Exhibit F

HAND TRACKING, FINGER IDENTIFICATION, AND CHORDIC MANIPULATION ON A MULTI-TOUCH SURFACE

by

Wayne Westerman

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical Engineering

Spring 1999

HAND TRACKING, FINGER IDENTIFICATION, AND CHORDIC MANIPULATION ON A MULTI-TOUCH SURFACE

by

Wayne Westerman

Approved: ______________________________
Neal Gallagher, Ph.D.
Chair of the Department of Electrical Engineering

Approved: ______________________________
Andras Z. Szeri, Ph.D.
Interim Dean of the College of Engineering

Approved: ______________________________
John C. Cavanaugh, Ph.D.
Vice Provost for Academic Programs and Planning

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

John Elias, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Charles Boncelet, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Phillip Christie, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

Kenneth Barner, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: ______________________________

John Scholz, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

Abundant thanks go to my adviser, John Elias, whose fond support, daily teamwork, and unfathomable hardware know-how gave me a unique foundation upon which to compose a dissertation.

Dr. Neal Gallagher, for inviting me to Delaware, ensuring my research in the Electrical Engineering Department and other parts of campus was always fully supported, offering weekly spiritual advice, and challenging me with proclamations of what could and could not be done. May he continue to carry his wisdom all across the country.

Dr. Charles Boncelet, Dr. Kenneth Barner, and Dr. John Scholz, for suggesting many helpful revisions to this manuscript.

Dr. Phillip Christie, for many fascinating lectures, and for challenging me to find the principles behind my inventions.

Dr. Rakesh, for making me write and understand mathematical proofs until they all look trivial.

Chris Thomas, for a most enduring friendship.

My piano teachers Beverly Stephenson and Ruth Anne Rich, for inspiring me with what the hands can do on a properly responsive instrument.

My typing assistants, Sarah Ruth Budd, Sara Levin, Mark Parsia, Denise Lemon, Barbara Westog, my sister Ellen, and my mother Bessie, for helping me to continue to program and publish for the past four years as I perfected less fatiguing methods for data entry.

Samuel Audet, for generously writing HotScroll, OS/2's only continuous scrolling software.

Brian Hall at Microedge, Inc., for adding a keystroke-saving variable-name-completion feature to Visual Slickedit at my bequest.

My fellow residents of Lovett Graduate House, for sharing the television when my mind was too weary for anything else.

I am so lucky to have such a loving, patient, and stable family, who always welcome me home twice a year even though I moved so far away. I thank my father for drawing me back to the farm for refreshing manual labor yet giving me time to develop one crazy project after another on my vacations, and also for enforcing the pragmatism and ethics of the frontier. I thank my mother for spicing my vacation diet with wholesome home-grown foods and swimming. I thank my grandmother Edna and her family for bringing poetry, history, and gentleness to my summers; may she always whisper from above how her family almost got transplanted to California. I thank my grandfather Walt for introducing electricity to our home town with only a fifth grade education; that was only the beginning.

This work was partially funded by a National Science Foundation Graduate Fellowship for Wayne Westerman.

This manuscript is dedicated to:

My mother, Bessie,

who taught herself to fight chronic pain in numerous and clever ways,

and taught me to do the same.

TABLE OF CONTENTS

Chapter 2

PROXIMITY IMAGE FORMATION AND TOPOLOGY

Limited hand and finger tracking experiments have previously been conducted with a variety of sensing technologies. This chapter begins with a review of these sensing technologies and explains why proximity sensing arrays are particularly well-suited for everyday applications of hand tracking. Then the chapter discusses proximity image pre-processing such as background object removal, sensor offset adaptation, and electrical noise filtering. The chapter concludes with a sampling of proximity images which illustrate the typical features and arrangements of hand contacts. This hand topology section is particularly important to the understanding of the contact segmentation and identification algorithms in Chapters 3 and 4, which rely heavily on relative contact shape and position constraints.

2.1 Related Methods for Hand Motion Sensing

Hand position and motion can conceivably be detected with mechanical or electromagnetic sensors attached to the hand, with remote optical or acoustical sensors, or with proximity or pressure sensors mounted on an object in the user's environment. At first glance the attached sensor methods seem advantageous because they can capture three-dimensional hand activity in free space, unconstrained by the physical form factor of an interfacing object. Data gloves and computer vision systems have been popular in virtual reality experiments for this reason. Such systems are clearly appropriate for capturing the free-space hand gestures and sign language as they appear in communication between humans, but several factors make them impractical for everyday human-computer interaction.

2.1.1 Free-Space Gestures

The first problem lies with holding or slowly adjusting hand position in free space. The quick, relative motions of sign language may be easy to perform, but holding the unsupported hands out in front of the body for extended periods is very tiring [152, 153]. In such postures fingertip positions are also somewhat unstable, so considerably less precision is possible than when some part of the hand or arm rests against a firm object. Also, it is very difficult for a computer to distinguish motions intended to be instructions for the computer from postural adjustments or gestures to co-workers. This is known as the gesture saliency problem. To appreciate the difficulty of this problem, consider how often we humans mistakenly think someone is gesturing at us when the gesture is actually intended for someone behind us or no one at all. If the direction of gaze of the sender is not known, determining the intended recipient of gestures is even more troublesome.

2.1.2 Data Gloves

Free-space motion sensing technologies have limitations as well. Though DataGloves [148] can potentially capture the entire range of finger flexion and extension, in practice the flexion sensors are imprecise yet expensive and cumbersome to wear. Furthermore, as a bodily attachment, gloves must often be removed when the user resumes non-computer tasks. This is both a practical disadvantage and an ergonomic disadvantage because it discourages users from taking rest breaks and mixing in non-computer tasks which rely on other muscle groups. FakeSpace, Inc. [36] markets *pinch* or *chord gloves* for virtual reality systems which detect contact between electrically conducting fingertip pads rather than general flexion and extension of the fingers. The lack of flexion sensors reduces cost, and consistent with the design philosophy of this dissertation, such physical fingertip contact turns out to be more reliable and easier to learn than free-space finger motion gestures [60].

2.1.3 Video Gesture Recognition

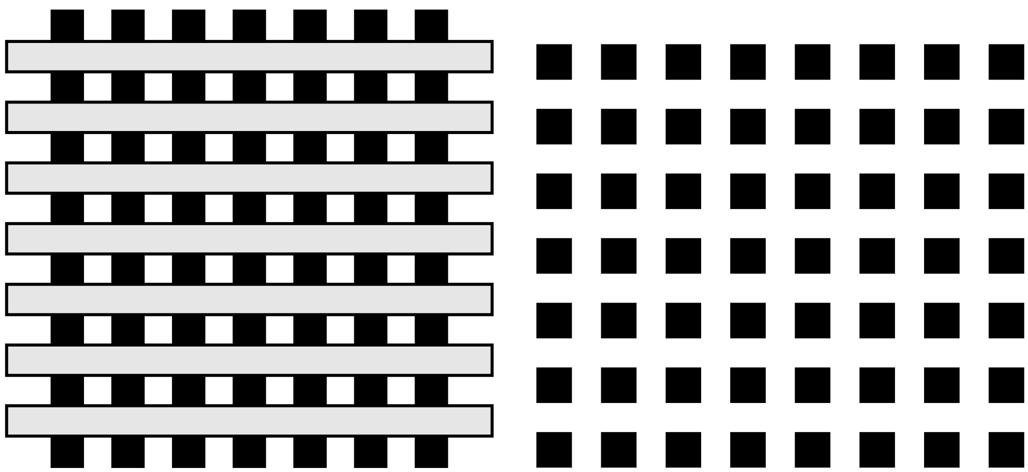
Computer vision technologies avoid the encumbrance of wearing gloves but cannot always infer fingertip location. Assuming decent lighting is available, much of the luminosity information that a video camera supplies is unnecessary for finger tracking, and must be filtered out with computationally intensive algorithms [115]. The body of the hand can occlude the fingertips at some camera and hand angles. Occlusion and limited camera resolution also make it very difficult to determine exactly when the fingers touch a surface.

2.1.4 Benefits of Surface Contact

Most importantly, the emphasis on hand tracking in three-dimensional free space ignores the long history of manipulating hand tools and musical instruments which provide rich haptic feedback as the tool is acquired. While economics may preclude customizing the shapes of general-purpose input devices as much as hand tools are customized, detection of contact with a physical surface provides, at the bare minimum, a clear demarcation between motions on the surface that the computer is intended to recognize and motions away from the surface that the computer should ignore. Though individual finger activity on a surface is constrained to two-and-a-half dimensions, Chapter 5 will demonstrate that extra degrees of freedom can be extracted from rotational and scaling motions of multiple fingers on a surface. For many applications the improved clarity of user intent and tactile feedback that surface contact imparts will more than make up for the slight reduction in movement freedom.

