Hand Occlusion with Tablet-sized Direct Pen Input

Daniel Vogel[1,2], Matthew Cudmore[2], Géry Casiez[3], Ravin Balakrishnan[1], and Liam Keliher[2]

[1] Dept. of Computer Science
University of Toronto, CANADA
{dvogel|ravin}@dgp.toronto.edu

[2] Dept. of Math & Computer Science
Mount Allison University, CANADA
{dvogel|mcudmore|lkeliher}@mta.ca

[3] LIFL & INRIA Lille
University of Lille, FRANCE
gery.casiez@lifl.fr

ABSTRACT

We present results from an experiment examining the area occluded by the hand when using a tablet-sized direct pen input device. Our results show that the pen, hand, and forearm can occlude up to 47% of a 12 inch display. The shape of the occluded area varies between participants due to differences in pen grip rather than simply anatomical differences. For the most part, individuals adopt a consistent posture for long and short selection tasks. Overall, many occluded pixels are located higher relative to the pen than previously thought. From the experimental data, a five-parameter scalable circle and pivoting rectangle geometric model is presented which captures the general shape of the occluded area relative to the pen position. This model fits the experimental data much better than the simple bounding box model often used implicitly by designers. The space of fitted parameters also serves to quantify the shape of occlusion. Finally, an initial design for a predictive version of the model is discussed.

Author Keywords: Hand occlusion, pen input, Tablet PC.

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

INTRODUCTION

Given our familiarity with using pens and pencils, one would expect that operating a tablet computer by drawing directly on the display would be more natural and efficient. However, issues specific to direct pen input, such as the user's hand covering portions of the display during interaction – a phenomena we term *occlusion* (Figure 1a) – create new problems not experienced with conventional mouse input [12].

Compared to using pen on paper, occlusion with pen computing is more problematic. Unlike paper, the results of pen input, or system generated messages, may be revealed in occluded areas of the display. Researchers have suggested that occlusion impedes performance [7,10] and have used it as motivation for interaction techniques [1,14,24], but as of yet there has been no systematic study or model to quantify the amount or shape of occlusion.

Certainly, any designer can simply look down at their own hand while they operate a Tablet PC and take the perceived occlusion into account, but this type of ad hoc observation is unlikely to yield sound scientific findings or universal design guidelines. To study occlusion properly, we need to employ controlled experimental methods.

In this paper we describe an experimental study using a novel combination of video capture, augmented reality marker tracking, and image processing techniques to capture images of hand and arm occlusion from the point-of-view of a user. We call these images *occlusion silhouettes* (Figure 1b). Analyses of these silhouettes found that the hand and arm can occlude up to 47% of a 12 inch display and that the shape of the occluded area varies across participants according to their style of pen grip, rather than basic anatomical differences.

Based on our findings, we create a five parameter geometric model, comprised of a scalable circle and pivoting rectangle, to describe the general shape of the occluded area (Figure 1c). Using non-linear optimization algorithms, we fit this geometric model to the silhouette images captured in the experiment. We found that this geometric model matches the silhouettes with an F1 score [18] of 0.81 compared to 0.40 for the simple bounding box which designers often use implicitly to account for occlusion. The space of fitted parameters also serves as to quantify the shape of occlusion, capture different grip styles, and provide approximate empirical guidelines. Finally, we introduce an initial scheme for a predictive version of the geometric model which could enable new types of *occlusion-aware* interaction techniques.
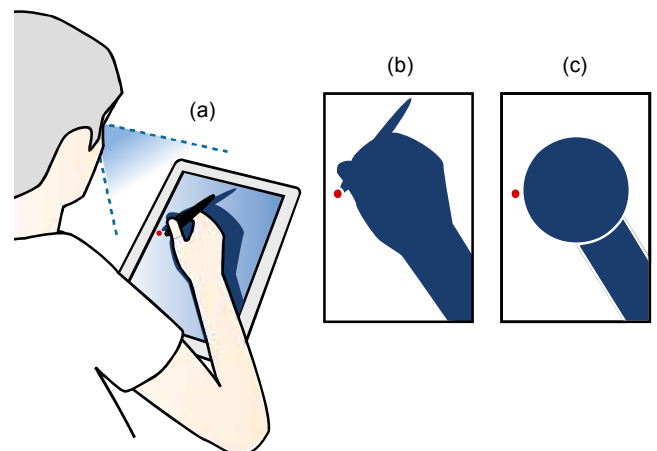

Figure 1: (a) Occlusion caused by the hand with direct pen input; (b) an *occlusion silhouette image* taken from the point-of-view of a user and rectified; (c) a simplified circle and rectangle geometric model capturing the general shape of the occluded area.

RELATED WORK

Few researchers have investigated occlusion directly, but many have speculated on its effect or considered it in the design of interaction techniques. Those who investigate it have done so without a strict control for occlusion. In practice it is very difficult to actually control without resorting to a different input paradigm such as indirect pointing.

Hancock and Booth [7] found that right-handed users selected targets more slowly when located East of the current pen position and attribute this effect to hand occlusion. However, the second slowest time occurred when targets were located in a North-East direction, so it is difficult to conclude. They recommend that pop-up menus should be placed South-West of the current pen location to minimize occlusion for right-handed users.

Based on experimental results, Forlines and Balakrishnan [4] argue that tactile pen feedback can make up for loss of visual feedback due to pen and hand occlusion. They also argue that occlusion is less problematic for serial compared to continuous input because the user can lift their hand to survey the display as part of the task.

Inkpen et al. [10] found a performance advantage and user preference for left-handed scrollbars with left-handed users. All participants cited occlusion problems when using the right-handed scrollbar and the authors note that some participants raised their grip on the pen or arched their hand over the screen to reduce occlusion.

There are several examples of pen interaction techniques that use occlusion as motivation. Ramos and Balakrishnan [14] designed a sinusoidal shaped slider that reduced occlusion from the user's hand. In Apitz and Guimbretières' [1] crossing based interface, they utilized a predominant right-to-left movement direction to counteract occlusion from right-handed users. Zeleznik and Miller [24] describe a tear-off menu technique to reduce occlusion problems.

In the related field of touch screen and tablet top interaction, occlusion is also cited as motivation. Shen et al. [16] discuss table top techniques to combat occlusion, including remote manipulation of objects and visual feedback that expands beyond the area typically occluded by a finger. Other strategies include: placing the hand behind [22] or under the display [23]; and shifting a copy of the intended selection area up and out of the area occluded by the finger [20].

