

# Minput: Enabling Interaction on Small Mobile Devices with High-Precision, Low-Cost, Multipoint Optical Tracking

Chris Harrison      Scott E. Hudson

Human-Computer Interaction Institute, Carnegie Mellon University  
5000 Forbes Avenue, Pittsburgh, PA 15213  
{chris.harrison, scott.hudson}@cs.cmu.edu



Figure 1. Our prototype *Minput*-augmented device running an audio player application. Two optical sensors on the back of the device enable x/y translation and rotational tracking on ad hoc surfaces, such as tables, walls, clothes, and the palm of one's hand.

## ABSTRACT

We present *Minput*, a sensing and input method that enables intuitive and accurate interaction on very small devices – ones too small for practical touch screen use and with limited space to accommodate physical buttons. We achieve this by incorporating two, inexpensive and high-precision optical sensors (like those found in optical mice) into the underside of the device. This allows the entire device to be used as an input mechanism, instead of the screen, avoiding occlusion by fingers. In addition to x/y translation, our system also captures twisting motion, enabling many interesting interaction opportunities typically found in larger and far more complex systems.

**ACM Classification:** H.5.2 [Information interfaces and presentation]: User Interfaces - Input Devices and Strategies, Interaction Styles, Graphical User Interfaces.

**General terms:** Human Factors

**Keywords:** Mobile devices, touch screens, optical tracking, pointing, input, sensors, spatially aware displays, gestures.

## INTRODUCTION

Small mobile devices offer the promise of significant computational power that can be carried with us into variety of circumstances and environments. As advances in electronics allow devices to become smaller and smaller, we begin

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.

Copyright 2010 ACM 978-1-60558-929-9/10/04...\$10.00.

to reach limits not of the electronics, but in the area needed to provide a usable human interface. Buttons, for example either begin to consume a significant fraction of available surface area, or become too small for comfortable and effective use. Techniques such as touch screens become less effective, especially for fine-granularity operations, when the size of a human finger begins to take up a significant fraction of the entire display.

Considerable work has attempted to address issues in this area. Although many approaches can operate in small to medium form factors, each suffers from at least some drawbacks. Vision-based approaches are perhaps the most compelling (see e.g., [9,15,16]). These, however, require the integration of a camera into, e.g., the backside, forcing the user to grasp the (small) device in a very particular fashion. Furthermore, vision processing is computationally expensive and error prone in dynamic contexts such as walking or waiting for the bus (with high levels of non-input related optical flow).

Acoustics offer another approach [7], although such methods will always have to contend with environmental noise and face privacy and social issues in shared settings. Accelerometers also suffer from high false positives in dynamic contexts, such as walking and riding on public transportation. Moreover, they tend to offer a lower level of fine control and expressivity – barring them from applications requiring varied and high accuracy interactions.

Several point solutions have attempted to address this problem. Of note is SideSight [2], which uses infrared proximity detection around the device periphery to perform multi-touch tracking. It is unclear, however, how well irregular

surfaces (e.g., the palm or clothing) affect tracking due to line-of-sight issues. NanoTouch [1] cleverly incorporates a touch-sensitive surface on the underside of a device, allowing for direct finger manipulation without screen occlusion. On a very small device, the focus of our efforts, it is not clear if one can even comfortably place two fingers on the underside for accurate multitouch gestures, such as pinching. More important, however, is that both systems essentially provide a 1:1 control-device (C-D) gain, tightly coupling the resolution of input to the size of the device.

### MINPUT

The mass production of optical mice has made the highly sophisticated sensors on which they rely very inexpensive. Additionally, advances in electronics and optics have yielded sensors that are both small and extremely precise. A generic optical mouse, costing only a few dollars, is capable of capturing and comparing surface images several thousand times per second. Often, this high resolution enables their use on a variety of surfaces - both traditional (e.g., mouse pads, tables) and ad hoc (e.g., palms, pants, bed covers, papers) (Figure 1). Vision-based interaction techniques (e.g., [9,15,16]) tend to heavily tax even modern mobile device processors and batteries. Fortunately, the optical sensors we employ have dedicated, highly efficient processors that handle most of the computation with negligible power consumption.

The central idea behind *Minput* is simple: place two optical tracking sensors on the back of a very small device. This allows the whole device to be manipulated for input, enabling many interesting interactions with excellent physical affordances [6]. This could allow for the creation of e.g., a matchbook-sized media player that is essentially all screen on its front side. The device could be operated using any convenient surface, e.g., a table or palm. The use of two tracking elements enables not only conventional x/y tracking, but also rotation [8], providing a more expressive design space. The latter motion is calculated by taking the difference in velocities of the two sensors.