2.1.5 Sensing Finger Presence

Technologies which have been applied to detecting finger or stylus contact include resistive membranes, surface acoustic wave, active optics and finger capacitance sensing (see Lee's 1984 Master's Thesis [88] for an early review). Most implementations are limited to unambiguous location of a single finger because they rely on what Lee calls "projective" sensor matrices. In a projective matrix (Figure 2.1a), one sensor element is allocated to each row and column at the edge of the active



a) b)

Figure 2.1: The two basic multi-touch proximity sensor arrangements. In a), "projective" row and column spanning sensors integrate across each row and column electrode and only need connections at the edges of the matrix. Touching fingertips can be counted by counting the maxima in the column signals assuming the fingertips lie in a roughly horizontal row unobstructed by thumb or palms. The square sensors in b) only integrate over the local square. The exact locations of any number of fingertip-sized contacts can be interpolated from the 2D array of square sensors, but a connection matrix must be run underneath the sensor array to connect the sensors to signal processing circuitry.

area. Finger presence anywhere along a row will register on that row's sensor, so that a finger affects roughly one row and one column sensor. While the total number of sensors needed is related to only the square root of the active area, multiple finger contacts can confuse these systems [88]. As was true in 1984, the surface acoustic wave and infrared touchscreens as well as capacitive touchpads on the market still suffer from this limitation.

Some devices on the market partially utilize multiple fingers despite the ambiguities of projective sensing. For example, touchpads manufactured by Logitech, Inc. [15, 78] for laptop computers are able to detect the presence of up to three fingertips. The patent to Bisset and Kasser [15] explains that this is done by assuming the fingers lie in a row and counting the number of maxima in the column projection. However, as will be seen in Figures 2.2 and 2.3 below, this projection maxima counting method becomes ambiguous for larger touch surfaces in which one hand part can intersect the same column as another, such as when both fingers and palms touch the sensing area or the hand rotates so fingers lie diagonally or in a column.

Figures 2.2 and 2.3 demonstrate the limitations of this projection approach compared to the two-dimensional arrays of sensors (Figure 2.1b) to be discussed in Section 2.1.7. Fingertip, thumb, and palm heel surface contacts are simulated with two-dimensional Gaussians of varying widths on the 2D square grid. The grid samples the Gaussians at 2.5 mm intervals such as would occur in a capacitive sensing array with moderate spatial resolution. The darkness of the squares is proportional to the finger capacitance or proximity sampled at the square. The projective signals which would be measured from the row and column spanning electrodes of Bisset and Kasser [15] are simulated by integrating over each row of the 2D array to obtain the horizontal bar plots to the left of each grid and by integrating over each column to obtain the vertical bar plots under each grid.

Figure 2.2 shows the projection sensing ambiguities which can occur when the fingertip row is not horizontal, but lies diagonally instead due to various hand

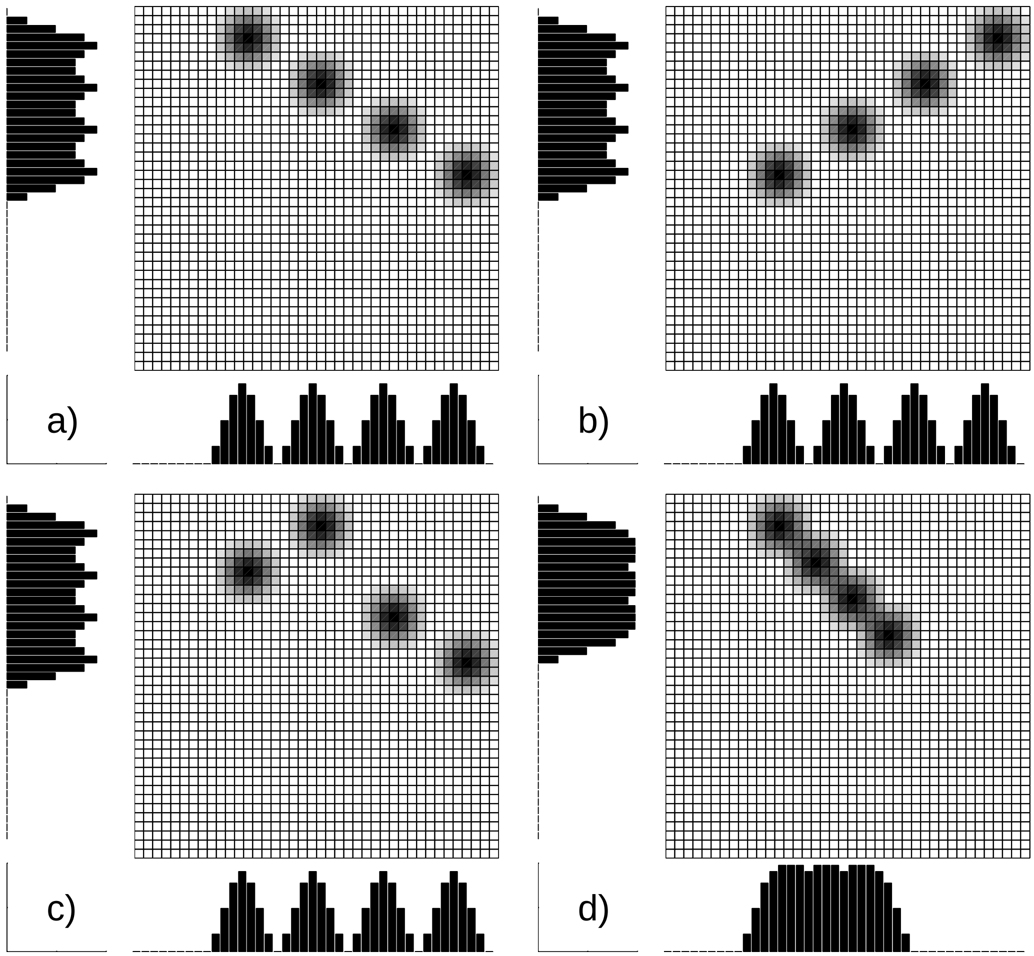


Figure 2.2: Projection sensor ambiguities for various diagonal arrangements of fingertips. The different fingertip contact arrangements shown on the square sensor grid in a)-c) all produce the same row and column projections (horizontal and vertical bar plots), preventing the projection method from determining the hand rotation, though it can still count the fingertip maxima. In d) the fingertips are so close together that the projection minima between fingertips disappear, preventing fingertip counting, though the diagonal minima are still discernable in the square sensor grid.

rotations. In Figure 2.2a-c four maxima appear in both the row and column projections (bar plots), indicating at least four objects are touching the surface, but the projections are the same in each case even though the fingertip arrangements (grid) differ. The same projections could be obtained from a 4×4 array of 16 fingertips also, though most human operators will not have that many fingertips. In Figure 2.2d the fingertips are so close together in their diagonal row that the projection maxima merge, though local maxima are still clearly separated by diagonal partial minima in the sampled 2D array.

Figure 2.3 shows how fingertip counting from projection sensors is occluded by the presence of thumb and palms in a neutral hand position. In Figure 2.3a four fingertips lie in a slight arc, producing four maxima in the column projections and one in the row projection. Figure 2.3b includes the thumb in nearly the same column as the index fingertip, causing an additional maximum in the row projection (horizontal bars) only. The index fingertip is removed in Figure 2.3c; because the thumb is still in the same columns, the number of projection maxima does not change, though the amplitudes change somewhat. Because the amplitudes also depend on how lightly each finger touches the surface, the change in projection amplitudes cannot reliably resolve this ambiguity; the amplitude changes could also be a result of a lightening in hand pressure. In Figure 2.3d the palms touch as well, leaving three maxima in the row projection but causing the column projection maxima to merge into just two. Therefore from the row projection one could surmise that some palms, the thumb, and some fingertips are touching, but one can no longer tell how many fingertips are touching because the palm column projections get integrated with and obscure the fingertip column signals.

As Lee points out, measuring projections from additional angles such as diagonals can help disambiguate multiple contacts, as is done in tomography systems,