Other researchers have cited problems with occlusion in unrelated experiments and usability studies. Grossman et al. [6] found that users sometimes moved away from the experimental target so they could invoke a hover widget without hand occlusion. Hinkley et al. [9] discovered that conventional GUI tooltips could be easily blocked by the hand. Hinckley et al. [8] found that users needed a chance to lift their hand to view the screen and verify progress when making a lasso selection. Dixon, Guimbretière, and Chen [2] located a start button below their main experimental stimulus to counteract hand occlusion. Ramos et al. [15] argue that accuracy is impaired when using a direct pen because of pen tip occlusion, but provide no evidence.

INITIAL OBSERVATIONAL STUDY

To investigate how occlusion affects usability in more detail, we conducted an initial observational study of Tablet PC interaction with realistic tasks and common software applications. This allowed us to discover a wider range of issues without knowing what they were beforehand. We considered using a performance-oriented experiment design, but controlling for occlusion a-priori is difficult without deeper knowledge about the shape and location of occlusion.

Twelve right-handed participants completed the study. A moderator guided then through an imagined scenario where an office worker must complete a presentation using a Tablet PC while away from their desk. They used typical office applications like a web browser, spreadsheet, and presentation tool. Text input was not required to complete the scenario. During the study, we asked participants to *think-aloud* as we recorded their actions with video and logged pen movements. The scenario took about 50 minutes to complete.

We found that occlusion likely contributed to user errors, led to fatigue, and forced inefficient movements:

- *Hidden Status Messages*. Several participants missed system status messages shown near the bottom of the display. In one case, a participant assumed a "file being saved" confirmation message had been shown beneath their arm, but in fact they missed selecting the save button.

- *Missed Previews*. The presentation application featured real time document previews when browsing text formatting menus. Unfortunately, many participants did not notice this feature, and some assumed their formatting choices had been successful, when they had mistakenly unselected the text behind their arm.

- *Inefficient Movements*. When dragging to highlight text or drawing a selection marquee, we observed large movement deviations past or away from the intended target when moving in a predominately left-to-right direction.

- *Occlusion Contortion*. Like Inkpen et al. [10], we observed participants occasionally arching their wrist while selecting formatting options to simultaneously preview document changes that would otherwise be occluded.

- *Rest Position*. We found that participants had a neutral rest position for their hand located at the right side of the display (all participants were right-handed). Participants commented that this enabled them to survey the display before a task, without their hand getting in the way.

Our results reinforce and expand those of previous research. Yet, we still do not have a thorough understanding of the fundamental characteristics of hand and arm occlusion. For this reason, we continued by examining the shape and area occluded by the hand in a methodical manner. These results could be used by designers to more effectively counteract the effect of occlusion with refined layouts or enable new types of occlusion-aware interaction techniques that compensate for occlusion in real time.

FORMAL EXPERIMENT

Our goal is to measure the size and shape of the occluded area of a tablet-sized display. To accomplish this, we record the participant's view of their hand with a head-mounted video camera as they select targets at different locations on the display. We then extract key frames from the video and isolate *occlusion silhouettes* of the participant's hand as they appear from their vantage point.

Participants

22 people (8 female, 14 male) with a mean age of 26.1 (SD 8.3) participated. All participants were right-handed and pre-screened for color blindness. Participants had little or no experience with direct pen input, but this is acceptable since we are observing a lower level physical behaviour.

At the beginning of each session, we measured the participant's hand and forearm since anatomical dimensions likely influence the amount of occlusion (Figure 2). We considered controlling for these dimensions, but recruiting participants to conform to anatomical sizes proved to be difficult, and the ranges for each control dimension were difficult to define.
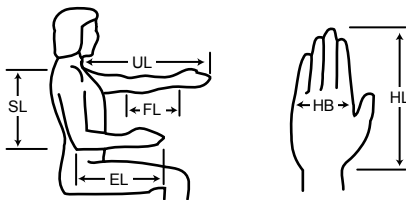
Figure 2. Anthropomorphic measurements (diagram adapted from Pheasant and Hastlegrave [13]).

- EL - elbow to fingertip length
- SL - shoulder to elbow length
- UL - upper limb length including hand
- FL - upper limb length, elbow to crease of wrist, EL - HL
- HL - hand length, crease of the wrist to the tip of finger
- HB - hand breadth, maximum width of palm

Apparatus

The experiment was conducted using a Wacom Cintiq 12UX direct input pen tablet. It has a 307 mm (12.1 inch) diagonal display, a resolution of 1280 by 800 pixels (261 by 163 mm), and a pixel density of 4.9 px/mm (125 DPI). We chose the Cintiq because it provides pen tilt information which is unavailable on current Tablet PCs.

We positioned the tablet in portrait-orientation and supported it such that it was at an angle of 12 degrees off the desk, oriented towards the participant. Participants were seated in an adjustable office chair with the height adjusted so that the elbow formed a 90 degree angle when the forearm was on the desk. This body posture is the most ergonomically sound according to Pheasant and Hastlegrave [13].

To capture the participant's point-of-view, we use a small head-mounted video camera to record the entire experiment at 640 × 480 px resolution and 15 frames-per-second (Figure 3a). The camera is attached to a head harness using hook-and-loop strips making it easy to move up or down so that it can be positioned as close as possible to the center of the eyes, without interfering with the participants' line of sight.

In pilot experiments, we found that we could position the camera approximately 40 mm above and forward of the line of sight, and the resulting image was very similar to what the participant saw.

Printed fiducial markers were attached around the bezel of the tablet to enable us to transform the point-of-view frames to a standard, registered image perspective for analysis. Details of the image analysis steps are in the next section.
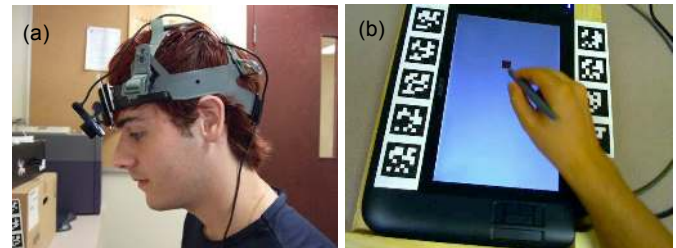
Figure 3. Experiment apparatus: (a) head mounted camera to capture point-of-view; (b) fiducial markers attached to tablet bezel (image is taken from head mounted camera video frame).
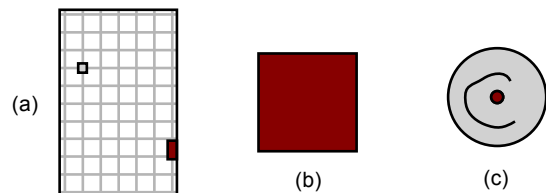
Figure 4. (a) 7 x 11 grid for placement; (b) square; (c) circle target (targets are printed actual size).

Task and Stimuli

Participants were presented with individual trials consisting of an initial selection of a *home target*, followed by selection of a *measurement target*.

