The Smart Phone: A Ubiquitous Input Device

Rafael Ballagas, Jan Borchers
Media Computing Group
RWTH Aachen University
Aachen, Germany
(ballagas, borchers)@cs.rwth-aachen.de

Michael Rohs
Institute for Pervasive Computing
ETH Zurich
Zurich, Switzerland
rohs@inf.ethz.ch

Jennifer G. Sheridan
Ubiquitous Computing Group
Lancaster University
Lancaster, England
sheridaj@comp.lancs.ac.uk

INTRODUCTION

Mark Weiser envisioned ubiquitous computing as a world where computation and communication would be conveniently at hand and distributed throughout our everyday environment. [1] As mobile phones are rapidly becoming more powerful, this is beginning to become reality. Your mobile phone is the first truly pervasive computer. It helps you to both keep in touch with others and to manage everyday tasks. Consequently, it's always with you. Technological trends result in ever more features packed into this small, convenient form factor. Smart phones can already see, hear, and sense their environment. But, as Weiser pointed out: "Prototype tabs, pads and boards are just the beginning of ubiquitous computing. The real power of the concept comes not from any one of these devices; it emerges from the interaction of all of them." Therefore, we will show how modern mobile phones (Weiser's tabs) can interact with their environment – especially large situated displays (Weiser's boards).

The emerging capabilities of smart phones are fueling a rise in the use of mobile phones as input devices to the resources available in the environment such as situated displays, vending machines, and home appliances. The ubiquity of mobile phones gives them great potential to be the default physical interface for ubiquitous computing applications. This would provide the foundation for new interaction paradigms, similar to the way the mouse and keyboard on desktop systems enabled the WIMP (windows, icons, menus, pointers) paradigm of the graphical user interface to emerge. However, before this potential is realized, we must find interaction techniques that are intuitive, efficient, and enjoyable for applications in the ubiquitous computing domain.

In this article, we survey the different interaction techniques that use mobile phones as input devices to ubiquitous computing environments, including two techniques that we have developed ourselves. We use the word "smart phone" to describe an enhanced mobile phone. In our analysis, we blur the line between smart phones and personal digital assistants (PDAs), such as the PalmPilot, because the feature sets continue to converge.

ANALYZING THE INPUT DESIGN SPACE

Foley, Wallace and Chan's classic paper [2] structures a taxonomy of input devices around the graphics subtasks that they are capable of performing (Position, Orient, Select, Path, Quantify, and Text Entry). We use this taxonomy as a framework to structure our analysis of smart phones as ubiquitous input devices. A detailed discussion of the requirements and general issues of each subtask can be found in the original text. Although Foley et al.'s analysis was done in the context of the desktop computing paradigm, these subtasks are still applicable to ubiquitous computing. They naturally carry over to situated display interactions, but their applicability is not limited to graphical interactions.

Position

During a positioning task, the user specifies a position in application coordinates. This subtask is often used as part of a command to place an entity at a particular position. Positioning techniques can either be continuous where the object position is continually changing during the positioning task or discrete where the position is changed at the end of the positioning task. We note that position could refer to screen position, or physical position in the real world. For example, the height of motorized window blinds can be adjusted using the position subtask.

The mobile phone has been used for positioning tasks in a variety of ways:

Continuous indirect translation with a trackpad. Remote Commander [3] enables individuals to use the touch screen on a PDA as a trackpad to control the relative position of a cursor on a remote display.

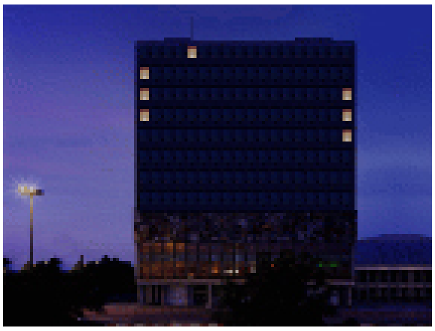

Figure 1. The view of the side of a building used to play the classic game Pong with buttons on the mobile phone controlling the paddle.

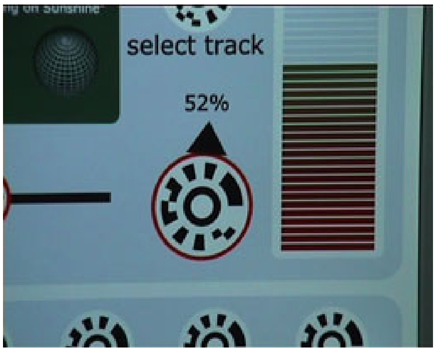

Figure 2. Using the phone to manipulate tagged widgets such as dials and sliders.

