Vision and Visual Interfaces
ABSTRACT

The presentation of information as a large display and the use of such displays to support collaboration in face-to-face activities have long been commonplace. Computationally enhanced displays relying on the form factor of whiteboards, surfaces, tables, benches, and desks now afford face-to-face computer-supported interaction and collaboration not possible with classical desktop or mobile computing. This chapter provides an introduction to research and developments in multitouch input technologies that can be used to realize large interactive tabletop or “surface user interfaces.” Such hardware systems, along with supporting software, allow for applications that can be controlled through direct touch or multitouch. Further, a review of gestural interactions and design guidelines for surface user interface design for collaboration is also provided.
Key words: tabletop, HCI, surface user interface, gestural interface, design guidelines.

1.1 INTRODUCTION

People have been using large displays to present information for centuries. Ancient civilizations, including American Indian, Egyptian, Greek, Roman, and Mayan, used wall posters of papyrus, wall paintings, frescoes, and murals to announce and advertise social, economic, and political events [1]. As manufacturing and printing technologies became more advanced, ever larger displays proliferated. Now the many large static displays we see around us include billboards, conference posters, signs, and even the Goodyear blimp. Generally, such printed or fabricated displays are referred to as static as they offer limited interaction and cannot be easily repurposed. The use of lighting, animatronics, and mechanical movement has brought to some of these static displays an aspect of dynamism.

Large displays such as chalkboards were introduced over two centuries ago and remain in use today. Now bulletin boards and whiteboards are ubiquitous as well. Each type of display is employed in the sharing and communication of information in both face-to-face and asynchronous settings. These display technologies are simple and low cost because they are reusable. While many wouldn’t consider a flat surface, such as a table, bench, or desk, to be a large display surface, it clearly is. Placing static objects or printed or written items on a blackboard or work surface allows one person or a group to easily organize large amounts of information and make them accessible to many others.

Display and projection technologies have transformed such static displays by adding live displayed or projected information. Novel forms of human computer interaction such as multitouch technology have further transformed these devices into interactive systems [3], as shown in Figure 1.1. Users can now interact directly with

FIGURE 1.1
Users sharing photos in Sharepic [2].

4 CHAPTER 1 Face-to-Face Collaborative Interfaces
the displayed objects by simply touching the display, thus creating a sense of immediacy and naturalness. This sense of immediacy gives rise to a range of human–computer interactions and provides support for multiuser multisurface collaborative activities [4].

Other technologies are further helping to bridge the digital–physical divide by coupling real objects (eyeglasses, phones, pens, or blocks) with projected information. The aim is to weave such new forms of interaction and computation into the fabric of our lives until they are indistinguishable from it [5].

Systems exhibiting ambient intelligence (AmI) must consider the broader range of inputs now possible with current desktop, mobile, and gaming devices and applications. In practice everyday objects will become sites for sensing, output, and processing along with user input [6]. Examples of data or knowledge AmI systems can rely on as input involve spatial information, identity, environment, activity, schedules, and agendas, along with data that can be mined and inferred from other measurements or historical analysis.

Broadly speaking the range of implicit inputs to an AmI system can be called context data whether sensed or inferred. Context includes information from a human (physiological state), the sensed environment (environmental state), and the computational environment (computational state) that can be provided to alter an application’s behavior [7, 8]. In contrast, explicit inputs from the user can include speech or movement of any kind. It is the explicit forms of human movement in face-to-face collaboration that are the focus of this chapter.

Large horizontal or vertical shared displays are an aspect of AmI or ubiquitous computing (UbiComp) in which computation is everywhere and computer functions are integrated in everything. As mentioned, computationally enhanced basic objects such as whiteboards, tables, benches, and desks allow enhanced forms of face-to-face computer-supported interaction and collaboration not possible with conventional desktop or mobile computing. For example, using digital photos the projected collaborative tabletop photo sharing in Figure 1.1 offers older people the social sharing and story telling common with physical photos. In Sharepic, users do not feel they are using a computer to collaborate; instead, the medium supports the actions directly. Ultimately, computer-supported face-to-face collaboration will become so common in bars, restaurants, schools, offices, and homes that no one will notice its presence, in spite of its power to enhance our lives.

The graphical user interface (GUI) of a personal computer relies on keyboard, screen, and mouse for input and output. GUIs typically offer the Window, icon, menu, pointer (WIMP) metaphor as an intuitive view of the computer, compared with the classical command line (textual) interface. Modern desktop operating systems are built around the GUI, WIMP, and keyboard/screen/mouse concepts, with interaction support for a single person. Technological advancements have moved computing beyond the desktop with the widespread adoption of mobile phones with keypads, PDAs with styli, tablets with styli, and touch screens. More recent developments have seen the adoption of game controller inputs or gesture-driven controls for game platforms.
In spite of these advancements, however, the tie between the single machine, the single application, and the single user remains. Typical desktop machines do not support two keyboards or two mice to allow people to enter information at the same time; game systems often split the screen, effectively carving out sections for each person; and combining personal devices (PDA with iPhone, say) into a collective ecosystem is often not possible.