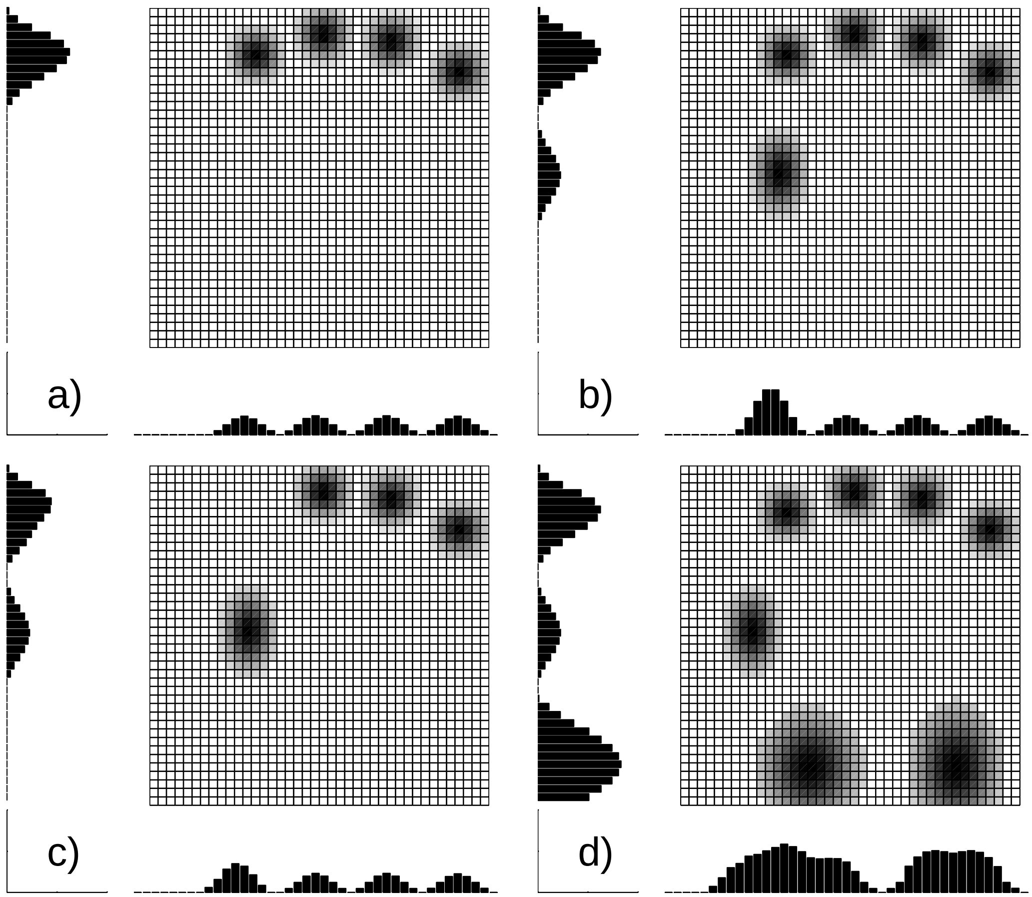


Figure 2.3: Ambiguities in projective sensing caused by presence of the thumb and palms in the same columns as fingertips. a) simply contains a slightly arched row of fingertips producing four column projection maxima (vertical bars at bottom) and one row projection maximum in the horizontal bars. Adding a thumb contact in b) adds a row maximum but not a column maximum because the thumb intersects nearly the same columns as the index fingertip. Removing the index fingertip in c) does not chance the number of projection maxima, meaning fingertips cannot be counted reliably in the presence of the thumb. Adding the palms in d) further obscures the fingertip row projection maxima, which get merged with those of the palms.

but details inside concave contacts will still be undetectable [88]. The number of unambiguously locatable contacts is generally one less than the number of projection angles utilized [88]. McAvinney's "Sensor Frame" [107, 108, 129], an attachment to the screen of a computer monitor which senses intersection of fingers with infrared beams from four directions, utilizes this tomography approach to unambiguously locate up to three fingers.

2.1.6 Tactile Imaging

This complex tomography approach can be avoided with a regular two-dimensional array of individually addressable sensors (Figure 2.1b), in which each sensor corresponds to a pixel in a "tactile image." Layered resistive-membrane pressure sensors can be constructed economically in this configuration, but their substantial activation force is ergonomically inferior to zero-activation-force proximity sensing. Another approach is to place a camera under a translucent tabletop and image the shadow of the hands [81, 110]. Unfortunately the bulky optics under the table will limit portability and leg room, and such systems cannot differentiate finger pressure [88]. Active optical imaging with an array of infrared transmitters and receivers on the surface could easily detect finger proximity, but would be prohibitively expensive and power consumptive.

2.1.7 Capacitance-Sensing Electrode Arrays

The remaining option is to measure the capacitance between the fingers and an insulated array of metal electrodes. The presence of a finger effectively increases the electrode capacitance to ground since the capacitance between the conductive fingertip flesh and an electrode plate is typically a few pF but the capacitance of the human body with respect to earth ground is relatively large (about 100pF) [88]. Since the capacitance between parallel plates drops quickly in inverse proportion to the distance between the plates, this technique can only detect fingers within a few millimeters of the electrodes. Spatial resolution increases dramatically as the fingers approach the electrodes. Precision of .2 mm can easily be obtained with 4 mm electrode spacings by computing a finger centroid, *i.e.*, interpolating between neighboring electrodes. The capacitive technique also indicates finger force up to a couple Newtons because the effective capacitor area increases as the fingertip pulp flattens against the surface [134]. While the limited proximity sensing range of electrode arrays ensures fingertip proximity information is clear and uncluttered, it also prevents detection of the finger joints and palms unless the whole hand is flattened against the surface.

Lee built the first such array in 1984 with 7mm by 4mm metal electrodes arranged in 32 rows and 64 columns. The "Fast Multiple-Touch-Sensitive Input Device (FMTSID)" total active area measured 12" by 16", with a .075mm Mylar dielectric to insulate fingers from electrodes. Each electrode had one diode connected to a row charging line and a second diode connected to a column discharging line. Electrode capacitance changes were measured singly or in rectangular groups by raising the voltage on one or more row lines, selectively charging the electrodes in those rows, and then timing the discharge of selected columns to ground through a discharge resistor. The principal disadvantage of Lee's design was that the column diode reverse bias capacitances allowed interference between electrodes in the same column. Even with 2048 electrodes and suitable interpolation between electrodes, the electrode spacing was probably too coarse to reproduce the fine mouse positioning achieved with current single-finger touchpads [46 48, 50, 51, 111]. Though its scanning rate depended irregularly on the number of and positions of surface contacts, for ten fingers it would have only been able to achieve 1-5 fps, which is much too slow for either typing or gesture applications.

Rubine [129, 130] reports seeing another multi-touch tablet demonstrated at AT&T in 1988 by Robert Boie which could detect all ten fingers. It boasted a 30 fps frame rate and resolution of 1 mil (.025 mm) in lateral position and 10 bits in pressure. Possibly it measured sensor capacitance with the synchronous detection technology in a 1995 patent by Boie *et al.* [17] that briefly mentions multi-touch tablets as an application.

2.1.8 The MTS's Parallelogram Electrode Array

The MTS contains a 16×96 electrode array (Figure 2.4) much like those in the above multi-touch tablets. It employs a special wedge electrode geometry to reduce the number of rows necessary by a factor of three without causing serious non-uniformities in vertical position interpolation. This reduction in electrode count speeds fabrication of research prototype arrays by lowering the discrete part count, but would not necessarily be beneficial for volume manufacturing techniques.

Rectangular electrodes (Figure 2.5) like those used by Lee [88] are more sensitive to vertical position changes near the top and bottom of the electrodes, where it is possible to interpolate between two electrodes, than in the middle of an electrode. If a finger is in the middle, the electrode is so tall that the electrodes above and below do not register enough signal to get a reliable interpolation.

In contrast, the vertically interleaved parallelogram electrodes interpolate via their physical geometry. The ratio of the horizontal cross-sections between electrodes in a column varies continuously with vertical location of an object (Figure 2.6a-d)). Though this improves uniformity of vertical interpolation compared to rectangular electrodes of the similar height, it also has the effect of vertically smearing signals, making it difficult to distinguish objects which appear in the same electrode column less than one row spacing apart. For research prototyping purposes this is tolerable because the fingers tend to lie in a row, no more than one per column. However, once in awhile the thumb or pinky pass behind and intersect columns of the other fingertips, becoming indistinguishable from the fingertip in front of them (see Section 2.3.3). Also, as is discussed in Appendix B, vertical interpolation biases do arise

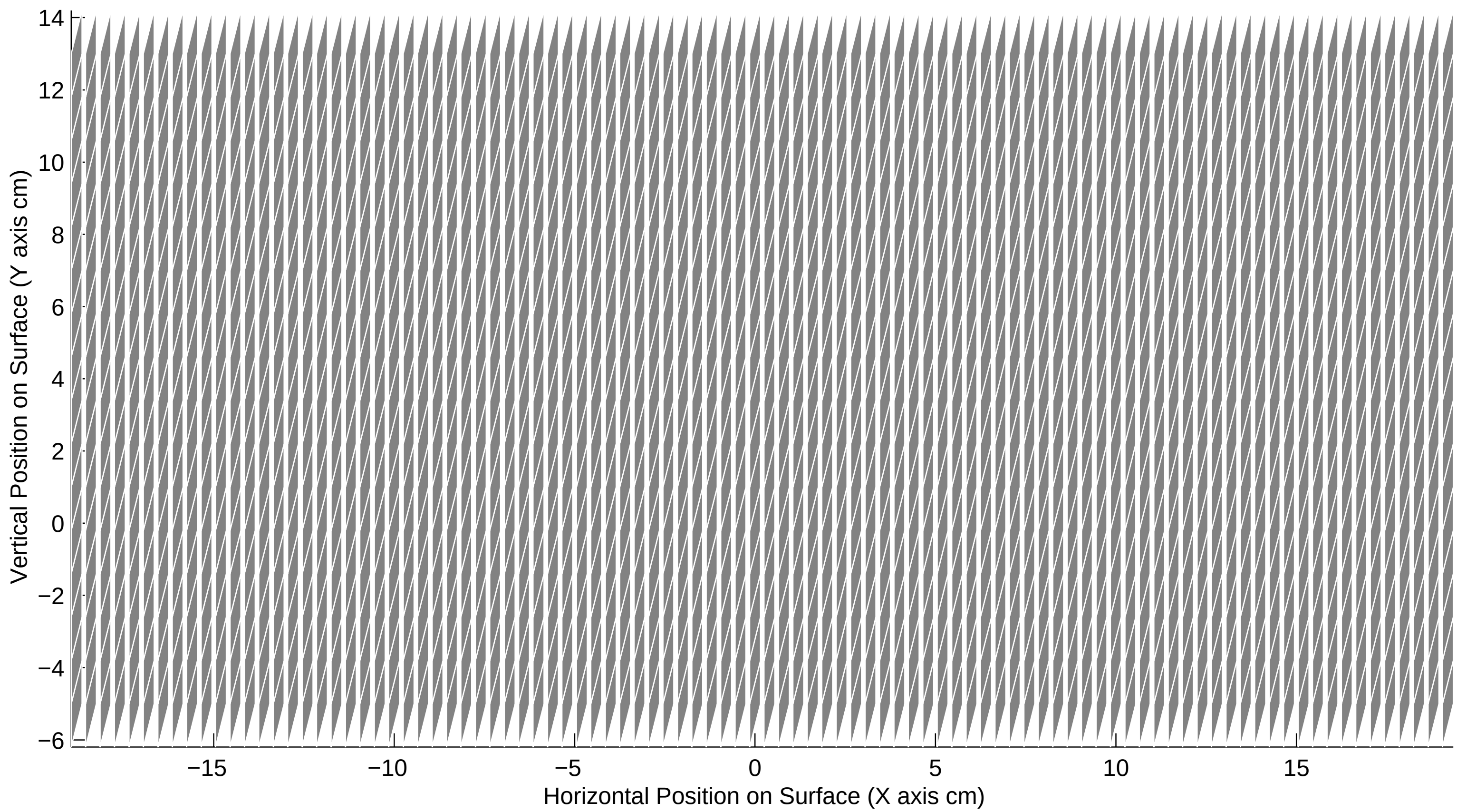


Figure 2.4: Diagram of electrode layout for the entire 16×96 parallelogram electrode array. Row pitch is 1.2 cm and column pitch is 0.4 cm, but electrodes are only 0.25 cm wide.