The 128 px tall and 64 px wide home target was consistently located at the extreme right edge of the tablet display, 52 mm from the display bottom. This controlled the initial position of the hand and forearm at the beginning of each trial. We observed participants instinctively returning to a similar rest position in our initial observational study.

The location of the measurement target was varied across trials at positions inscribed by a 7×11 unit invisible grid (Figure 4a). This created 77 different locations with target centers spaced 122 px horizontally and 123 px vertically.

We observed two primary styles of pen manipulation in our initial observational study: long, localized interactions where the participant rested their palm on the display (such as adjusting a slider), and short, singular interactions performed without resting the hand (such as pushing a button). Based on this, our task had two types of target selection: *tap* – selection of a 64 px square target with a single tap (Figure 4b); and *circle* – selection of a circular target by circling within a 28 px tolerance between a 4 px inner and 32 px outer radius (Figure 4c). The circle selection is designed to encourage participants to rest their palm, while the tap selection can be quickly performed with the palm in the air. The different shapes for the two selection tasks were intended to serve as a mnemonic to the user as to what action was required.

The circle selection used an ink trail visualization to indicate progress. Errors occurred when the pen tip moved beyond the inner or outer diameter. We wanted this to be difficult enough to require a palm plant, but not tedious. In practice, participants took at least half-a-second to circle the target, which seemed to be enough to plant the palm.

At the beginning of each trial, a red home target and a gray measurement target were displayed. After successfully selecting the home target, the measurement target turned red and the participant selected it to complete the trial. We logged all aspects of pen input, including pressure and tilt.

Design

We presented 3 blocks of trials for each of the two tasks. A block consisted of 77 trials covering each target position in the grid, making 3 repetitions for each grid position and task type. Trials were presented in randomized order within a block and the presentation order of tasks was balanced across participants. Before beginning the first block of a task, the participant completed 40 practice trials. In summary:

> 2 *Tasks* (*Tap, Circle*) × 3 *Blocks* × 77 Target Positions
> = 462 data points per participant

IMAGE PROCESSING

To transform the point-of-view video into a series of occlusion silhouettes, we performed the following steps with custom built software (Figure 5):

Frame Extraction. We extracted video frames taken between successful down and up pen events for the tap target, or just before the circular target was completely circled. To do this, we had to synchronize the video with the data log. We used a visual time marker which functions similar to a movie clapperboard. The time marker is a large red square containing a unique number. When this square is tapped, it disappears and a timestamp is saved to our data log. After the experiment, we scrubbed through the video and found the video time where the time marker disappeared. Then, using linear interpolation between bounding time marks, we located the corresponding video frame for a given log time. In most cases, the frame captured the pen at the intended target location, but occasional lags during video capture produced a frame with the pen separated from the target location.

Rectification. We used the ARToolkitPlus augmented reality library [21] to track the fiducial markers in each frame and determine the location of the four corners of the display. In practice, this sometimes required hand tuning when the markers were occluded by the hand or were out of frame due to head position. Using the four corner positions, we unwarped the perspective using the Java Advanced Image [17] functions `PerspectiveTransform` and `WarpPerspective` with bilinear interpolation, and cropped it to a final 267×427 px image. Note that due to our single camera setup, the unwarping will shift the image of the hand down slightly relative to the actual eye view. As an example, if the eye position is at the end of a vector 500 mm and 50° from the centre of the tablet, and the camera is located 40 mm above and forward of the eye, the unwarped image of a point on the hand 40 mm above the tablet will be shifted down by 6.2 mm (about 4 px in our unwarped image). The exact error

will vary according to participant size and grip style, but the values above are typical. Rather than try to compensate for this slight shift and possibly introduce additional errors, we accepted this as a reasonable limitation of our technique.

Isolation. We used simple image processing techniques to isolate the silhouette of the hand. First, we applied a light blur filter to reduce noise. Then we extracted the blue color channel and applied a threshold to create an inverted binary image. We were able to use the blue channel to isolate the hand because the camera's color balance caused the display background to appear blue (it was actually white). Since the color space of skin is closer to red, this made isolating the hand relatively easy. To remove any edge pixels from the display bezel, we applied standard dilation and erosion morphological operations [3]. Finally, we filled holes based on the connectivity of pixels to produce the final silhouette.
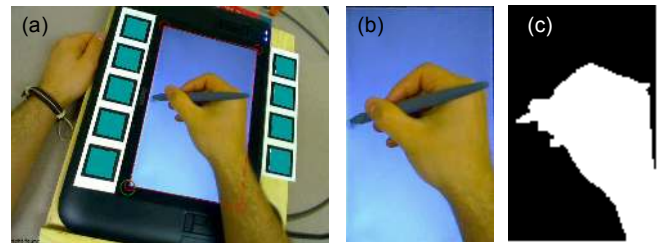

Figure 5. Image processing steps: (a) frame extraction; (b) rectification; (c) silhouette isolation.

RESULTS

Unfortunately, lighting and video problems corrupted large portions of data for participants 7, 14, 21, and 22 making isolation of their occlusion silhouettes unreliable. Capture problems with participant 8 corrupted the first block, but we kept this participant and their remaining blocks. In the end, our analysis included 18 out of the original 22 participants (6 female, 12 male) with a mean age of 26.3 (SD 8.4). In addition, we removed data trials when capture lag produced silhouettes more than 20 mm from the target location (7.8% of trials). These types of problems are typical when using video capture to generate empirical data: it is difficult to produce the same kind of "clean" data generated by experiments recording straightforward variables such as performance time and errors. Researchers attempting similar work should recruit extra participants and run multiple trials as we did, to ensure a reasonable amount of clean trials can be obtained.

Participants occasionally produced errors (mean 4.4%), but we included the silhouette regardless. Since each target must be successfully tapped or circled before continuing, the final video frame for an error trial would not differ. Also, the logged pen tilt values were very noisy, in spite of silhouette images suggesting tilt should be more uniform. Our attempts to filter them were unsuccessful, and we were forced to leave them out of our analysis.