Many AmI scenarios suggest radically new affordances and modes of interaction in the simplest and most basic operations of our daily lives. Two examples are turning a door handle while receiving haptic feedback on workflow status and clothing items that inform washing machines of their cleaning instructions. By contrast, the class of interface we consider here relies on common (e.g., face-to-face) interactions in well-understood social contexts such as meetings, or gatherings in a restaurant, as shown in the Microsoft Surface in Figure 1.2. Technologies to realize such computationally enhanced experience include computer vision, speech recognition, image processing, RFID, and projected and embedded displays. However, making active many of the surfaces we interact with on a daily basis requires the computational support and power they offer to reside at the periphery of our attention and go unnoticed until needed. The challenge here is not providing the next-generation mouse and keyboard but instead ensuring that sites that already support natural face-to-face collaboration operate fluidly and seamlessly once computation is introduced.

Interactive displays that support multiple users and simultaneous interaction in typical face-to-face activities will become standard. A display’s form factor typically defines its social benefits [10]. A mobile phone display can be comfortably viewed

FIGURE 1.2
Users sharing a map on the Microsoft Surface. Courtesy Microsoft.
by one user at a time; a large poster board is visible to multiple users simultaneously; an interactive surface such as the DiamondTouch [11] or the FTIR display shown in Figure 1.3 affords simultaneous interaction by many users. A display’s physical constraints, such as real estate, orientation, and mass, strongly affect the social interaction it supports, and this is further constrained by the users’ visual angle, territoriality, and capability to reach content and manipulate the display [3]. Interactive displays have been studied extensively in related work. Here the focus is on how touch or multitouch interactive displays can support face-to-face collaboration in AmI scenarios.

1.2 Background

The classical GUI has been successfully used since its inception, but the limitations of basic mouse and keyboard input are obvious, leading to the demand for new input and interaction modalities that will enhance the interface. Available since the early 70s in various forms, touch-sensitive surfaces are one solution. The PLATO IV, developed at the University of Illinois in 1972 and one of the first single-touch screens introduced, has a 16 × 16 array of touch-sensitive locations. One of the first multitouch systems was the flexible machine interface [12]. Here the touch-sensitive surface consisted of a frosted-glass sheet with special optical properties. When the sheet was touched, a black spot appeared that could be captured by the camera behind the sheet and recognized using simple computer vision.

Importantly, this was only an input device; the user viewed the visual feedback on a different device.

Buxton\(^3\) underscored the difference between touch tablets and touch screens as follows:

- **Touch tablets.** The touch-sensitive surface is not used for feedback display.
- **Touch screen.** The touch-sensitive surface is overlaid by the display.

The *soft machine* concept [13] builds on the classical hard machine concept (e.g., an oven or washing machine). It uses graphics software to generate images of controls (e.g., keys, pushbutton switches, and slides on the screen) to give the appearance of a hard machine. To ensure the illusion, the soft machine display is touch-sensitive, enabling a user to operate its controls by direct touch, as if they were physical controls. This connection between display and action increases the sense of immediacy in the interface.

The XEROX 5700 Electronic Printing System [14], the first commercial soft machine, has obvious parallels with the current generation of multitouch display prototypes and systems. In the original, all controls of the soft machine were on a black-and-white touch screen display. The Lemur,\(^4\) an advanced example of a soft machine, released in 2003 is a multitouch music controller device with all controls (knobs and sliders) virtually created on the display. This interface is customizable to user preferences. Other advances such as haptic output can overcome the lack of tactile feedback on soft versions of hard machines. Sony's tactile touch screen feedback is based on an actuator, constructed as a multilayer sandwich of thin (0.28-\(\mu\)m) piezoceramic films, that “bends” the screen. The haptic effect can be coupled with visual or audio feedback to simulate the feel of buttons that click [15].

Rather than a dedicated sensing surface, alternative approaches often rely on computer vision for sensing. VideoPlace is a vision-based system capable of tracking hands and thus recognizing and interpreting a rich set of gestures [16]. Users are placed against a neutral background, making it possible to process their silhouette image. The system can detect when they “touch” the graphical objects projected on the wall in front of them and react to the contact. The concepts demonstrated in VideoPlace are used in current camera-based multitouch systems that are able to “see” in front of the display. Diffuse illumination or capacitance-based systems (described in detail in Section 1.4) can discriminate between touch and hover.

The Digital Desk [17] uses optical techniques to sense hands, fingers, and objects being moved on the desk's surface. It operates by projecting images on the desktop where documents can be placed. Using an overhead camera, it reacts to interaction with the objects on the desk and can scan the documents placed there. Although

---


this is a front-projection system and occlusion can occur, the Digital Desk supports multitouch interaction through the use of both hands to rotate or scale documents or to manipulate objects on the desk.

Akin to the Digital Desk is Play Anywhere, a top-projection camera input system that allows multitouch user interaction [18]. It relies on a short-throw digital projector for top projection, an infrared illuminant, and an overhead camera to view the desk. These are all placed on the same stand, which makes this setup very portable because it can transform any flat surface into a multitouch sensitive display. The system can detect and track hands, fingers, and objects using a visual bar code scheme. The infrared illuminator illuminates the environment and causes a finger to cast an IR shadow on the display. If the shadow disappears, the system considers that the finger is touching the surface. By measuring the distance between the finger and its shadow, the system can also detect a hover.