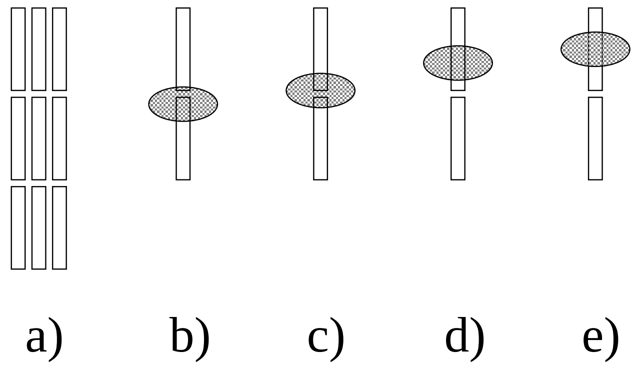


Figure 2.5: A 3×3 section a) of rectangular electrode array. Vertical interpolation between top and bottom electrodes works in b)-c) but not in d)-e).

for small contacts which are not centered on or between columns of the parallelogram electrode array. Thus a commercial product, especially one which attempts to recognize a handwriting grip or stylus, would have to abandon the electrode count savings of this scheme for traditional square electrodes and a smaller row spacing.

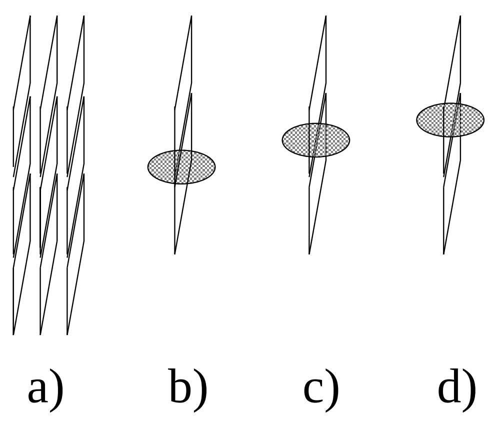


Figure 2.6: Vertical interpolation on the parallelogram electrode array is uniform in a)-d) since ratio of hatched cross sections on top and bottom electrodes changes gradually.

2.1.9 No Motion Blur on MTS

Another important characteristic of the MTS is that the sensing array multiplexes much of the integration, buffering and quantization circuitry. Therefore the capacitance of each electrode is measured over a relatively short period of a few hundred microseconds compared to the total array scanning period of ten to twenty milliseconds. This contrasts with the CCD arrays typically used in video cameras which integrate incoming photons at each pixel over most of the period between readouts. An advantage of the MTS's relatively short integration time is that MTS proximity images do not exhibit motion blur. However, if the scanning rate is not fast enough, quick finger taps over an electrode can occur entirely between measurements of that electrode and be completely missed. When tapping key regions during touch typing, fingers usually remain on the surface for at least 50 ms, but the scan period must be somewhat smaller than this for reliable detection. During the experiments conducted for this dissertation, the array scan frequency or frame rate has been set to 50 fps (corresponding to a period of 20 ms), which ensures that each finger tap shows up in at least one scan. However, at this rate the peak finger pressure as the fingertip bottoms out onto the surface in the middle of the tap cannot be measured accurately because the single scan detecting the tap might occur near the beginning or end of the tap cycle when the finger is barely touching the surface. Minor changes to the scanning hardware can easily push the frame rate to 100 fps, which will allow peak finger pressure to be measured fairly accurately even for extremely quick taps.

2.2 Tactile Image Formation and Background Removal

While designing a tactile sensor array for robotic fingertips nearly 20 years ago, Danny Hillis [59] realized how much easier touch imaging is than computer vision:

> ... analyzing a tactile image is like analyzing a visual image with controlled background, illumination, and point of view ... the properties that we actually measure are very close, in kind, to the properties that we wish to infer.

Comparing background segmentation techniques in vision-based and tactile hand imaging systems will verify his insight.

2.2.1 Optical Image Segmentation

Ahmad's real-time 3D hand tracker [3] segments the background by matching image patches to known skin color histograms, but to keep up with frame rates (30 frames per second) it must limit the skin search region and adaptively subsample the image. Finger positions are obtained by fitting ellipses to the segmented hand patches. The total hand patch area weighted with a centered Gaussian roughly indicates the distance between hand and camera. Ahmad also tries to recover finger joint angles, information which data gloves give directly, by finding fingertips and learning an inverse mapping from fingertip and palm position to intermediate joint angle. This feature of the tracker becomes unstable due to fingertip detection failure if the hand is not roughly normal to the camera.

The Digital Desk [154 157] is a system pioneered at Xerox for combining interaction with paper and digital documents. The system contains both a computer screen projector and zoomable cameras mounted high above the user's desk. The cameras both track hands and recognize text from paper documents lying on the desk. Since the vision system cannot determine exactly when fingers actually touch the desk surface, a microphone is placed under the desk to "hear" finger taps and thus emulate mouse clicks. Crowley and Coutaz [30] consider color, correlation tracking, principal components and active contours for following a pointing object on a digital desk. In the correlation method, a previous image of a fingertip is used as a reference template for correlations with the next image. The new finger position is indicated by the amount of template image shift which minimizes the sum of squared differences between template and image. Again, the computational costs of the correlation limit the template search region and thus the maximum trackable finger speed.

2.2.2 Methods for Proximity Image Formation

Background segmentation of proximity images from electrode arrays is much easier because extraneous objects are not expected to be visible in the background. Paper or plastic left over the electrodes do not register on capacitive proximity sensors, nor do small metal objects unless they are deliberately grounded. However, spatial non-uniformities in the parasitic capacitances of discrete components and signal lines may cause background measurements at each electrode to differ. Unlike background signals caused by extraneous external objects, such background non-uniformities are not expected to change over time. A local offset calibration or adaptive thresholding scheme can cancel these fixed sensor disparities. Once these sensor offsets are taken into account and electrical noise is filtered, the proximity image can simply be thresholded to identify regions of fleshy contact. Note that single-finger projective touchpads do utilize offset adaptation but do not have to segment the image into fleshy contact regions; they simply compute a global centroid from measurements of all row and column electrodes.

2.2.2.1 Binary Tree Scanning

Lee's binary tree scanning algorithm [88] combines noise filtering and thresholding in hardware by analog grouping and summation of electrode capacitance measurements. The array is recursively subdivided into rectangular electrode groups of decreasing size via bisection starting with the whole array. Thresholds are calibrated during device initialization for each electrode group at each size, or level, in the recursion. During subsequent scanning, subrectangles are scanned only if the parent rectangle's threshold is exceeded. Once the recursion reaches a measurement which passes threshold at the single electrode level, a finger position is computed as the centroid of the recursed electrode capacitance and its eight neighboring electrode capacitances. Advantages of Lee's scheme are: not every electrode in the array need be separately scanned each pass, and grouping of many electrodes at the beginning of the scan tends to average out noise. The disadvantage is that small, light contacts can be lost among the large electrode groups if the large group thresholds are marginally too high.

2.2.2.2 Brute Array Scanning

Both digital and analog processing speeds have increased enough since Lee's prototype was built that the scanning overhead concerns have become negligible, especially in light of the additional finger tracking and gesture recognition algorithms which the MTS must execute. Keep in mind that though the number of discrete components necessary for an electrode array may make it seem large, the number of "pixels" is still small compared to even a low-resolution digital camera image. For this reason, and to ensure even brief, light finger contacts are captured, the MTS employs a brute force electrode scan to form a complete proximity image before applying standard digital filtering techniques.