This configuration allows *Minput* to operate like a spatially aware display [5]. Previous systems, however, have tended to be large and complex. For example, [4] and [13] were tethered to high-cost and stationary tracking systems, while [11] used an equally immobile augmented table and vision system. *Minput* provides much of the same capability, but in an inexpensive, low-power, compact and mobile form.

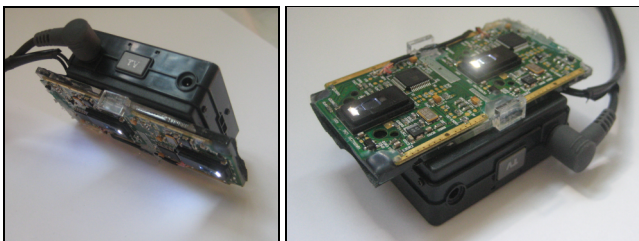


Figure 2. Two optical sensors operate on the underside of our proof-of-concept *Minput* device.

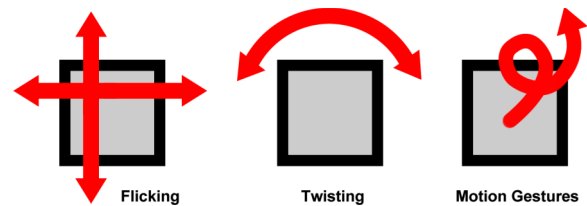


Figure 3. Possible gestures include omni-directional flicking, twisting clockwise and counter clockwise, and motion paths.

### PROTOTYPE

To investigate the usability and accuracy of our input approach, we constructed a small prototype device (Figure 2). For a display, we used a NHJ VTV-101 TV wristwatch modified to receive video from a conventional desktop computer (where interface control and rendering for our proof-of-concept device took place). The device features a 1.5" TFT LCD (30x23mm) with a resolution of 280x220. On the underside of the device, we mounted optical sensors extracted from two SlimG4 mice. The sensor and optics package is a diminutive 9x17x3mm, allowing it to be readily integrated into mobile device hardware. At the heart of the sensor is an ATA2198 processor, manufactured by At-Lab (<http://atlab.co.kr>), which samples at 3.4kHz (800 CPI). Translation data from the two sensors is transmitted over USB to the aforementioned PC.

### INPUT MODALITIES

*Minput* is an enabling technique on top of which numerous, distinct input modalities can be built. To illustrate this, we highlight three interaction techniques we believe to be of particular utility: gestures, virtual windows, and cursor control. We also introduce a new interaction: twisting for zooming and selection.

#### Gestures

The high precision motion captured by our approach makes gestures a strong candidate for input. As a proof of concept, we developed software that detected two basic forms: flicking and twisting (Figure 3). To flick, users simply rapidly swipe the device in a particular direction. We primarily used up, down, left and right, but diagonals and other angles are possible. Twisting is achieved by rotating the device around its center point. This feels much like twisting a physical knob, and offers many of the same affordances. More complex gestures are, as they are with mice or styli, eminently performable.

In piloting, we observed two distinct ways people perform such gestures. Some users held the device above the surface. When a gesture was to be performed, it was only then that the device made contact with the surface (for a moment). It then returned to a central, hovering position. Conversely, some users tended to prefer resting the device on the surface. This allowed gestures to be performed immediately. However, after the gesture was complete, users lifted the device, re-centered it, and placed it back on the surface. The latter is similar to clutching in mice (e.g., when the edge of the mouse pad is reached). In both methods, contact with the surface acts as a clutch for input.

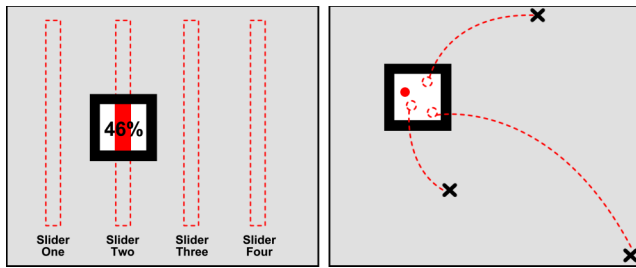


Figure 4. Left: The small display is mapped onto a larger physical surface. Right: the larger physical surface is mapped onto the smaller device.

### Virtual Windows and Zooming

*Minput* also allows for the device to be treated like a window looking onto a larger physical space (see, e.g., [4,9,17] for additional details and interactions). Consider, for example, four virtual sliders situated on a common surface, as depicted in Figure 4 (left). Users can switch between these different controls by physically moving the device left or right, to the corresponding spatial locations. Once situated on the desired control, they can alter the value by manipulating the device (e.g., up/down or twisting). *Minput* also offers very fluid interaction with zoomable interfaces (including multi-scale virtual windows). Twist is used for zoom - an analog operation for which it is well suited.