Occlusion Ratio

We define the *occlusion ratio* as the percentage of occluded pixels within all possible display pixels. We used a ratio, rather than actual area, for unit independence. The actual area can be computed using the display area of 42,543 mm^2.
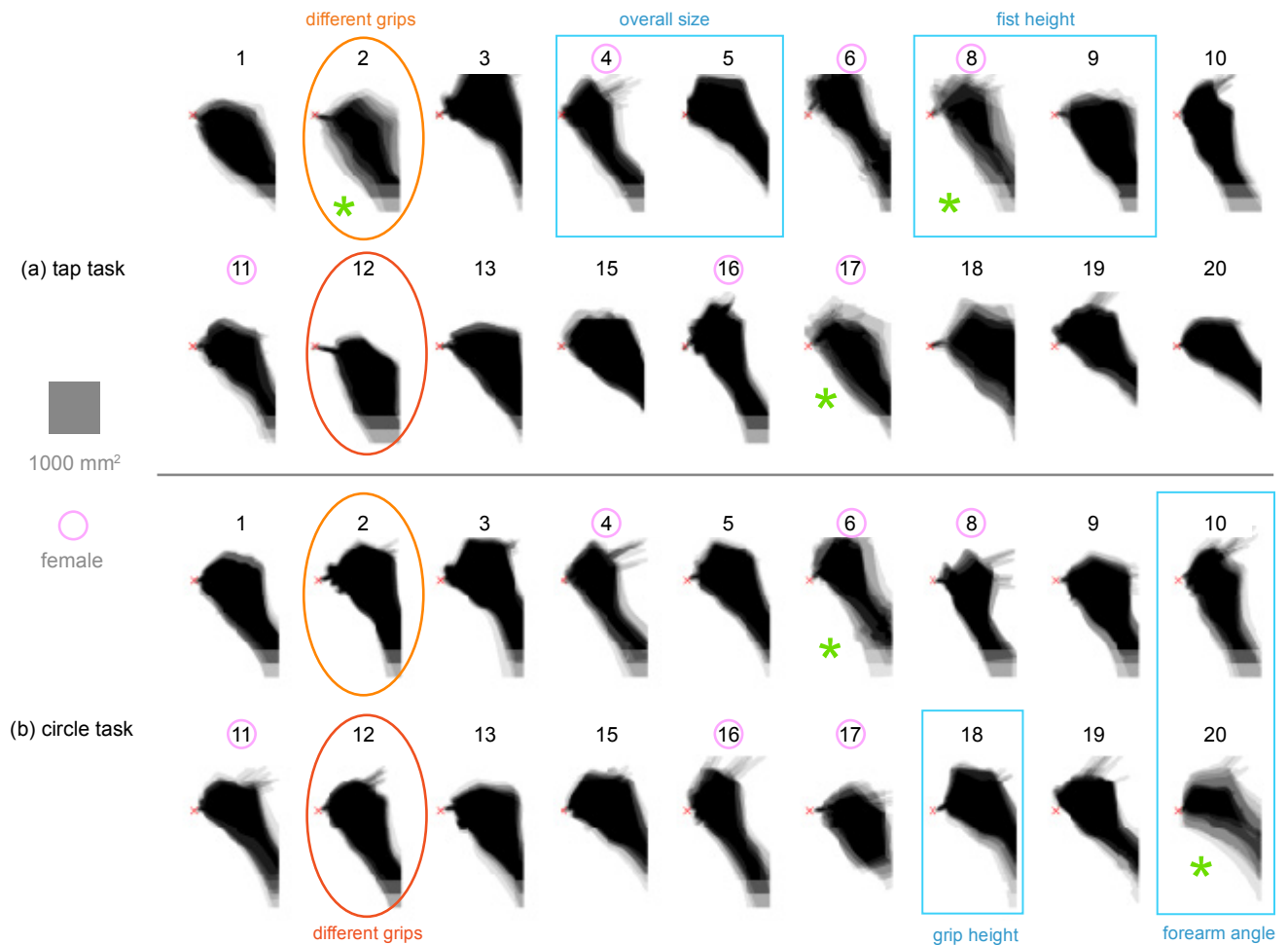
Figure 6. Occlusion shape silhouettes for each participant for: (a) tapping task and (b) circling task. 9 samples from 3 pen positions at middle-left portion of display; see text for discussion of participant and task comparison highlights.

Since occlusion ratio varies according to pen location, we calculate the occlusion area for each X-Y target location in the 7 x 11 grid. Not surprisingly, we found the highest occlusion ratios when the pen was near the top left of the display. However, the highest value did not occur at the extreme top, but rather a short distance below (Figure 7). The highest values did not differ greatly by task with 38.6% for circle (SD 6.2) and 38.8% for tap (SD 14.2). Participant 1 had the highest occlusion ratio with 47.4% for tap and 46.3% for circle.
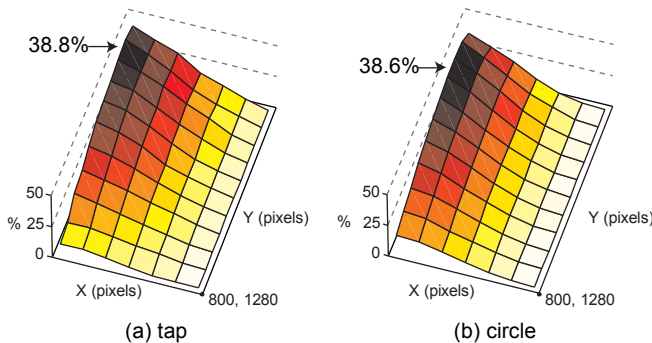

Figure 7. Occlusion ratio, plotted by X-Y display location for: (a) tap task; (b) circle task.

These mean ratios may reflect a sampling bias among our participants since controlling for aspects such as anatomical size and pen grip style is difficult to do a-priori. To help address this, we compare occlusion ratios given participant size.

Influence of Participant Size

We established a simple size metric S to capture the relative size of each participant's arm and hand compared to the general population. S is the *mean* of three ratios between a participant measurement and 50th percentile values from a table of anthropomorphic statistics[1]. We use measurements for shoulder length (SL), hand length (HL), and hand breadth (HB). Since tables of anthropomorphic statistics are divided by gender, we compute S for men and women using different 50th percentile values. We found mean S values of 0.99 (SD 0.04) and 1.01 (SD 0.06) for men and women respectively, indicating that the size of our participants was representative.

We expected to see a relationship between S and the maximum occlusion ratio since, larger hands and forearms should cover more of the display. However, a plot of S vs. maximum occlusion ratio does not suggest a relationship (Figure 8).

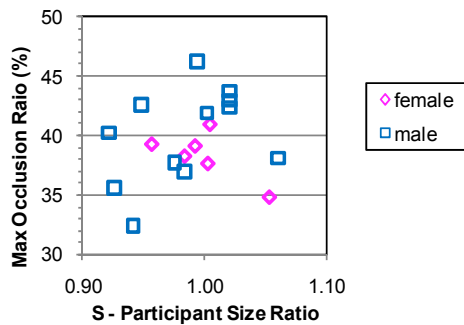[1] Anthropomorphic statistics for U.S Adults 19 to 65 years old [13].