2.2.2.3 Sensor Offset Adaptation

Sensor offset calibration will fail during device initialization if the user's hands are already on the board. Since there may not be a time when the fingers are known to be absent, the MTS continuously updates each electrode offset with the minimum of readings from that electrode. Suppose $A_{ij}[n]$ is the raw tactile proximity measured from the electrode at row i, column j during scan cycle n. Then the local offsets O_{ij} can be updated as:

$$O_{ij}[n] = \min(A_{ij}[n], O_{ij}[n-1]) \tag{2.1}$$

The offset-corrected image E is then:

$$E_{ij}[n] = A_{ij}[n] - O_{ij}[n] \; \forall i, j : \; 0 <= i < E_{rows}, 0 <= j < E_{columns} \tag{2.2}$$

Since capacitance measurements always return to baseline when fingers are removed, the offsets will correct themselves by decreasing as soon as fingers are lifted. The danger of this method is that negative electrical noise spikes can cause inadvertent lowering of the offsets. Local offsets which are too low lead to false positive proximity indications, just as offsets which are too high cause finger contacts to be missed. The MTS compromises by decreasing offsets only when at least three low proximities are read consecutively and by allowing very slow recovery, over about a minute, should an offset get lowered too far:

$$O_{ij}[n] = \min(\max(A_{ij}[n], A_{ij}[n-1], A_{ij}[n-2]), (O_{ij}[n-1] + \beta)) \tag{2.3}$$

where the max operation provides immunity to single negative noise spikes and a tiny β gives a slow recovery rate. Even with a tiny β, hands which are left resting on the board a few minutes will appear to fade. To prevent this, β is further decreased for those electrodes which the system confidently identifies as underlying a fleshy contact. These offsets quickly adapt to the minimum baseline capacitance so any readings above the offsets can be modeled as the flesh proximity magnitude plus minor Gaussian background noise.

2.2.2.4 Proximity Image Filtering

While Lee [88] electrically averaged the capacitances of entire rectangular groups of electrodes to combat noise before threshold testing, the MTS electrode array is much less noisy than Lee's device. Furthermore, to take full advantage of the electrode array resolution, groups should conform to finger contact shape electrode by electrode rather than be constrained to rectangular groups which poorly fit the oval shape of most hand contacts. Therefore, the MTS only employs slight spatial diffusion of each offset-corrected image to combat electrical noise. Then it applies significance threshold and local maximum tests to each diffused pixel to detect the center of each hand contact, as further described in Chapter 3.

2.3 Topology of Hand Proximity Images

To illustrate typical properties of hand contacts as they appear in proximity images, Figures 2.7 2.10 contain sample images captured by the prototype array of parallelogram-shaped electrodes. Shading of each electrode darkens to indicate heightened proximity signals as flesh gets closer to the surface, compresses against the surface due to hand pressure, and overlaps the parallelogram more completely. Notice that the proximity images are totally uncluttered by background objects; unlike optical images, only conductive objects within a couple millimeters of the surface show up at all. Background sensor offsets have already been removed from each image, and background electrical noise levels are so low as to not be visible with the given grayscale intensity map. Certain applications such as handwriting recognition will clearly require finer electrode arrays than indicated by the electrode size in these sample images. In the discussion that follows, the proximity data measured at one electrode during a particular scan cycle constitutes one "pixel" of the proximity image captured in that scan cycle.

In this section and the rest of this dissertation, the term "proximity" will only be used in reference to the distance or pressure between a hand part and the surface, not in reference to the distance between adjacent fingers. "Horizontal" and "vertical" refer to x and y directional axes within the surface plane. Proximity measurements are then interpreted as pressure in a z axis normal to the surface. The direction "inner" means toward the thumb of a given hand, and the direction "outer" means towards the pinky finger of a given hand. For the purposes of this description, the thumb is considered a finger unless otherwise noted, but it does not count as a fingertip. "Contact" is used as a general term for a hand part when it touches the surface and appears in the current proximity image, and for the group and path data structures which will represent it in Chapter 3.

2.3.1 Flattened Hand Image Properties

Figure 2.7 shows a right hand flattened against the surface with fingers outstretched. This flattened hand image includes all of the hand parts which can touch the surface from the bottom of one hand, but in many instances only a few of these parts will be touching the surface, and the fingertips may roam widely in relation to the palms as fingers are flexed and extended. At the far left is the oblong thumb which tends to slant at about 120°.

The columnar blobs arranged in an arc across the top of the image are the index finger, middle finger, ring finger and pinky finger. Since the fingers are fully extended, the creases at finger joints cause slight undulations in proximity along each column, though smearing by the parallelogram electrodes obscures this effect somewhat. Flesh from the proximal finger joints, or proximal phalanges, appears as the particularly intense undulations at the bottom of the index, middle, and ring finger columns. Since the fingers are fully flattened, flesh from the forepalm calluses is also visible as small clusters below the proximal phalanges, near the vertical level of the thumb.

The inner and outer palm heels cause the pair of very large contacts across the bottom of the image. These palm heels tend to be quite large, mildly oblong, and oriented diagonally. Unless the center of the palm is intentionally pushed against the surface, a large crease or proximity valley clearly separates the inner and outer palm heels. Even though image resolution is fairly low, it is clear that the fleshy contacts from different parts of the hand have subtly contrasting geometric properties. All the hand contacts are roughly oval-shaped, but they differ in pressure, size, orientation, eccentricity and spacing relative to one another.

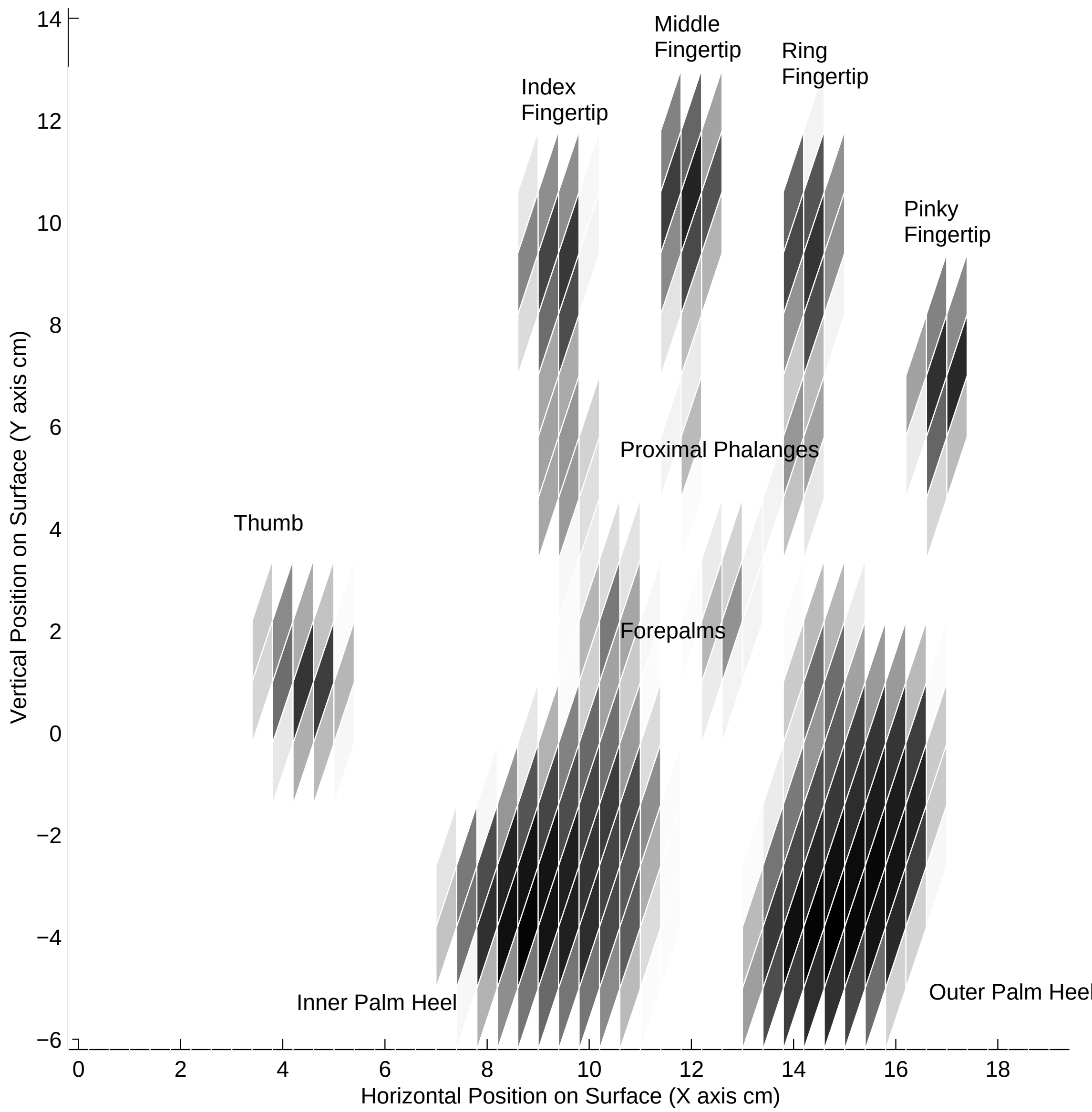


Figure 2.7: Offset-corrected proximity image of right hand flattened onto the surface with fingers outstretched and all hand parts labeled.

