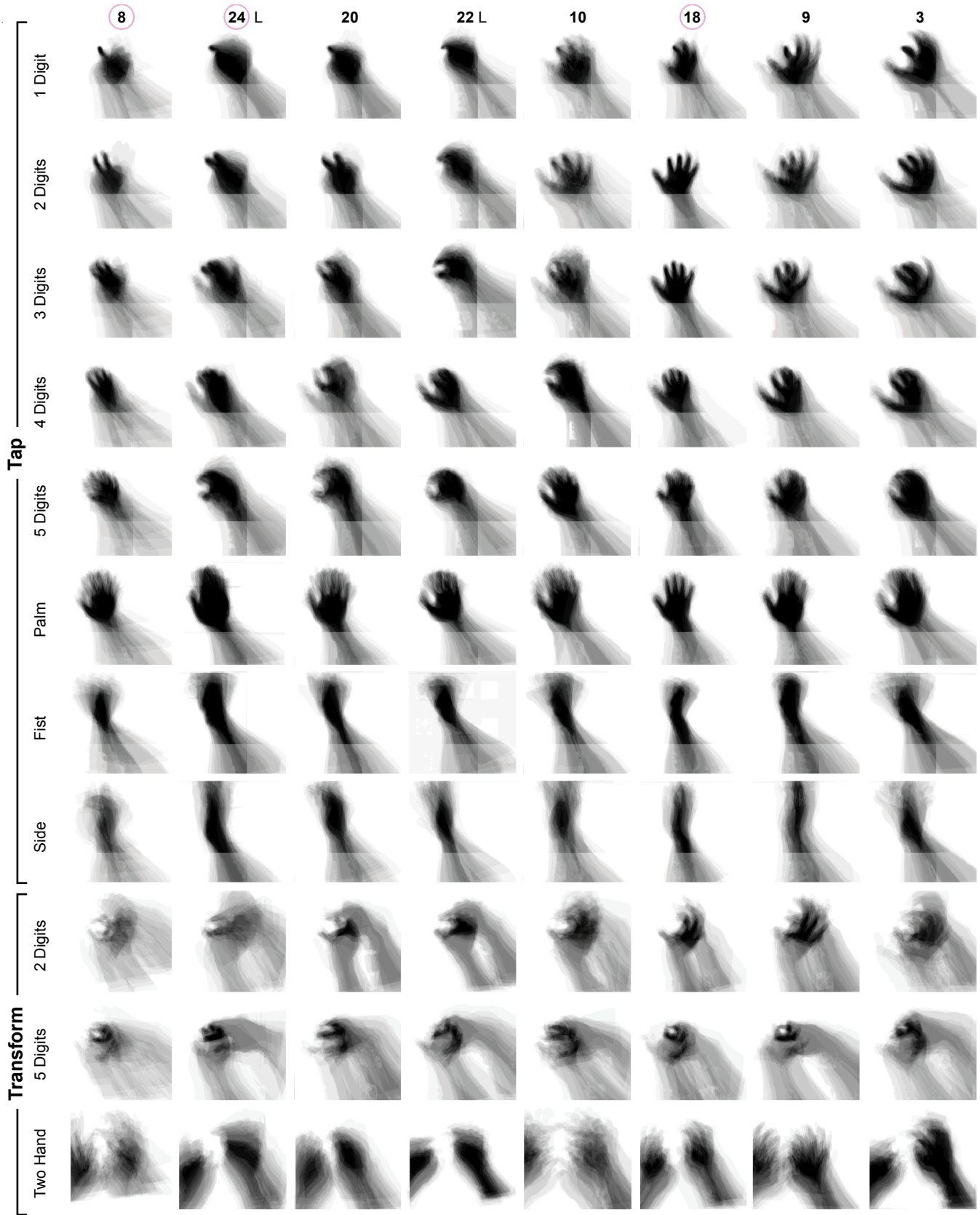
Figure 9. Selected participant mean occlusion shapes by *Posture* for *Tap* and *Transform* tasks. Individuals are in columns with participant numbers adored with a pink circle for females and a 'L' for left-handed participants (shown mirrored)