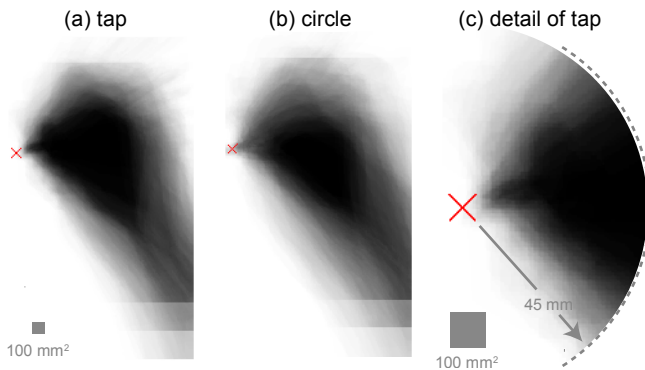Figure 8. Participant size (S) vs. max occlusion ratio.

Figure 9. Mean occlusion shapes: (a) tap task; (b) circle task; (c) detail of tapping mean shape.
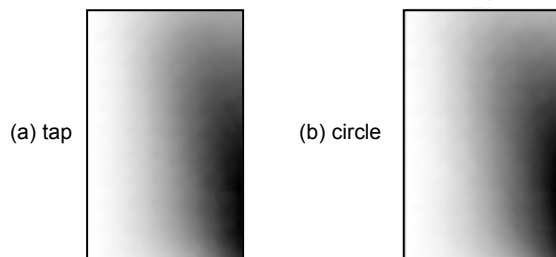

Figure 10. Pixels most likely to be occluded given a uniform distribution of pen positions: (a) tap task; (b) circle task (darker pixels are occluded more often).

Occlusion Shape

Although occlusion ratio gives some sense of the scope of occlusion, it is the shape of the occluded pixels relative to the pen position that is most useful to designers. Figure 6 illustrates the mean shapes for participants for circling and tapping tasks. Since the captured image of the forearm and hand are increasingly cropped as the pen moves right and downward, we illustrate shapes for positions sampled near the middle-left portion of display.

It is immediately apparent that occlusion shape varies between participants. There are differences which are likely due to anatomical size, possibly related to gender: compare how slender female participant 4 appears compared to male participant 5. Some participants adopt a lower hand position occluding fewer pixels above the target: contrast the height of participant 8 with participant 9. The forearm angle also often varies: for example, participant 20 has a much higher angle than participant 10. A few participants grip the pen far away from the tip, occluding fewer pixels around the target: participant 18 in the tapping task is one example.

When comparing individual participant shapes between the tap and circle tasks, the visual differences are more subtle and inconsistent. For example, we expected the higher speed of the tapping task to create a more varied posture resulting in blurry mean shapes. This seems to be the case for participants 2, 8, and 17, but there are contrary examples when circling shapes are more blurred: see participants 6 and 20. Only participants 2 and 12 seemed to adopt very different postures for tapping (low) and circling (high).

The combined participant mean shape gives an overall picture of occluded pixels near the pen position across all participants (Figure 9). As with individual participants, differences between tasks are subtle. The tapping task mean shape appears slightly larger, higher, and sharper compared to the circling task. In both cases, the typically occluded pixels form a circular blob centered far to the right of the pen tip with fewer pixels occluded in the immediate vicinity of the pen's position (Figure 9c).

Pixels Most Likely to be Occluded

Another way to view occlusion shape is to look at which display pixels are most likely to be occluded given a distribution of pen positions. To create a simple baseline for analysis, we assume that the probability of accessing any position on the display is uniform. Under this distribution, commonly occluded display pixels across participants and target positions form a cluster of frequently occluded pixels emanating from the lower two-thirds along the right edge (Figure 10). There appears to be no difference between circle and tap tasks.

A uniform distribution of pen positions is not representative of common application layouts: consider the frequency of accessing menus and toolbars located along the top of the display. With this in mind, the often occluded pixels near the bottom right may be even more likely to be occluded.

DISCUSSION

The results of our experiment reveal four main findings:

1. A large portion of the display can be occluded depending on pen position; with our participants it was typically as high as 38%, but could range up to 47%.

2. The pixels immediately below the pen position are not occluded by the hand as much as we expected, but more pixels are occluded above the pen tip horizon than previously thought. Given that our experimental set-up tended to shift the capture silhouette down slightly, this could be even more pronounced than what we observed.

3. Individuals seem to have a signature occlusion silhouette, but comparing silhouettes between different individuals can reveal large differences.

4. There appears to be no simple relationship between the size of the occluded area and anatomical size.

The Impact of Grip Style

The largest differences in occlusion silhouettes are due to the different styles of pen grips used by our participants (Figure 11). We searched the motor behaviour and graphonomics literature for a definitive classification of pen grip. Greer and Lockman [5] observed three different styles of pen grips used by adults, but do not describe these in detail. With our par-

ticipants, we found that grip style varied predominately across three dimensions: size of fist, angle of pen, and height of grip location on pen. We believe it is these characteristics of grip style that interact with anatomical measurements and ultimately govern occlusion area.
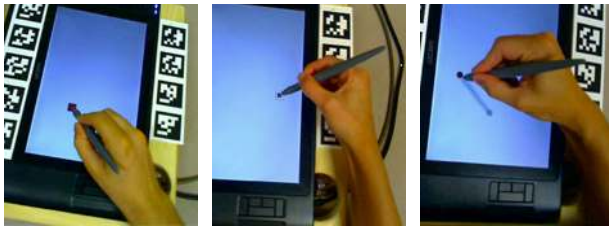

Figure 11. Grip styles: (a) loose fist, low angle, medium grip height; (b) tight fist, high angle, high grip height; (c) loose fist, straight angle, low grip height.

Left-handed Users

We conducted a small follow-up study with two left-handed users. Similar to Hancock and Booth's finding with performance [7], we found that the left-handed data mirrored the right-handed individuals.

Influence of Clothing

We gathered our data for sleeveless participants to maintain a consistent baseline, but we recognize that size of the occlusion silhouette could be much larger when clothed (consider using a tablet while wearing a loose fitting sweater or jacket). As a general rule, Pheasant and Hastlegrave [13] suggest adding 25mm to all anatomical dimensions for men and 45mm for women to account for thickness of clothing.

GEOMETRIC MODEL OF OCCLUSION SHAPE

The experiment revealed that the occlusion shape was somewhat uniform within a participant and high level similarities appeared across participants. We wondered if a simple geometric model could describe the general shape and position of the occlusion silhouettes. If so, by fitting this model to the actual silhouettes, the resulting model parameters could serve as empirical guidelines for designers. Moreover, this geometric representation could form the basis for a predicative version of model: in real time, a system would be aware of occluded portions of the interface without the aid of elaborate sensors. For example, imagine an interface that knows when a status message is occluded, and re-displays it as a bubble in a nearby non-occluded area instead.