2.3.2 Properties of Hands in the Neutral Posture

Figure 2.8 shows a proximity image for all fingers and palms of both hands

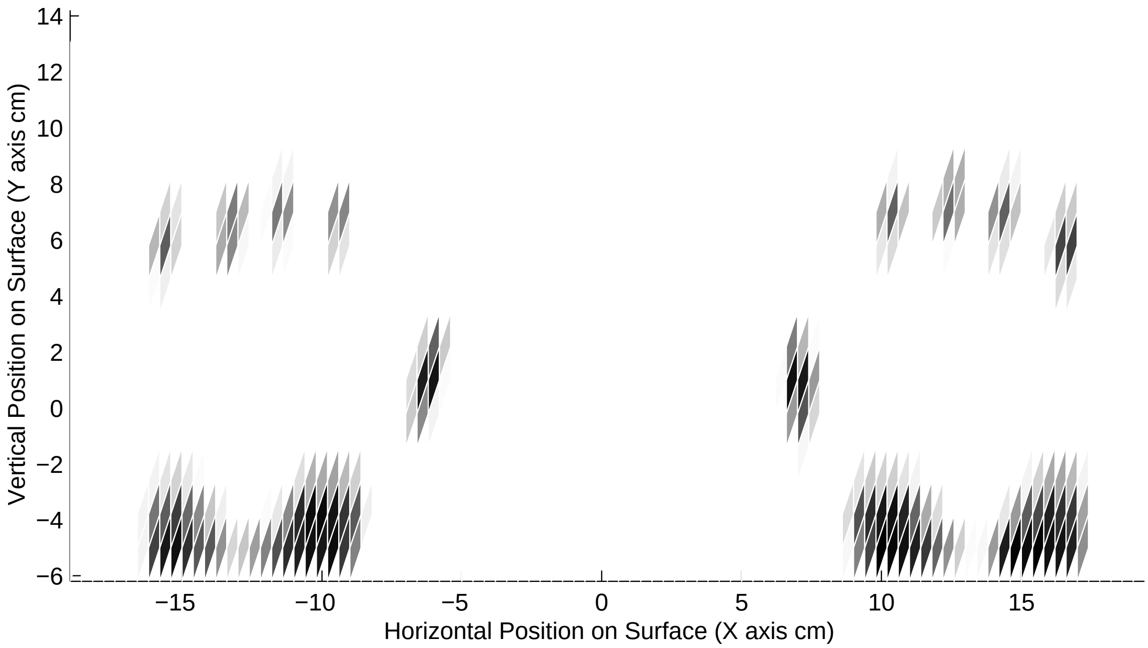


Figure 2.8: Proximity image of both hands resting on the surface in their respective neutral or default postures.

resting in what will be known hereafter as their default positions. Since these positions correspond to the most neutral hand and finger postures, with wrist straight and fingers curled so fingernails are normal to the surface, gestures are most likely to start from this hand configuration. Note that since fingers are curled, the proximal phalanges and forepalms are far above the surface and not visible. Because the fingers are slightly spread in this neutral posture, all fleshy contacts are clearly separated by at least one electrode at the background or zero proximity level. Since only the tips rather than the lengths of the fingers are visible, the fingers appear much shorter than in Figure 2.7, and would appear circular if not for vertical smearing by the parallelogram electrodes. However, the finger widths remain fairly constant regardless of contact elongation. Also, the electrodes at the center of each fingertip do not appear as dark as the central thumb and palm heel electrodes because, in this case, the fingertips contacts are not tall enough to fully overlap any of the parallelograms, limiting the proximity signal regardless of their distance from the surface. The palm heels appear somewhat shorter than in Figure 2.7 since only the rear of the palm can touch the surface when fingers are flexed, but the separation between the palm heels is unchanged.

The fact that the intermediate finger joints connecting fingertips to palms, *i.e.*, the lengths of the fingers, do not appear in this commonly occurring proximity image has further consequences. While such lack of intermediate hand structure simplifies determination of the fingertip centroid, it is also the main shortcoming of capacitive proximity sensing in terms of hand gesture recognition. Reliably establishing finger or even hand identity when intervening hand structure is missing from the proximity images poses the most challenging problem of the work described in this dissertation. This challenge is the subject of Chapter 4.

2.3.3 Partially Closed Hand Image Properties

For a tracking system to support a wide range of hand gestures, it must tolerate contact shapes and juxtapositions which vary from the default. The two extremes to be considered in this work are the previously discussed flattened hand and the partially closed hand shown in Figure 2.9. Here the thumb is pushed directly behind the index finger, but vertical smearing by the wedge electrodes may cause thumb and index finger to appear as a single unseparable contact. Unlike the default hand posture in Figure 2.8, adjacent fingertips are so close together as to be distinguishable only by slight proximity valleys or saddle points between them. At the given horizontal electrode spacing, the saddle points between adjacent fingertips may only be separated by a single column wide. Any segmentation algorithm must use the partial minima in the horizontal direction to distinguish these fingertips. In

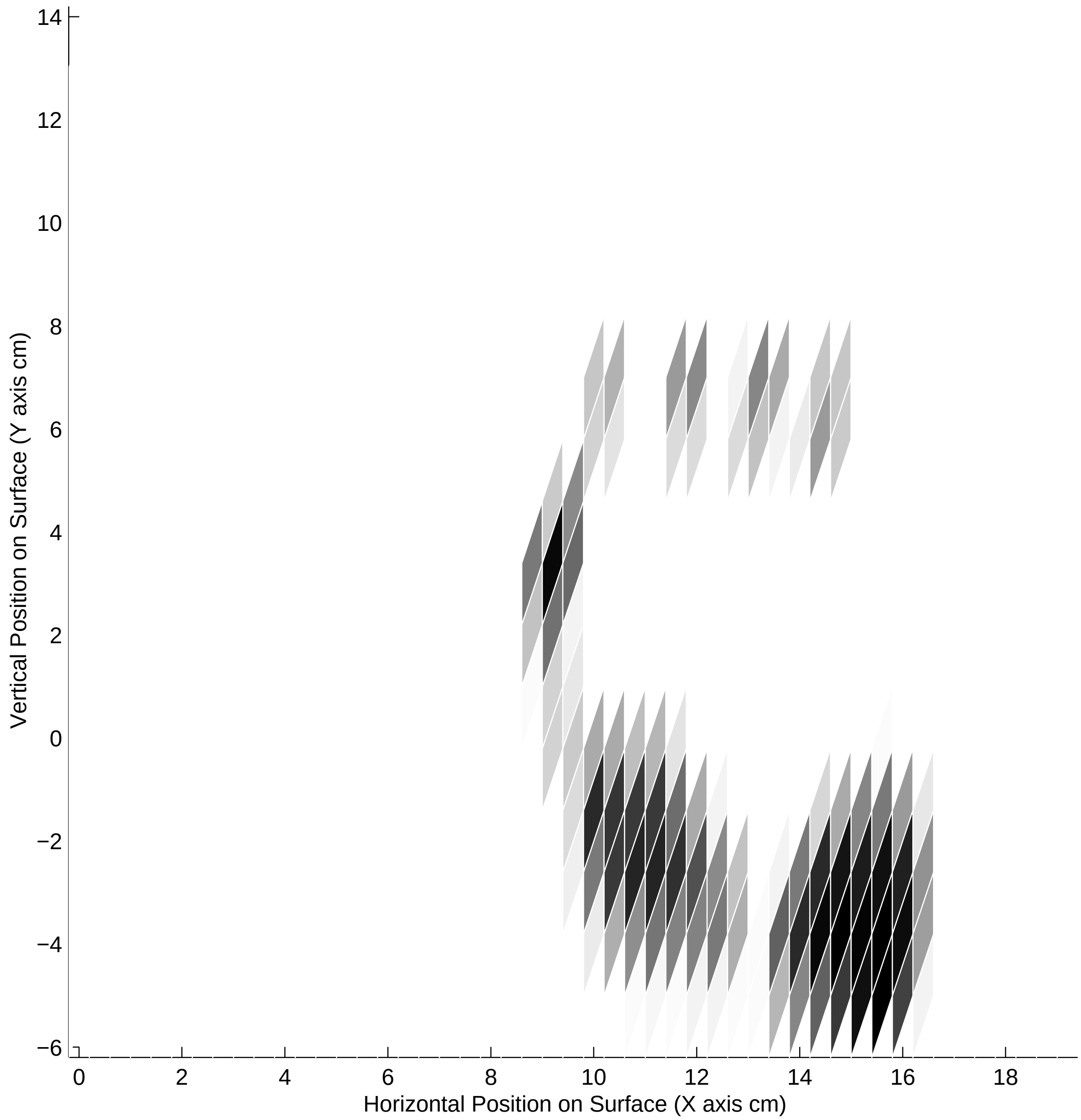


Figure 2.9: Proximity image of a partially closed hand with fingertips squished together.

case the fingertip row is rotated, partial minima in diagonal directions must also be detected. This conflicts with the segmentation needs of palms, which may contain spurious partial minima due to minor variations in sensor gain or flesh proximity across their large areas. All partial minima within palm contacts should be ignored except the large crease between the palm heels.

2.3.4 Pen Grip Image Properties

Figure 2.10 is a proximity image of a right hand in a pen grip configuration, which is particularly comfortable and dexterous for handwriting or freehand drawing. The thumb and index fingertip are pinched together as if they were holding a pen, but in this case they are touching the surface instead. Actually the thumb and index finger appear the same here as in Figure 2.9. However, the middle, ring, and pinky fingers are curled under as if making a fist, so the knuckles from the top of the fingers actually touch the surface instead of the finger tips. The curling under of the knuckles actually places them behind the pinched thumb and index fingertip, very close to the palm heels. The knuckles also appear larger than the curled fingertips of Figure 2.9 but the same size as the flattened fingertips in Figure 2.7. These differences in size and arrangement are sufficient to distinguish the pen grip configuration from the closed and flattened hand configurations. Though the contact segmentation and identification methods presented in this dissertation extend to the pen grip configuration with minimal modification, a higher resolution sensor array without vertically smearing parallelogram electrodes is needed to accurately discern the pinched fingers.