More recent examples of touch-sensitive surface technologies include the Apple iTouch/iPhone product line and the SMART Table,⁵ (shown in Figure 1.4). The SMART Table uses digital vision touch (DViT), which relies on small cameras embedded in a device around the rim of the display [19]. When an object enters the field of view (FOV), the angle within the FOV of the camera is calculated. The SMART Table is multitouch and multiuser, designed for primary education applications. Other examples include ThinSight [20], developed for small LCD panels only, which supports multitouch input using hands and objects (or a stylus in N-trig’s⁶ case). ThinSight uses a 2D grid of retro-reflective optosensors (containing an IR light

---

⁵Copyright 2001 SMART Technologies ULC. All rights reserved.
emitter and an optically isolated IR light detector, allowing it to emit light and to
detect the intensity of incident light at the same time) placed behind an LCD panel.
When a reflective object is placed in front of the panel, a part of the light is reflected
back and detected by the optosensor. The data generated by the sensor grid is a low-
resolution grayscale “image” that can be processed to extract information about the
number of contact points and their position. N-trig uses a capacitive sensing system
(i.e., signals are emitted that capacitively couple with users when they touch the sur-
face) mounted in front of the LCD panel to detect touch, and a specialized sensor to
detect and track the stylus.

1.3 SURFACE USER INTERFACE

A tangible user interface (TUI), initially termed a graspable user interface [22], as
shown in Figure 1.5, is one that integrates both the representation and the control
of computation in the same physical artifact. It helps provide physical form to
computational artifacts and digital information, and, unlike a GUI, it “…makes
information directly graspable and manipulable with haptic feedback” [21]. The
informal definition of a touch screen given in Section 1.2 can be made more rigorous
and broad when defined as a “surface user interface” (SUI), which relies on a self-illu-
minated (e.g., LCD) or projected horizontal or vertical interactive surface coupled
with computation control on the same physical surface. As with a TUI, the outputs
from and inputs to a SUI are tightly coupled. SUI-based touch screens range from
small personal devices such as the iPhone to large public interactive surfaces such
as the DiamondTouch.

SUIs are used in public places (kiosks, ATMs) or in small personal devices (PDAs,
 iPhones) where a separate keyboard and mouse cannot or should not be used. Basic
SUIs have been common for over 20 years in the form of interactive kiosks, ATMs,
and point-of-sale systems, which rely on touch-screen technology with simple button
interfaces. The basic technologies in many products, such as the Nintendo DS with

FIGURE 1.5
TUI instantiations of GUI elements [21]. Courtesy Brygg Ulmer.
its resistive touch screen, were established in hardware research decades ago. The recent heightened interest in SUIs such as the iPhone generally stems from the low cost of production and the interaction styles they afford beyond the desktop paradigm.

The current generation of SUIs suitable for face-to-face interaction are built on LCD displays or form-factored into walls or coffee tables. In their current form they cannot be considered a “basic object” in Weiser’s vision. However, display technologies are now ubiquitous, and if SUI interaction styles can be woven into the environments and activities of everyday life and their industrial design improved, we can then achieve *invisibility in action*. As noted previously, the size of the display often determines the social benefits it supports. Thus, for face-to-face collaboration larger interactive horizontal or vertical surfaces are key. These have been researched and developed for over 20 years and from a technology perspective can be classified as front-projected, rear-projected, and self-illuminated.

For front-projected displays, the Digital Desk [17] is the seminal example. One application is a calculator based on a paper sheet with printed buttons. Pressing the paper buttons is recognized by a vision system and the current total is projected into the total square. DiamondTouch from MERL [11], Sony’s SmartSkin [23], AudioPad [24], and TANGerINE [25] all rely on front projection onto a touch-sensitive surface or use computer vision for sensing. Front-projected systems may suffer from occlusion of the projected image or camera line of sight caused by body parts. Capacitive systems such as SmartSkin and DiamondTouch are described further in Section 1.4.2.

Rear-projected SUIs avoid the problem of occlusion of the projected image or a camera’s line of sight. These large touch-sensitive SUIs can be seen in nightly newscasts, during political elections, and in other media. TouchLight, shown in Figure 1.6, relies on a rear-projected display and an infrared illuminant on a semi-transparent acrylic plastic plane fitted with HoloScreen material [26]. The reacTable has a horizontal orientation and can operate with physical objects and fiducial markers [27]. FTIR-based displays [28] rely on the total internal reflection (TIR) of light being frustrated by touch (see Section 1.4.1). The Microsoft Surface7 is a commercial-level multitouch, multiuser interactive surface. Based on experience with TouchLight [26] the Microsoft Surface is realized using diffused illumination technology (described in detail in Section 1.4.1). For a comparison of diffused illumination with frustrated total internal reflection of light see Figure 1.10. The Microsoft Surface can detect input from multiple hands and fingers, and it uses multiple cameras so its video input has high resolution. In addition, it can detect and identify objects and their position on the table. Recent advances such as the UlteriorScape [29] use a combination of multiple rear projectors and lumisty film for viewpoint-dependent face-to-face collaborative displays.

Self-projected systems are typically developed around LCD or plasma screen technologies as shown in Figure 1.4. Advances such as ThinSight [20] may result in

---

every future large screen being multitouch enabled. Alternatively, the OLED manufacturing process may be enhanced to form-factor IR LEDs and sensing elements directly into the same space as the pixels. Such developments would radically alter the nature of surface user interface research and development if all displays were multitouch ready.

1.4 MULTITOUCH

Multitouch SUIs can be mounted vertically on walls or horizontally on tables. They are capable of sensing the location of finger(s) when contact with the surface is made. The size of the surface can vary from a few inches to a few feet diagonally [30]. Buxton described the key features of SUIs in terms of touch tables which...
remain true today. Because multitouch SUIs have no mechanical intermediate device that can get lost or damaged, they are also appropriate for pointing tasks in moving or vibrating environments. Users can interact with all ten fingers at once compared to just one mouse per hand. Because of their simple one-piece construction they are suitable for intense use in clean or dirty environments.