### Cursor Control

Finally, it is possible to map the device's spatial position on a larger surface to a cursor position on the screen. As illustrated in Figure 4 (right), to move the cursor to the bottom right of the screen requires translating the device to the bottom right of the surface. Unlike [1,16] and conventional touch screens, which are forced to operate with a 1:1 C-D gain, we can appropriate large surfaces from the environment to offer a high C-D gain (e.g., 5:1), offering extremely precise interaction (Figure 5). For example, it is possible to hit one-pixel targets with *Minput* without special mechanisms (e.g., [10,14]). Selection (i.e. "clicking") could be achieved with a twisting motion, a button on the side of the device, or by tapping the whole device.

### EXAMPLE APPLICATIONS

To demonstrate the immediate feasibility and accuracy of our sensing approach, we built several, proof-of-concept applications. Many are combinations of the techniques described in the previous sections.

#### Hierarchical Interfaces

Foremost, *Minput* readily supports hierarchical navigation. This capability opens the interface design space to nearly everything seen on contemporary mobile devices.

Our first application was an iPod-like audio player. We employ vertical lists of items (artists, songs), which are navigated by dragging the device upwards and downwards (Figure 1, left). Selecting an item (entering an artist, playing a song) is achieved with a right flick. Users are able to traverse backwards by left flicking. A volume control acts as a home screen, the value of which can be altered by turning the device like a knob (Figure 1, right).

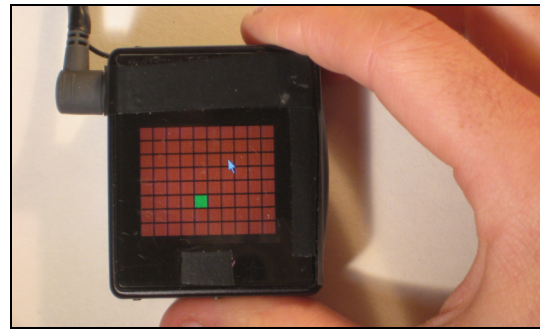


Figure 5. *Minput* supports conventional cursor control, appropriating ad hoc surfaces for high C-D gain movement.

We also created a simple photo album viewer to illustrate an alternative, gestures-only, navigation interface. Instead of a continuous lists, flicks are used to traverse between different albums. Once a desired album is located, it can be entered with a clockwise twist. Users are then limited to navigating photos in that album. To leave the "directory", users can perform a counterclockwise twist at any time.

Additionally, the two aforementioned applications are entirely gestural, and require no "clicks". This allows users to contiguously grip the device with a single hand, and without the need to reposition (fingers or otherwise) or reach for buttons. This motion-only interaction also means users can grip the device in any number of configurations they find comfortable or convenient.

#### Scrolling and Zooming

The prevalence of large content and small screen sizes has made scrolling and zooming common operations on mobile devices. Apple's iPhone, for example, uses finger drags to move the focus and pinch gestures to change the scale. *Minput* can replicate this capability through one-handed positional movement and twisting gestures. The latter is responsible for controlling the zoom level; a clockwise twist is used to zoom in, while counterclockwise rotation zooms out (Figure 6). After completing a zoom gesture, the device can be lifted and reoriented. If desired, the interface can be graphically counter-rotated such that the content remains properly oriented on screen.

We created two demonstration, *Minput*-driven, photo-browsing applications. The first displays a single, high-resolution photograph. The device acts like a small window, looking at only a part of a much larger image [4,9,17].



Figure 6. A user can pan around a document by translating the device. Zoom level is controlled by twisting clockwise or counterclockwise. After a motion is completed, the device can be lifted and reoriented (similar to mouse clutching).

Users are able to explore the picture by moving the device in any direction; twisting controls zoom. The second photo application lays out a grid of photographs. Flicks allow users to move up, down, left and right, one photo at a time. Twisting gestures control how much of the grid is visible at any given moment (single photo, 2x2, 3x3, or 5x5 grid).

We also created a mock web browser (Figure 6). *Minput*, coupled with the ability to click (either by tapping the device, or through a separate button, perhaps on the side of the device), enables an immediate method for navigating the web on very small displays - ones without directional inputs or touch screen ability.

### Conventional Interfaces

*Minput's* ability to act like a high-precision mouse makes conventional WIMP-like interfaces readily possible. This could allow devices with very small displays to support potentially rich interfaces – ones that would breakdown with touch screen input (e.g., accurately clicking targets like those shown in Figure 5). Additionally, *Minput's* high-accuracy positional movement could also prove useful in other, more specialized contexts, where gestures and virtual window interactions are too coarse. For example, selecting a sentence of text from a paragraph (for, e.g., search, copy/paste) requires precise two-dimensional positioning.