2.3.5 Comfortable Ranges of Hand Motion

Given that the MTS prototype has the form factor of a standard computer keyboard and is similarly placed on a desk, lap or workbench to operate from a sitting or standing posture, the ranges of hand position and rotation expected during

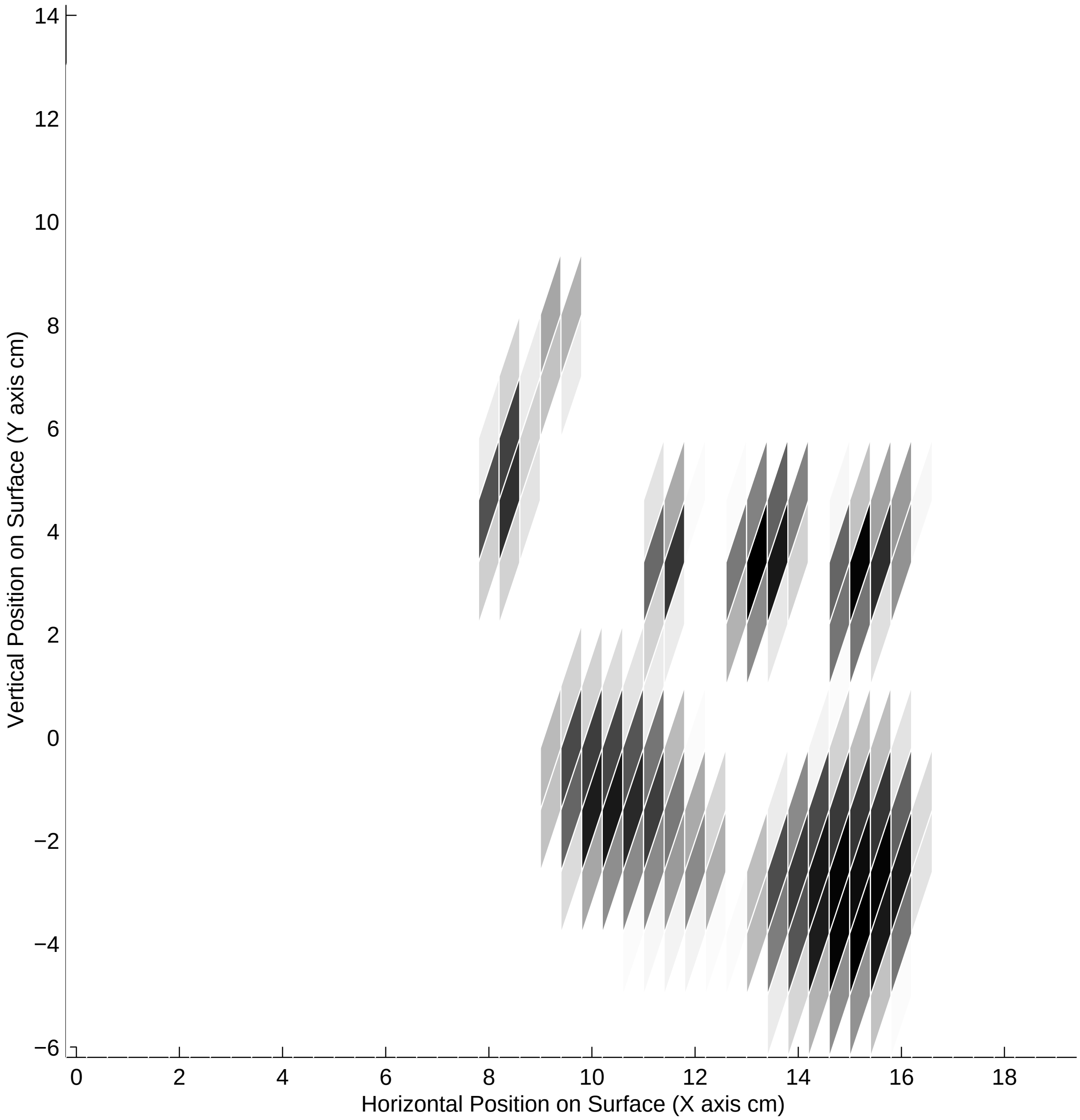


Figure 2.10: Proximity image of a hand with inner fingers pinched and outer fingers curled under towards the palm heels as if gripping a pen.

normal operation are fairly limited. When only one hand is on the surface, its maximum inward rotation can occur when it crosses to the opposite side of the surface, as shown in Figure 2.11. This situation maximizes the inward rotation of both the forearm about the elbow and the hand about the wrist. The maximum

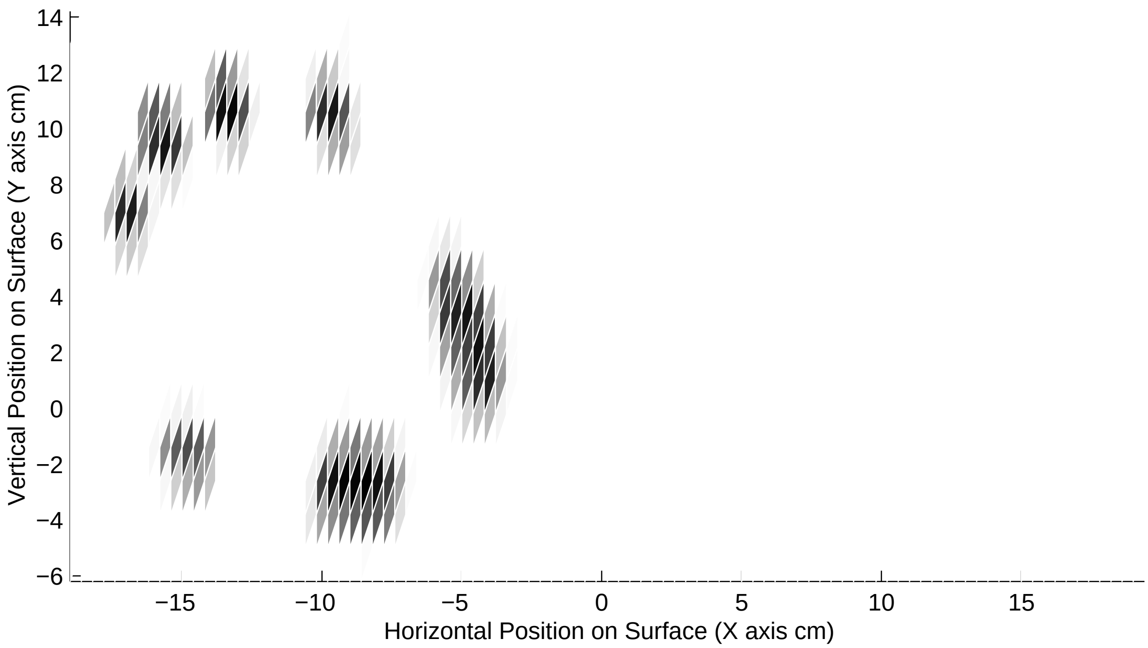


Figure 2.11: Proximity image of right hand at far left of sensing surface and rotated counter-clockwise to its biomechanical limit.

clockwise or outward rotation occurs from the default hand position with forearm parallel to the vertical surface axis, as shown for the right hand in Figure 2.12. Further rotations are only possible through contortions of the whole body or if the operator's torso is not facing the apparatus.

When both hands are on the surface, hand position is even further limited by the fact that operators are not expected to let the hands cross over or overlap one another. Figure 2.13 shows the maximum leftward position of the right hand when the left hand is in its default position. For some operations only part of a hand may remain in the active sensing area, as shown for the row of right hand fingertips at

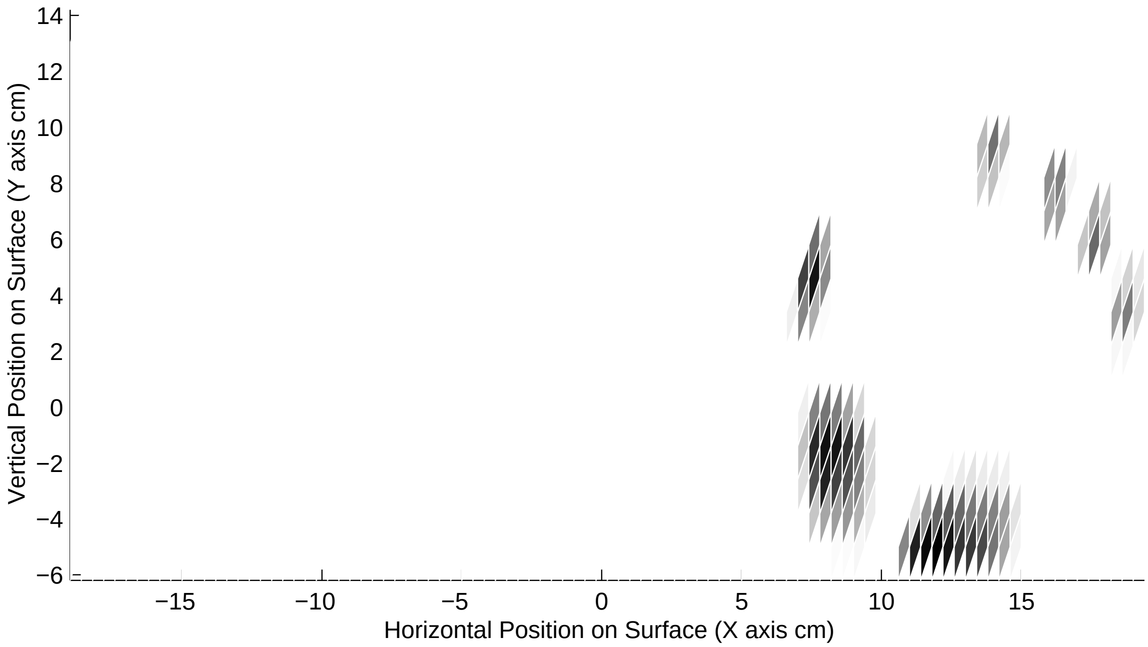


Figure 2.12: Proximity image of right hand at far right of sensing surface and rotated outward to its biomechanical limit.