There are many ways to approach modeling the shape of the occlusion silhouettes. Perhaps the most straightforward approach is to assume pixels below and to the right of the pen's position are occluded, an approach which some designers and researchers seem to use implicitly. We refer to this as a *bounding rectangle model* (Figure 12c). This model is constant relative to the pen's position and requires no other input, but the accuracy is poor. At the other end of the spectrum, we could create a model with a flexible shape such as one composed of Bézier spline segments (Figure 12a). While this would certainly yield a very accurate representation of the occluded area, the huge number of parameters would make fitting and interpreting the model difficult and hence impractical for creating empirical guidelines. Our aim then is to cre-

ate a simple model with a small number of parameters, yet still produce a reasonable degree of accuracy.

Scalable Circle and Pivoting Rectangle Model

We noticed that the occlusion silhouettes produced by the experimental data often resembled a lopsided circle for the fist, a thick narrowing rectangle sticking out the bottom for the arm, and, with some participants, there was also a thinner rectangle puncturing the top of the ball for the pen. This meant that a single *oriented* bounding box would be unlikely to capture all grip styles accurately. Our first approach then, was to create a geometric model using an ellipse for the fist, an isosceles trapezoid for the arm, and a rectangle for the pen. However, even this model had 11 parameters and automatically fitting the geometry to our experimental data was problematic. Instead, we simplified our representation further to an offset circle and a rectangle with only the following 5 parameters (also illustrated in Figure 12b):

- q is the offset from the pen position **p** to the circle edge,
- r is the radius of the circle over the fist area,
- Φ is the rotation angle of the circle around **p** (expressed in degrees where $\Phi = 0°$ when the centre is due East, $\Phi = -45°$ for North-East, and $\Phi = 45°$ for South-East),
- Θ is the angle of rotation of the rectangle around the centre of the circle (using the same angle configuration as Φ),
- w is the width of the rectangle representing the forearm.

Note that the length of the rectangle is infinite for our purposes. If we were building a model for larger displays, this may become another parameter, but at present we are concerned with tablet-sized displays like the portable Tablet PC.
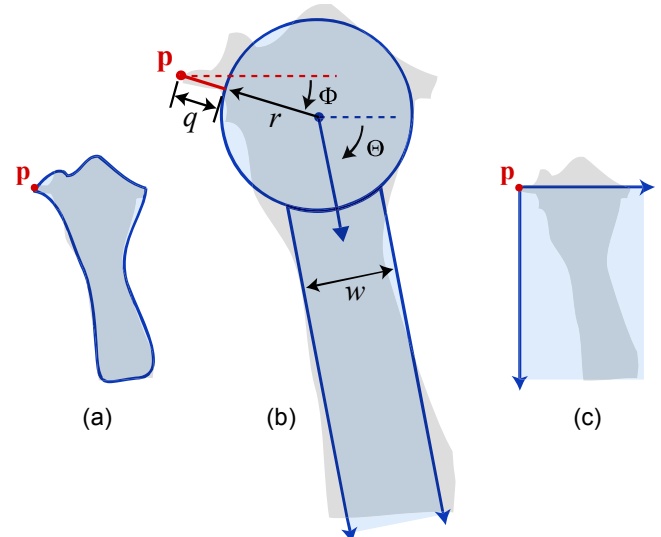
Figure 12. Three occlusion shape models: (a) Bézier spline; (b) circle and rectangle; (b) bounding rectangle. **p** is the position of the pen.

Fitting the Geometric Model to Captured Silhouettes

For each silhouette image from our experiment, we use non-linear optimization techniques to set the five parameters of the geometric model so that it "fits" over the silhouette as accurately as possible. Note that other optimization algo-

rithms, or other fitting techniques, can be used – we describe our process as an example of one possible procedure.

To guide the optimizers to an *optimal fit*, we create an objective function. Our objective function returns 0 for a perfect fit, when the geometry matches the silhouette exactly, and increases as the alignment diverges. This is computed using two area ratios: the percentage of the silhouette image not covered by the geometry, and the percentage of geometry area not covering the silhouette. We give slightly more weight to the first ratio to favour covering more occluded pixels at the potential cost of covering non-occluded pixels as well. To compute these area ratios, we converted the silhouette binary images to polygons and computed the ratios analytically. The inverse would have worked as well, converting the geometric model to a binary image and "counting pixels" to calculate the ratios. To reduce the chance of the optimizer finding anatomically improbable configurations, we constrained the possible angles for Θ and Φ to be in (0, 90) and (-90,90) respectively. We also added smaller objective terms to encourage a smaller rectangle width w and shorter distance from circle to pen position q.

One problem during our initial optimization attempts was caused by cropped occlusion silhouette images. As the pen moves towards the bottom right, more and more of the forearm and fist are outside the display area and were cropped during image processing, making it difficult for the optimizer to find an optimal placement of the geometry. We solved this by fitting the geometry in two stages for each participant and target type (circle and tap). In the first stage, we optimized all parameters using 3 pen positions near the upper left portion of the display, since the hand and forearm would not be cropped. Using these values, we found mean values for r and w. In stage two, we locked r and w to these mean values and optimized over the remaining parameters. We rationalize this two-stage strategy by arguing that the size of silhouettes produced by the fist and forearm is unlikely to vary greatly according to X- and Y-coordinate, but their position and angle may change. If we had silhouette images capturing the entire image of the fist and forearm including parts outside the display, we would not have needed this step.

We ran the optimization using two algorithms in sequence over all target locations except the rightmost where the hand was completely off the display. First, a pattern search algorithm found a candidate neighbourhood for the global minima, and then a standard gradient search found a local minima (see [19] for algorithm descriptions). We could not use gradient search alone since our objective function produced a rough error surface. The total time for optimization was approximately 12 hours on a 2.66 GHz quad processor.

Testing the Accuracy of the Geometric Model

We use precision-recall plots and F1 scores, standard measures used in information retrieval [18], to test our geometric model's fidelity. This can be justified by considering the geometric model as a binary classifier which labels each pixel as occluded or not occluded. High precision means that pixels labelled by the model as occluded are actually occluded, but other occluded pixels may have been missed. High recall

means that the model is correctly labelling occluded pixels, but could also be labelling non-occluded pixels as occluded.

By plotting the results of each fitted silhouette in precision-recall space, we get a sense for how well the model performs (Figure 13). A near-perfect model will have a concentration of points in the upper right corner and an F1 score close to 1. We calculate mean F1 scores across all cases.

Our geometric model has a mean F1 score of 0.81 (SD 0.20) and the precision-recall plots suggests very high recall, but some falloff for precision (Figure 13b). This precision falloff is expected since we designed our optimization objective function to fit the model in a more conservative manner, favouring covering more occluded pixels at the potential cost of covering non-occluded pixels. A designer would probably be more comfortable over compensating for occlusion, but this is a limitation. We included the bounding box model as a baseline comparison. It has a F1 score of 0.40 (SD 0.20) and a precision-recall plot (Figure 13a) suggesting a poor fit in terms of both precision and recall.