### EXPERIENCE

In order to get a preliminary understanding of the sensitivity and usability of our approach, we presented our prototype to eight beta testers (four female, mean age 23.1) who had not seen or used the device. The testers were allowed to play with each application, including the music player, photo album browser, single photo navigation with zoom, photo grid navigation with zoom, and mock webpage navigation. During the session, users were encouraged to provide commentary on their impressions. When needed, verbal instructions were given to help users operate the applications. Questions were also asked to elicit feedback.

Reactions were overwhelmingly positive. People consistently used words like “natural” and “intuitive” to describe the interactions, with several noting that they “understand completely how to use it” within a few minutes of using the device.

The twist feature was particularly popular. Most testers found it to be very natural to perform and conceptually logical. People, unprompted, likened it to the twisting of a lens on a camera, or the twisting of a screw (to “drill” in or out). However, users found it less intuitive for selection, instead, finding flicking more natural. Finally, several users commented that the physicality of the device was a nice property, offering many of the affordances of their corresponding physical counterparts. One tester suggested that this property made flicking with *Minput* more intuitive than the finger flicking as implemented on the iPhone.

### CONCLUSION

We have presented *Minput*, an input and sensing technique that enables intuitive and high-precision interaction on very small devices. This offers similar performance and capa-

bilities to that of considerably more sophisticated systems, but at a fraction of the cost and complexity, and in a form factor palatable for mobile computing. We describe our robust prototype and applications developed for it. We conclude with a brief overview of feedback from eight users, who uniformly understood and appreciated the interactions.

### ACKNOWLEDGMENTS

This work was supported in part by grants IIS-0713509 and IIS-0840766 from the National Science Foundation.

### REFERENCES

1. Baudisch, P. and Chu, G. Back-of-device interaction allows creating very small touch devices. In *Proc. CHI '09*. 1923-1932.
2. Butler, A., Izadi, S., and Hodges, S. SideSight: multi-"touch" interaction around small devices. In *Proc. UIST '08*. 201-204.
3. Cho, S., Murray-Smith, R., and Kim, Y. Multi-context photo browsing on mobile devices based on tilt dynamics. In *Proc. MobileHCI '07*. 190-197.
4. Fitzmaurice, G. W., Zhai, S. and Chignell, M. H. Virtual Reality for Palmtop Computers. *ACM Trans. on Information Systems*, 11(3): 197-218 (1993).
5. Fitzmaurice, G. W. Situated Information Spaces and Spatially Aware Palmtop Computers. *Comm. of the ACM*, 36(7): 38-49 (1993).
6. Fitzmaurice, G. W., Ishii, H., and Buxton, W. A. Bricks: laying the foundations for graspable user interfaces. In *Proc. CHI '95*. 442-449.
7. Harada, S., Landay, J. A., Malkin, J., Li, X., and Bilmes, J. A. The vocal joystick: evaluation of voice-based cursor control techniques. In *Proc. ASSETS '06*. 197-204.
8. Masui, T., Tsukada, K., Sii, I. MouseField: A Simple and Versatile Input Device for Ubiquitous Computing. In *Proc. UbiComp '04*. 319-328.
9. Mooser, J., You, S., and Neumann, U. Large document, small screen: a camera driven scroll and zoom control for mobile devices. In *Proc. ISD '08*. 27-34.
10. Olwal, A., Feiner, S., and Heyman, S. Rubbing and tapping for precise and rapid selection on touch-screen displays. In *Proc. CHI '08*. 295-304.
11. Olwal, A. LightSense: enabling spatially aware handheld interaction devices. In *Proc. ISMAR '06*. 119-122.
12. Ronkainen, S., Häkkinen, J., Kaleva, S., Colley, A., and Linjama, J. Tap input as an embedded interaction method for mobile devices. In *Proc. TEI '07*. 263-270.
13. Tsang, M., Fitzmaurice, G. W., Kurtenbach, G., Khan, A., and Buxton, B. Boom chameleon: simultaneous capture of 3D viewpoint, voice and gesture annotations on a spatially-aware display. *ACM Trans. Graph.* 22, 3 (Jul. 2003), 698-698.
14. Vogel, D. and Baudisch, P. Shift: a technique for operating pen-based interfaces using touch. In *Proc. CHI '07*. 657-666.
15. Wang, J., Zhai, S., and Canny, J. Camera phone based motion sensing: interaction techniques, applications and performance study. In *Proc. UIST '06*. 101-110.
16. Wigdor, D., Forlines, C., Baudisch, P., Barnwell, J., and Shen, C. Lucid touch: a see-through mobile device. In *Proc. UIST '07*. 269-278.
17. Yee, K. Peephole displays: pen interaction on spatially aware handheld computers. In *Proc. CHI '03*. 1-8.