As the most important advantage over mouse/keyboard interfaces is the possibility of multiple points of contact, experiments have been undertaken on two-handed input [31]. One experiment was composed of two tasks: selection and positioning and navigation and selection. The results show that for the first task all but one subject used both hands simultaneously and that parallel execution of subtasks led to a decrease in execution time from about 2.35 ms to 1.85 ms for novice users and from 1.8 ms to 1.6 ms for expert users. The second task was to select specific words from a document. The subjects were split again into novice and expert, but the general trend was the same: two-handed input completion times were lower than those for one-handed input by 15% for experts and 25% for novices. Also, the difference between expert and novice users when using two-handed input decreased significantly from 85% to 32% for one-handed input. These experiments demonstrate that splitting tasks between two hands leads to increased productivity, and that users tend to use both hands if possible, as such behavior is natural. When the opportunity presents itself and the task is suitable, users naturally revert to coordinated multi-hand interaction.

1.4.1 Camera-Based Systems

One of the first attempts at a large interactive display based on camera input was the Liveboard [32], which, although it used a stylus-based interaction method, was built on the principles used by the following prototypes. The pen used was cordless and had four buttons to control different states. These states gave Liveboard a slight advantage in this area over finger-touch displays. The pen sent a beam of optical light to the pen detection module behind the screen, from where the digitized pen position readings were transmitted to the PC through a serial port. The pen was accurate to less than 1 mm and could transmit 90 \textit{xy} positions per second. Figure 1.7 is a diagram of the Liveboard prototype. A positive feature of Liveboard was that it could be operated from a distance; this allowed natural gestures such as sweeping motions to be used to execute operations, including scrolling.

HoloWall is a vertical touch-sensitive display [33], and as seen in Figure 1.8 its configuration is straightforward. The image together with IR light is projected from the back of the screen. When the hand approaches the screen, the IR light is reflected and the IR-sensitive camera captures the point of touch. The input image from the camera is processed and the touch points are separated from the background. HoloWall’s advantage over Liveboard is that it can detect more than one point of contact and is easier to implement even for large displays. Because the IR
light is projected from the back, it is not always possible to distinguish between a touch and a very close hover, as the light is reflected even when close to the surface.

TouchLight [26], as shown in Figure 1.6, relies on a setup similar to HoloWall’s. However, here two cameras are used behind the screen and there is no diffusing projection surface. Instead, a different type of material is used on top of the acrylic sheet—namely, a DNP HoloScreen, which is a refractive holographic film that scatters light from a rear projector when the light hits it at a particular angle and is transparent to all other light. As there is no diffusing surface, the cameras can see beyond the surface and the system can achieve high-resolution imagery of objects on the surface. TouchLight supports user and face recognition capabilities and the scanning of documents. By using two cameras, TouchLight can determine how far objects are from the screen. After lens distortion correction, the images from both cameras
are fused to obtain the final input image, which is used to recognize multiple touches and over-time gestures.

The metaDESK [34], developed as part of the Tangible Bits project [35], is a complex system with an almost horizontal top surface (Figure 1.9). This drafting table orientation is similar to that in many of the systems so far described (back-projected screen, IR lamps, and cameras) with the addition of active and passive lenses and a range of physical objects for interaction with the surface. The largest component of the metaDESK is a back-projected graphical surface for display of 2D geographical information within the Tangible Geospace prototype. The active lens is an arm-mounted flat-panel display and serves to display 3D geographical information projected in 2D on the desk’s surface. This display panel is tracked by an Ascension Flock of Birds’ 6-DOF magnetic-field position sensor. The passive lenses are created out of fiber-optic cluster material using a wooden frame, which makes it possible to visually simulate an independent display surface. The physical objects, or “phicons” (physical icons), are made of transparent acrylic backed with a “hot mirrors” material that makes them visible to the IR cameras but transparent to the eye. A computer vision system composed of two cameras located inside the desk performs the sensing in metaDESK together with magnetic-field position and electrical-contact sensors. The entire system is coordinated by three networked computers. metaDESK provides a user interface that supports physical interaction with digital information projected on a near-horizontal screen.

**FIGURE 1.9**
metaDESK system hardware architecture. Courtesy Jun Rekimoto.
The systems described thus far fall into the broad diffused illumination category, as shown in Figure 1.10(a). Both a camera and an IR illuminant are placed behind the screen/display surface. As objects (hand, finger, face, paper, etc.) approach the screen, they reflect the IR light back to the camera, which typically has a filter to limit the range of IR light it sees. The video from the camera is then passed to
software that determines blobs and tracks them, and determines gestures and other higher-order user actions. These hardware elements form only a limited portion of an overall SUI. Once gestures and actions are determined these must be passed to an end-user application.

FTIR multitouch displays rely on the TIR of light being frustrated (F) by a user’s touch [36, 37, 28]. TIR occurs at the meeting point of light with mediums of a lower index of refraction. Depending on the angle of incidence, light is refracted and, when the angle is higher than a certain threshold, TIR occurs. When a different material (skin, say) makes contact with this medium it frustrates the TIR and scatters light out of its waveguide at the contact point, as shown in Figure 1.10(b).