Note that our geometric model is only one of many potential models. For example, although we ruled out the *oriented* bounding box initially, it may be satisfactory in some situations. Evaluating our model, or any others, in real applications remains future work.
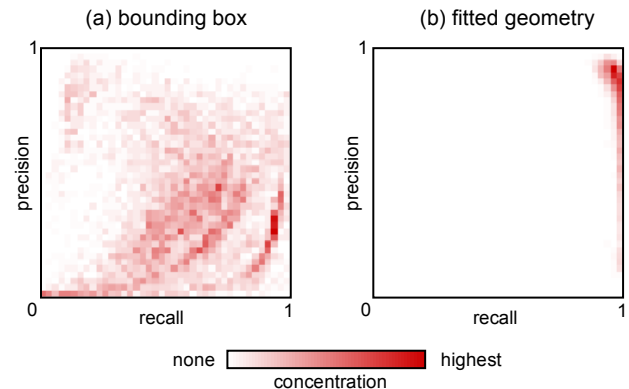

Figure 13. Precision-recall concentration plots illustrating performance of: (a) simple bounding box model; (b) fitted circle and rectangle geometry. More points in the upper right indicate better performance.

Space of Fitted Parameters

We can use the space of optimized parameters to further investigate the shape of the occlusion silhouettes. To enable comparison with Figure 6, in Table 1 we summarize the participant mean parameters for the circle task across the same 9 pen positions at the middle-left portion of the display. This focuses our comparison on positions in which the entire hand and forearm silhouette is captured without cropping and reduces variance from parameters, such as Θ (the forearm angle), as it changes across pen coordinate positions.

For the most part, the fitted parameters match visual intuition from the mean silhouette images in Figure 6. For example, a low value of Φ indicates a high grip and a high value of Φ indicates a low grip: the two lowest Φ values of -25.5 and -21.4 for participants 6 and 16 match the high grips seen in Figure 6, the high Φ values of 12.1 and 11.9 for participants

9, and 17 match their low grips. Likewise, q captures how close participants hold the pen relative to the pen tip: high q values of 28.3 and 26.1 for participant 2 and 17 indicate they hold the pen far from the tip, and low q values of 5.3 and 4.1 for participants 16 and 19 indicate the opposite. A comparison of other mean parameters with the silhouettes in Figure 6 reveals similar patterns.

We expected more variance in parameter values between participants than within a participant. For the most part, this was the case, but there are exceptions. Participants 6 and 20 have high variance, but we expected this from their blurry mean silhouettes in Figure 6. The high variance for participant 17 is somewhat surprising; we speculate that this may be due to image cropping caused by the grip style.

	q	r	Φ	Θ	w
1	12.3 (2.2)	61.5 (1.4)	10.1 (3.7)	58.0 (2.3)	58.9 (1.8)
2	28.3 (3.8)	64.0 (6.6)	-4.9 (3.8)	63.5 (3.6)	62.7 (2.3)
3	14.9 (2.5)	64.5 (1.1)	-13.9 (2.6)	57.7 (3.7)	72.8 (3.3)
4	7.1 (4.7)	50.3 (0.8)	-7.4 (5.1)	60.1 (3.3)	49.0 (2.2)
5	17.6 (4.3)	59.9 (0.8)	-7.9 (3.9)	53.8 (2.1)	61.8 (2.2)
6	15.3 (4.8)	58.4 (13.3)	-25.5 (8.1)	60.5 (5.0)	58.9 (5.5)
8	14.1 (5.5)	53.8 (1.4)	8.6 (6.3)	68.6 (4.0)	50.1 (1.7)
9	21.5 (1.9)	63.2 (1.1)	12.1 (4.5)	62.0 (3.6)	59.5 (1.1)
10	9.5 (3.4)	55.3 (1.7)	-1.6 (3.9)	69.2 (3.2)	54.9 (4.0)
11	15.5 (2.7)	56.8 (1.1)	-7.0 (5.9)	53.8 (5.4)	56.1 (3.3)
12	14.9 (3.5)	59.5 (0.8)	1.9 (3.5)	61.5 (2.6)	57.0 (3.5)
13	23.9 (3.8)	65.4 (1.2)	7.6 (4.1)	56.6 (3.4)	61.0 (5.7)
15	13.0 (3.8)	64.6 (1.6)	-9.1 (2.7)	45.8 (5.0)	63.9 (3.9)
16	5.3 (3.0)	52.6 (1.8)	-21.4 (6.9)	61.0 (3.2)	50.7 (2.4)
17	26.1 (7.9)	63.4 (6.8)	11.9 (7.7)	39.9 (23.3)	60.5 (19.1)
18	23.5 (2.9)	60.4 (0.9)	-16.4 (3.9)	46.9 (6.5)	73.1 (1.8)
19	4.1 (3.5)	58.6 (1.0)	-11.0 (2.8)	48.1 (3.8)	48.7 (2.9)
20	11.2 (5.4)	56.5 (9.5)	-11.9 (12.4)	39.0 (8.3)	58.6 (1.4)
all	**15.5** (7.9)	**59.5** (6.2)	**-5.1** (12.3)	**55.6** (10.8)	**59.0** (8.5)

Table 1. Summary statistics of fitted geometric model parameters for each participant for circle task (9 samples from 3 pen positions at middle-left portion of display, standard deviations shown in parenthesis).

DESIGN IMPLICATIONS

Our findings suggest three main design implications.

1. Avoid showing visual feedback or related widgets in the area described by the circular area to the right of the pen (see Figure 9c).

2. Avoid displaying status or alert messages in the bottom right area of the display since it may be often occluded by the hand (Figure 10).

3. When designing for occlusion, be aware that real users have a wide range of pen grips and postures.

We can use our fitted geometry model parameters to make implications 1 and 2 more specific. Assuming our experimental sample is representative, the mean parameter values across

all participants (bottom row of Table 1) could form a universal mean configuration for the geometric model (Figure 14). In practice, this may not be the most accurate solution given implication 3 and because these mean values include only a subset of pen positions to avoid introducing higher variance, but it may suffice as a rough guide for designers.
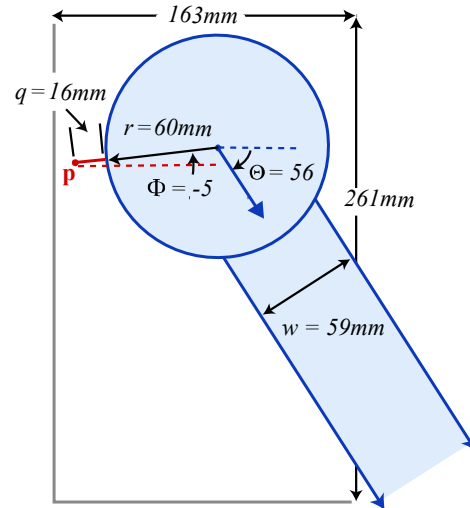

Figure 14. Mean configuration for the geometric model (with our display size context as context).