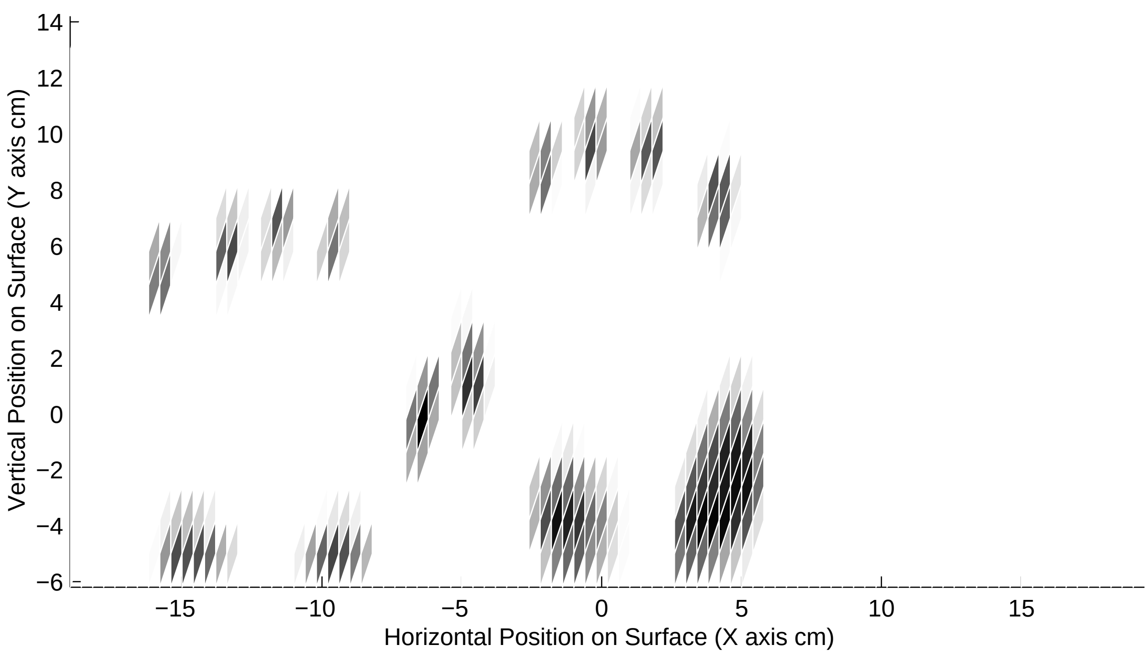


Figure 2.13: Proximity image of left hand in default position and right hand up against it.

the bottom middle of the surface in Figure 2.14. Though it is hard to imagine how

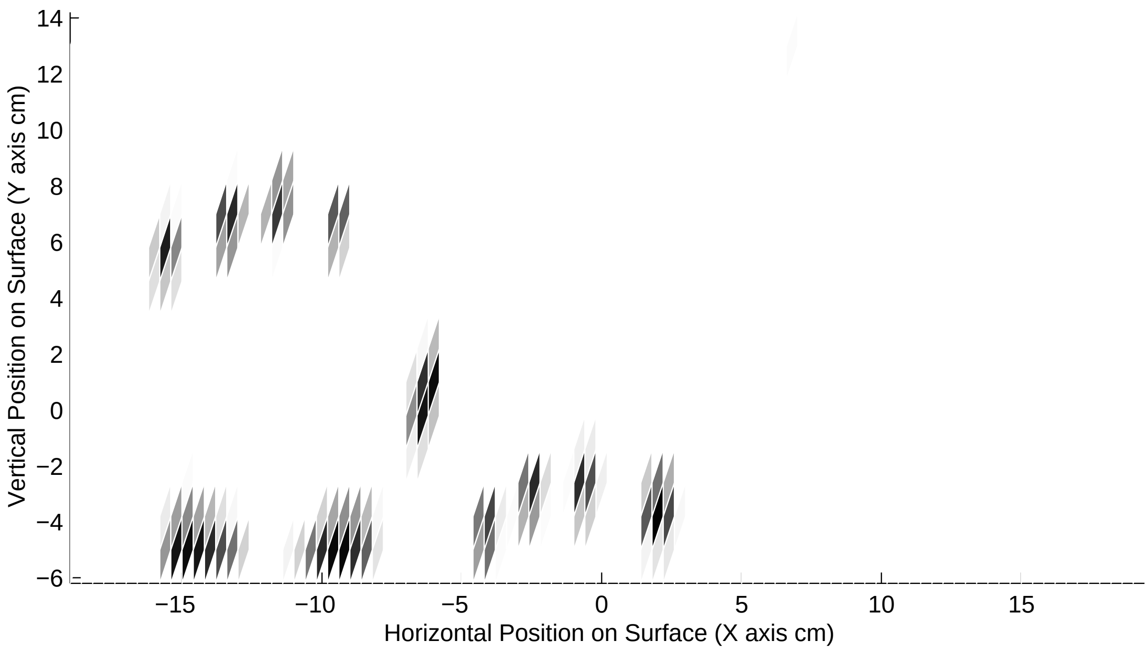


Figure 2.14: Proximity image of left hand in default position and right hand moved down so only fingertips remain in active sensing area.

this would be useful, the fingertips can also lie over the top of the active sensing area as in Figure 2.15, so only the thumb and palms remain visible.

2.4 Conclusion

Capacitance-based proximity sensing has many advantages over other hand motion sensing techniques. These advantages include precise detection of flesh contact with a surface, zero-force activation, avoidance of mechanical encumbrances, prevention of fingertip occlusion, and absence of background scene clutter. An array of a few thousand electrodes is sufficient to detect and uniquely determine the positions of any number of contacts from the undersides of both hands. Though each electrode has a constant sensor offset which must be removed, a large MTS can have signal-to-noise ratios as high as its tiny touchpad cousins.

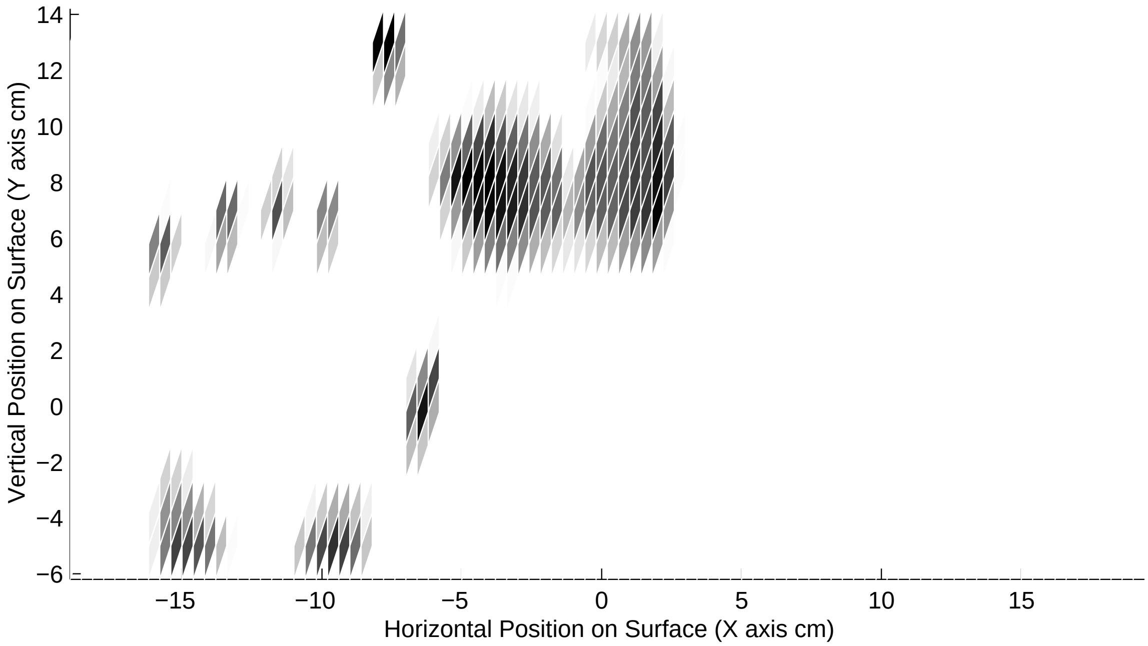


Figure 2.15: Proximity image of left hand in default position and right hand moved up so only thumb and palms remain in active sensing area.

The MTS offers a previously unexplored compromise between the rich tactile and force feedback of a mechanical keyboard or joystick and the feedback void of free space hand gestures. The proximity signals measured by the MTS correspond almost exactly to the operator's own sensations of engaging and sliding the hand across the surface. Even though hand proximity images contain ambiguities due to the lack of sharp edges between flesh contacts and the absence of intervening hand structure, the results of Chapters 3 and 4 will show that these ambiguities are surmountable. Ultimately such a unique, close correspondence between the sensations of the operator and the proximity imaging system can support much faster and more accurate gesture recognition than video-based systems.