Han’s FTIR multitouch prototype [28] uses a sheet of acrylic with diamond-polished edges lit by high-power IR LEDs to effect the TIR. This is akin to the FTIR system shown in Figure 1.3. A camera behind the screen is equipped with a matching bandpass filter to cover just the frequency of the light emitted from the diodes. It captures the light frustrated by any object touching the acrylic, so multiple contact points can be detected. Here the camera captures at 30 fps with a 640×480 resolution, so image processing can be easily accomplished in real time on a commodity PC. The system works well in combination with rear projection when a projection screen is placed on top of the acrylic. This configuration ensures that there is no disparity between the display and interaction surfaces.

The success of Han’s work owes more to the fluidity of prototype multitouch software applications than to any novelty of the FTIR approach itself [28]. Generally, a compliant surface is necessary between the acrylic and the projection screen as there is a slight delay after the moment the finger is lifted and until the contact effect completely dissipates. Without this, the finger leaves traces on the screen. The choice of compliant surface is particularly sensitive in an FTIR setup; to date there are no perfect solutions.

UlteriorScape [29] uses a combination of multiple rear projectors and lumisty film for viewpoint-dependent face-to-face collaborative displays. Multiple projectors are arranged so that light impacts a shared display along with secondary displays formed from the lumisty film. This allows both personal and public displays to be combined using a single rear-projected setting.

An alternate to the DI and FTIR approaches is DViT. The digital vision touch technology (DViT) from SMART Technologies [19] relies on camera input and image processing. Here two to four cameras are placed in the corners of the screen, facing diagonally across the surface rather than toward it. When the user touches the display using a finger or pointing device, the contact point coordinates are calculated by triangulation using the angle of the contact point relatively to all cameras. Once processed, the contact point is sent to the application as mouse clicks or as “electronic ink.” This system enables the use of fingers instead of pointing devices such as a mouse or stylus, which makes it easy and intuitive for untrained users. The fact that the cameras are in the corners means that there is no technology in the display itself so the system is resistant to extended use. The drawback with this technology is that because it does not readily support multiple contact points, any additional contacts after the first one is detected may be ignored.
1.4.2 Capacitance-Based Systems

Camera-based systems as described in Section 1.4.1 typically cannot distinguish the input of different users—one touch looks very much like another. Software solutions to this problem are emerging in which intelligent decision making or user observation plays a key role in accurate user disambiguation. In contrast, DiamondTouch is a touch-sensitive input device that distinguishes the simultaneous input of multiple users (Figure 1.11(a)) [11]. Unlike other touch-sensitive surfaces, DiamondTouch is not sensitive to objects placed on the surface as it is not pressure sensitive. The top layer of the surface is made out of a durable material (e.g., PVC).

The basic layout of DiamondTouch can be seen in Figure 1.11(b). The table surface is built on a layer of eight embedded antennas, constructed from electrically conductive material, that are insulated from each other. A different electrical signal is sent to each one, and when touched, signals are capacitively coupled from beneath the touch point, through the user, into the receiver unit assigned to that user. Each antenna occupies a single area of the table and can be clearly identified. Initially the size of the antenna was $5 \times 5\, \text{mm}$ so a single touch activates at least three to four antennas in the same area.

As with all front-projection technologies, one of the problems with DiamondTouch is that a user’s hand interferes with the image (e.g., cast shadows), as shown in Figure 1.11(a). Another problem is the need for each user to maintain contact with her assigned antenna to ensure an effective capacitive coupling and hence identification. In spite of these issues, many applications, including planning, device coordination, mapping [38], and photo sharing [2], have been built using the DiamondTouch hardware.

An alternate form of capacitive sensing is employed by the SmartSkin [23]. The SmartSkin surface, as shown in Figure 1.12, is built on a mesh of transmitter/receiver electrodes. This enables identification of the contact point with the table as well as measurement of the distance between the hand and the table. When a hand approaches the table it capacitively couples the electrodes and causes a wave signal.

![FIGURE 1.11](MERL DiamondTouch: (a) collaborative multiuser application; (b) schematic. Courtesy Circle Twelve.)
This makes it possible to measure the distance from the hand to the touch point. In practice, a dense sensor mesh enables such a system to determine the shape of the hand, which in turn increases the input capabilities.

Interaction techniques explored in SmartSkin include

- **Mouse emulation with distance measurement.** A hand’s position is tracked in 2D to emulate the mouse, and the distance from the table is used as a mouse click. A threshold distance is used to distinguish between a mouse click and a release.
- **Shape-based manipulation.** The shape of the hand is used to move objects on the display without touching them.
- **Capacitance tags.** The use of objects coated with a conductive material was also explored. When placed on the table alone these tags have no effect, as they are not grounded. However, once a user touches them they are sensed. Such tags are used to realize knobs and sliders as TUI controls.
While capacitive systems are now prolific in small displays, including the Nintendo DS, the Apple iPhone, ATMs, and kiosks, they are currently a niche area of research and development for large-scale collaborative systems. Because of their inability to support rear projection and because of the actual or perceived cost of the hardware, camera-based systems now dominate for sensing multitouch input on large displays.

1.5 GESTURAL INTERACTION

A gesture is the movement of a body part to express meaning. Typical gestures such as pointing, waving, or nodding are made by the hand or the head as appropriate. Both simple gestures such as pointing and complex gestures, such as in sign language rely heavily on the cultural, geographical, or linguistic frame of reference for their interpretation. Gesture recognition is the interpretation of human gestures using various inputs and computational processing. Almost any type of input device or system can be used to collect data on a user “gesture.”