Towards A Predictive Model

To address the limitation with the geometric model described above, we developed an initial scheme for a predictive geometric model which would adjust according to pen coordinate position and user grip style. This would provide designers with more accurate models to fine tune layouts according to a custom set of users and display positions, in addition to enabling new types of occlusion-aware interfaces introduced above. This model is a work in progress, and we briefly describe it here with initial test results.

We simplified the problem significantly by first assuming constant values could be found for Φ, q, r, and w for each user. We imagine a short calibration process where a user fine tunes the size and position of the rendered geometric model shapes to match their grip style, indirectly setting these values. Note that the usability of the model may be limited if the calibration process is too long or cumbersome. Also, if reliable pen tilt information is available, we believe that Φ could be determined directly.

In early model prototypes, we found that an incorrect Θ could drastically alter the position of the forearm and throw the model off. To correctly model Θ, we use a kinematic model [11] to estimate the posture of the forearm. This requires several simplifying assumptions to make the problem tractable, and required us to estimate the position of the torso and arm segment lengths using gender-specific anthropometric data. In spite of these approximations and simplifying assumptions, testing our model against the experimental data yielded an F1 score of 0.77 (SD 0.16). More work remains to be done to test this model in real world applications and compare its performance against static occlusion models such as the mean version of our geometric model or bounding box.

CONCLUSIONS AND FUTURE WORK

Previous researchers and designers have, for the most part, made implicit assumptions like the bounding box model to determine what areas of the display are likely occluded by the silhouette of the hand and forearm. To move beyond such assumptions, we have provided a systematic study of occlusion using experimental data gathered with a head mounted video camera, augmented reality marker tracking, and image processing techniques. Our results suggest that the shape of occlusion varies among participants due to anatomical size and grip style, and for the most part, individuals adopt a consistent posture for long and short interactions. Moreover, the general shape of occluded pixels is higher relative to the pen than previously thought. Despite varying occlusion shapes across users, we were able to develop a simple five parameter geometric model that captures the general shape of the occluded area and use the space of parameters to characterize and further quantify the shape of occlusion.

It is important to note that we have focused on occlusion resulting from a typical, neutral posture. Inkpen et al. [10] report that users will contort their hand to overcome occlusion problems. We are currently investigating this phenomenon.

Extending our results to very large vertical displays or very small hand-held devices has yet to be explored. In terms of our geometric model, we expect that most parameters relating to grip style are unlikely to change (q, Φ, r, and w)**,** but the values of these parameters would differ as the size and orientation of the display diverges from the tablet-sized display used in our study. In the case of large displays, the variance of parameter values would likely increase substantially.

As future work, we plan to refine the predictive version of the model introduced here, and utilize it to create occlusion-aware interfaces. We also plan to contrast our findings with occlusion silhouettes resulting from touch input.

REFERENCES

1. Apitz, G. and Guimbretiére, F. CrossY: a crossing-based drawing application. *Proceedings of UIST*, ACM (2004), 3-12.

2. Dixon, M., Guimbretière, F., and Chen, N. Maximizing Efficiency in Crossing-Based Dialog Boxes. *Proceedings of SIGCHI*, ACM (2008), 1623 - 1632.

3. Dougherty, E.R. *An Introduction to Morphological Image Processing*. Bellingham, Wash., USA, 1992.

4. Forlines, C. and Balakrishnan, R. Evaluating tactile feedback and direct vs. indirect stylus input in pointing and crossing selection tasks. *Proceedings of SIGCHI*, ACM (2008), 1563-1572.

5. Greer, T. and Lockman, J.J. Using writing instruments: invariances in young children and adults. *Child Development 69*, 4 (1998), 888(15).

6. Grossman, T., Hinckley, K., Baudisch, P., Agrawala, M., and Balakrishnan, R. Hover widgets: using the tracking state to extend the capabilities of pen-operated devices. *Proceedings of SIGCHI*, ACM (2006), 861-870.

7. Hancock, M.S. and Booth, K.S. Improving menu placement strategies for pen input. *Proceedings of Graphics Interface*, (2004), 221-230.

8. Hinckley, K., Guimbretiére, F., Agrawala, M., Apitz, G., and Chen, N. Phrasing techniques for multi-stroke selection gestures. *Proceedings of Graphics Interface*, (2006), 147-154.

9. Hinckley, K., Zhao, S., Sarin, R., et al. InkSeine:In Situ search for active note taking. *Proceedings of SIGCHI*, ACM (2007), 251-260.

10. Inkpen, K., Dearman, D., Argue, R., et al. Left-Handed Scrolling for Pen-Based Devices. *International Journal of Human-Computer Interaction 21*, 1 (2006), 91-108.

11. Isaacs, P.M. and Cohen, M.F. Controlling dynamic simulation with kinematic constraints. *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, ACM (1987), 215-224.

12. Meyer, A. Pen computing: a technology overview and a vision. *SIGCHI Bull 27*, 3 (1995), 46-90.

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

14. Ramos, G. and Balakrishnan, R. Fluid interaction techniques for the control and annotation of digital video. *Proceedings of UIST*, ACM (2003), 105-114.

15. Ramos, G., Cockburn, A., Balakrishnan, R., and Beaudouin-Lafon, M. Pointing lenses: facilitating stylus input through visual-and motor-space magnification. *Proceedings of SIGCHI*, ACM (2007), 757-766.

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

17. Sun Microsystems. Java Advanced Imaging API. http://java.sun.com.

18. Van Rijsbergen, C.J. *Information Retrieval*. London ; Toronto, 1979.

19. Venkataraman, P. *Applied Optimization with MATLAB Programming*. New York, 2002.

20. Vogel, D. and Baudisch, P. Shift: a technique for operating pen-based interfaces using touch. *Proceedings of SIGCHI*, ACM (2007), 657-666.

21. Wagner, D. and Schmalstieg, D. ARToolKitPlus for Pose Tracking on Mobile Devices. *Proceedings of 12th Computer Vision Winter Workshop (CVWW'07)* , (2007).

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

23. Wigdor, D., Leigh, D., Forlines, C., et al. Under the table interaction. *Proceedings of UIST*, ACM (2006), 259-268.

24. Zeleznik, R. and Miller, T. Fluid inking: augmenting the medium of free-form inking with gestures. *Proceedings of Graphics Interface*, (2006), 155-162.