On an SUI, gestures are typically classified as single-point, multipoint, multi-hand, and whole-hand. SUIs that incorporate gesture recognition often suffer from problems of feedback and visibility—for example, “How do I perform the gesture,” “How do I know what I just did was interpreted?” In this regard standards are important. This section reviews a standard gesture set that is well understood across SUI-based systems and that serves as a reference set in further SUI development.

FingerWorks has developed a diverse range of input gestures for their iGesture Pad. These include simple mouse emulations—point, click, double-click, right-click, drag/select, and scroll,—and more advanced gestures for Web browsing—touch and slide for back and forward, zoom in or zoom out—relying on all five fingers expanding or contracting on the table as shown in Table 1.1. File operation commands—open, save, new, print—and simple editing commands—cut, copy, paste—are also supported, as described in Table 1.2. Basic application commands mapped to gestures include exit, switch between applications, and minimize, as shown in Table 1.3. FingerWorks was acquired by Apple and the influence of their gesture research and development can be seen across the Apple product line from multitouch gestures on the MacBook trackpad to the iPhone.

Although these examples are all multitouch but single-hand input, as the iGesture Pad was used for input purposes only, they indicate the diversity of gestures that can be mapped into input. Not all gestures are natural and intuitive, but, depending on the task at hand, users can adapt and learn.

\[8\text{Accessed July 2008 from: http://www.fingerworks.com.}\]
<table>
<thead>
<tr>
<th>Gesture Type</th>
<th>Graphics</th>
<th>Action</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse Emulation</td>
<td><img src="image" alt="Mouse Emulation Gesture" /></td>
<td>Tap any two adjacent fingers.</td>
<td></td>
</tr>
<tr>
<td>Scroll</td>
<td><img src="image" alt="Scroll Gesture" /></td>
<td>Touch and slide four fingers up/down. “Roll” fingers for fine scrolling.</td>
<td></td>
</tr>
<tr>
<td>Double-click</td>
<td><img src="image" alt="Double-click Gesture" /></td>
<td>Tap three adjacent fingers once.</td>
<td></td>
</tr>
<tr>
<td>Right-click</td>
<td><img src="image" alt="Right-click Gesture" /></td>
<td>Tap thumb, middle, and ring fingers.</td>
<td></td>
</tr>
<tr>
<td>Drag/select</td>
<td><img src="image" alt="Drag/select Gesture" /></td>
<td>Touch and move three fingers.</td>
<td></td>
</tr>
<tr>
<td>Point</td>
<td><img src="image" alt="Point Gesture" /></td>
<td>Touch and move any two adjacent fingers.</td>
<td></td>
</tr>
<tr>
<td>Web Browsing</td>
<td><img src="image" alt="Web Browsing Gesture" /></td>
<td>Back</td>
<td>Touch and slide thumb and three fingers left.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Forward</td>
<td>Touch and slide thumb and three fingers right.</td>
</tr>
</tbody>
</table>

(Continued)
Table 1.1  Mouse and Web-Browsing Emulation Gestures—cont’d

<table>
<thead>
<tr>
<th>Gesture Type</th>
<th>Graphics</th>
<th>Action</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zoom in</td>
<td></td>
<td>Touch and expand thumb and four fingers.</td>
<td></td>
</tr>
<tr>
<td>Zoom out</td>
<td></td>
<td>Touch and contract thumb and four fingers.</td>
<td></td>
</tr>
<tr>
<td>Find</td>
<td></td>
<td>Touch and pinch thumb and two fingers.</td>
<td></td>
</tr>
</tbody>
</table>


Table 1.2  Editing and File Operation Gestures

<table>
<thead>
<tr>
<th>Gesture Type</th>
<th>Graphics</th>
<th>Action</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Editing</td>
<td></td>
<td>Cut</td>
<td>Touch and pinch thumb and middle finger.</td>
</tr>
<tr>
<td>Copy</td>
<td></td>
<td>Tap thumb and middle finger.</td>
<td></td>
</tr>
<tr>
<td>Paste</td>
<td></td>
<td>Touch and expand thumb and middle finger.</td>
<td></td>
</tr>
<tr>
<td>Gesture Type</td>
<td>Graphics</td>
<td>Action</td>
<td>Description</td>
</tr>
<tr>
<td>---------------</td>
<td>----------</td>
<td>----------</td>
<td>----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Undo/redo</td>
<td><img src="undo.png" alt="Image" /></td>
<td>Touch and slide thumb and middle finger up/down. Slide quickly and crisply to undo just one step, or gradually for multiple steps.</td>
<td></td>
</tr>
<tr>
<td>Tab/back tab</td>
<td><img src="tab.png" alt="Image" /></td>
<td>Touch and slide thumb and middle finger right/left. Slide quickly and crisply for just one tab, or gradually for repetitive tabs.</td>
<td></td>
</tr>
<tr>
<td>File Operation</td>
<td>Open</td>
<td>Touch and rotate counterclockwise thumb and three fingers.</td>
<td></td>
</tr>
<tr>
<td>Close</td>
<td><img src="close.png" alt="Image" /></td>
<td>Touch and rotate clockwise thumb and three fingers.</td>
<td></td>
</tr>
<tr>
<td>Save</td>
<td><img src="save.png" alt="Image" /></td>
<td>Touch and contract thumb and three fingers.</td>
<td></td>
</tr>
<tr>
<td>New</td>
<td><img src="new.png" alt="Image" /></td>
<td>Touch and expand thumb and three inner fingers.</td>
<td></td>
</tr>
<tr>
<td>Print</td>
<td><img src="print.png" alt="Image" /></td>
<td>Prespread hand, then touch and further expand thumb and three outer fingers.</td>
<td></td>
</tr>
</tbody>
</table>

Source: [http://www.fingerworks.com](http://www.fingerworks.com).
1.6 GESTURAL INFRASTRUCTURES

The gesture and speech infrastructure (GSI) [39] was built using DiamondTouch and DVIT technologies as described in Section 1.4. While it is a multimodal system (speech and gesture) the gestures GSI supports are very simple: one-finger pointing for selection, dragging, and panning; two-finger stretching for zooming in and out of a region; fist-stamping to create new objects; and palm-down wiping to delete objects. For example, a two-handed gesture might be multiple-object selection by surrounding an area with upright hands, used in many multitouch applications for grouping objects or as area selection. Sharepic includes a range of multihand and multitouch gestures such as two-handed copy and two-person reset (wherein two people can collaborate on a single two-handed (per person) gesture) [2].

RoomPlanner is furniture layout application developed with DiamondTouch technology [40]. It is designed for face-to-face work by two people sitting across a table who have a 2D overhead view of the room and the furniture. The interaction techniques here are more diverse than those in GSI or Sharepic. They are classified as follows:

- Single-finger input techniques for tapping and dragging (to select and move furniture in the room and to interact with the context-sensitive menu)
- Two-finger input techniques for rotation and scaling

<table>
<thead>
<tr>
<th>Gesture Type</th>
<th>Graphics</th>
<th>Action</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application Control</td>
<td><img src="image" alt="Gesture Illustration" /></td>
<td>Show desktop</td>
<td>Touch spread thumb and three fingers. Slide left.</td>
</tr>
<tr>
<td>Exit application</td>
<td><img src="image" alt="Gesture Illustration" /></td>
<td>Touch and rotate clockwise spread thumb and three fingers.</td>
<td></td>
</tr>
<tr>
<td>Switch application</td>
<td><img src="image" alt="Gesture Illustration" /></td>
<td>Spread hand, then touch three fingers and thumb and slide left or right. Slide crisply to advance just one window, or gradually to scroll through whole list.</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.3 Application Control Gestures

Source: [http://www.fingerworks.com](http://www.fingerworks.com).
Single-hand techniques:
- Flat hand on the table to rotate the display
- Vertical hand to sweep the furniture pieces (as one would sweep objects across a real table top)
- Horizontal hand on the table to display the properties of objects found in that area
- Tilted horizontal hand with front projection to create a physical space where private information can be projected

Two-handed techniques:
- Both hands vertical, moving away from each other to spread furniture around the room
- Both hands vertical, moving toward each other to bring objects in to the area delimited by them
- Corner-shaped hands used to create a rectangular editing plane

1.6.1 Gestural Software Support

The NUI Snowflake system addresses the needs of third parties in the development of interfaces for collaborative multitouch interaction. While supporting both FTIR and DI, it uses computer vision and image processing to recognize motion, gestures, and multitouch interaction. Figure 1.13 shows a multiuser, multi-input collaborative photographic application running on an NUI Horizon device.

The DiamondSpin Toolkit [38] also supports third parties in the development of interfaces. Along with support for DiamondTouch, its main feature is a real-time polar-to-Cartesian transformation engine that enables around-the-table interactions. Among the applications built on top of the toolkit is Table for N, designed for a small number of people at a table creating, sketching, annotating, manipulating, and browsing various types of documents, including text, html, images, and video clips. Another application is UbiTable, which enables rapid display ecosystem formation.
and content exchange between the UbiTable and other devices such as laptops, cameras, and PDAs.

CollabDraw uses the DiamondSpin Toolkit to enable collaborative art and photo manipulation [41]. Its hardware support is based on DiamondTouch. The notion of cooperative gesturing is introduced here, referring to gestures from multiple users interpreted by the system as one command. The system recognizes six basic hand inputs: single-finger, two-finger, three-finger, flat palm, single-hand edge, and two-hand edges. Collaborative gestures are also implemented by combining two or more basic gestures from different users. Eleven collaborative gestures were explored in this first prototype. For example, for drawing a stroke on the canvas, a touch on the surface with one finger by a user is necessary; in parallel another user can change the stroke’s width or color. An interesting feature of the system is the ability to create collaborative gestures when more than two people are sitting at the table. One such gesture is stroke erasing, when all users around the table rub their palms back and forth to clean up the canvas. This is akin to the two-person reset introduced earlier [2].

1.7 TOUCH VERSUS MOUSE

The move from a standard mouse or even a multimouse setup to a fully touch-driven interface can introduce many human–computer interaction issues, such as keyboard input, accuracy of control, and fatigue. Preliminary user studies have been carried out to compare the differences between touch and mouse input and one-handed and two-handed input [42]. In line with earlier results [31], these experiments show the advantage of two hands, albeit, in this case, only over the use of two mouse devices. In the first part of the experiment user preference and performance for one-touch input, and mouse input were tested. As in prior studies [43] target selection time was very close between the two input modes. There was a slight advantage in touch mode, but this mode also had more selection errors, particularly when the distance to the target increased. The docking task showed an advantage of the mouse, suggesting that dragging the finger across long distances on the surface is inefficient. According to the first part of the experiment, for single-mode input the mouse was indicated and preferred by users. The results for the second experiment (a bimanual task) showed that users had trouble tracking the mouse pointers on the screen, leading to higher task completion times.

One user study [44] suggested that mouse pointers should be visually connected to give more clues about their position and to decrease the concentration required for operation. Direct-touch input using both hands had better completion times, but whereas there were no errors for close targets, touch was more error prone once the distance to the target increased. This was not the case with two-mouse input, where the error rates were constant, as in the first experiment.
Some of the errors with two-handed input were caused by the need for symmetrical movement of the hands; usually the nondominant hand is harder to control for precise tasks. As in real life, most tasks requiring both hands are asymmetrical [45], and an experiment focusing on different task types may show fewer errors with two-handed input.

Another problem with touch is the lack of precision in selection. Techniques for precise selection have been proposed [46, 47, 48, 43].

These studies provide evidence that multitouch is preferred to multimouse input, and that, even though in interactions where single-touch is sufficient, the mouse is more efficient. Touch input has other advantages such as reduced hand fatigue and enabled awareness of users action in a multiuser context.

1.8 DESIGN GUIDELINES FOR SUIs FOR COLLABORATION

Although the use of SUIs for supporting collaboration is a relatively new area of research, various design guidelines have been proposed [49, 50, 51]. This section discusses the key elements of these guidelines with respect to face-to-face interaction while keeping the original terminology from those sources.

Hardware Setup

- **Size.** Collaborative face-to-face systems must be considerably larger than traditional displays to support a number of users working together.
- **Configuration.** A tabletop or a wall display is task dependent. In the case of tabletop displays, which are more functional for face-to-face collaboration, the display space can be split into shared and private areas.
- **Input.** Each user should have at least one input mode (touch in the case of a touch-sensitive surface). To date, only the DiamondTouch can distinguish between different users’ input. However, with the personal display spaces clearly separated, user identification can be realized at the application level with other multitouch systems as well.
- **Resolution.** Large displays suffer from low resolution for both output and input. The use of multiple projectors is possible but their cost is high. Touch input also has low resolution, and selecting small icons with the finger can be a challenge. Special techniques for precise selection should be considered depending on the face-to-face task at hand [46, 47, 52].
- **Interactive response.** In the case of face-to-face tasks, interaction can be computationally intensive. Preprocessing must be considered to reduce response delay.
- **Support for transitions between tabletop collaboration and external work.** There must be easy file transfer to personal devices for individual work.
Support for the use of physical objects. The TUI prototypes described in Section 1.4 are examples of support for input through different objects. Also, special areas can be created on the sides of tables for objects that are not input devices.

Application-Level Guidelines

Support for mental models. To interact freely with interface objects, users should be able to rearrange them according to their preferences for efficiency and comfort.

Representation changes. Allow users access to various representations of the data, according to individual preference. As was shown in [53] seeing more than one representation of the same data can have a positive effect on the decision-making process.

Task history. Keep track of interactions during the session for later discussion and reference.

Perception. Some of the issues to be considered are user viewing angle, global and local legends, and strategic label placement [54]. Another issue is rotating representations so that all users at the table can have a clear view.

1.8.1 Designing the Collaborative Environment

According to [55] two types of basic operation are necessary within groupware systems for a fluid collaboration between users: coordination and communication.

Group coordination guidelines include:

- **Workspace organization.** Public and private display and work spaces are necessary.

- **Fluid transitions between activities.** There should not be much effort required to switch between different activities or stages within the same operation. This could entail provision of more than one input mode or addition of interface objects that can be easily manipulated.

- **Information access.** Rights and restrictions should be in place for different actions or access to data.

- **Collaboration styles.** Various collaboration styles should be supported. Multiple copies of documents must be made available for individual work, and concurrent access and interaction with shared data are necessary.

The communication guidelines are

- **Support for easy transitions between personal and group work.** Give users distinct work areas for individual work during group sessions.

- **Support for interpersonal interaction.** Take into account the fundamental mechanisms that people use in collaborative interactions. The ergonomics of the system must be considered to make it suitable for collaborative work.

- **Privacy.** Users should have control over how much of the data they want to expose to others.
1.9 CONCLUSIONS

The following usability issues, if ignored in practice, can hinder the development of any collaborative SUI-based application. They must be addressed each time an SUI-based system is to be considered.

- The user’s stylus, fingers, and hands may partially occlude the interface.
- Interface elements may be difficult to select because of stylus, finger, or hand size.
- Users may suffer fatigue due to the range of human motion required.
- The screen surface can be damaged or dirty.
- There may be a lack of tactile feedback from passive screen surfaces.
- Calibration between the display (projector or LCD) and the sensing elements can become misaligned.

This chapter provided an overview of the core research and developments in multitouch display systems for surface user and face-to-face collaborative interfaces. We showed how these developments have impacted each other and the relative merits of each. We also discussed the notion of a surface user interface and how various forms of technology can be used to realize it for collaborative interaction. Tables 1.1 through 1.3 illustrated the types of gesture, action, and description SUIs should support. The chapter concluded with design guidelines for SUIs and how they can be applied to future interface design.

The ultimate goal of surface user interfaces in collaborative face-to-face activities is for people not to feel they are using a computer; instead, the visual elements should naturally support their actions. Ultimately, SUIs will become so commonplace in everyday life that no one will notice their presence. They will be aesthetic, powerful, and enhance our lives but so too will they be commonplace, obvious, and boring.

REFERENCES