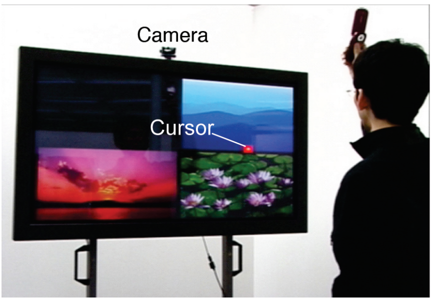

Figure 3. User waves the phone screen in front of a camera to control cursor position.

Continuous indirect translation with velocity-controlled joystick. A return-to-zero joystick controls the velocity of a continuously repositioned object. Zero displacement of the joystick corresponds to zero velocity. Silfverberg et al. [4] have done an in-depth study of isometric joysticks on handheld devices to control the cursor on a public display. Many of today's mobile phones are shipping with binary resolution, 4 to 8 direction, return-to-zero joysticks.

Continuous indirect translation with accelerometers. Accelerometers are beginning to emerge in handheld devices such as Samsung's SCH-S310 mobile phone with an integrated 3-D accelerometer. Rock 'n' Scroll [5] allows users to scroll (e.g. through an electronic photo album) by tilting the handheld device. Although this technique was used to interact with an application directly on the device, it could clearly be extended to positioning tasks in ubiquitous computing environments.

Continuous indirect translation with directional step keys. The location of an object is controlled by using up, down, left, right step keys for 2-D applications, plus in and out for 3-D. In the Blinkenlights project [6], users played the arcade classic Pong using the side of a building as a large public display. Each window equalled one pixel on the 18x8 pixel display (shown in Figure 1). Players connected to the display by making a standard call to a phone number. Pressing the number 5 on the phone keypad moves the paddle up, and the number 8 moves it down.

Continuous direct translation with camera tracking. Madhavapeddy et al. [7] present camera-based interactions involving tagging interactive GUI elements such as sliders and dials (see Figure 2). In manipulating the position and orientation of the phone camera, the user can position a graphical slider, or orient a graphical dial. An analogy can be drawn to the classic light pen with a tracking cross. As the light pen moves to a new position, the cross follows the motions of the pen. Tracking may be lost if the pen is moved to fast, but can be easily resumed by repositioning the pen back to the tracking cross. Madhavapeddy et al.'s interactions rely on the tagged GUI widget instead of a cross for tracking.

In Madhavapeddy et al.'s positioning technique the phone is responsible for tracking, and the environment is responsible for displaying and refreshing the tracked images. Other interactions have been created with these roles swapped. For example, smart phones have been augmented with laser pointers, as in [8], making them suitable for positioning tasks described by Olsen et al. [9] that use a camera in the environment to track the laser.

Continuous indirect translation with camera tracking. The C-Blink [10] system allows users to position a cursor on a large display using a mobile phone with a colored screen as shown in Figure 3. The user runs a program on the phone that rapidly changes the hue of the phone screen and waves the phone in front of a camera mounted above the large display. The displayed hue sequence encodes an ID to support multiple users. The camera tracks the absolute position of this signal in the camera image to control the cursor on the display.

Figure 4. Point & Shoot interaction: (Left) The phone screen is used to aim at a puzzle piece on a large situated display. (Middle) Pressing the joystick indicates a selection and a Visual Code grid is briefly superimposed to compute the target coordinates in the captured photo. (Right) The grid has disappeared and the target puzzle piece is highlighted indicating successful selection.

Discrete direct translation using a camera image. The Point & Shoot [11] interaction technique is illustrated in Figure 4. The locus of attention is on the phone screen. You aim using a cross hair displayed over a live camera image on the mobile phone. To reposition the cursor, you press and release the joystick button while aiming at the desired position with the cross hair.[1] This triggers a grid of tags to shortly superimpose over the large display contents for the camera, as can be seen in the middle of Figure 4. The grid is used to derive a perspective independent coordinate system to determine the selected point with pixel-level accuracy, which is enabled by the special properties of the Visual Code tags. [12] An analogy can be drawn to the classic light pen with position discretely determined by displaying a raster scan when the user clicks a button on the light pen. When the raster scan is sensed by the pen, the position of the pen is known because of a tight coupling between the pen clock and display clock. In Point & Shoot, a visual tag grid replaces the functionality of the raster scan. This technique is robust to different display technologies and loose coupling between camera and display. Point & Shoot needs only one visual tag entirely in the camera view to establish a coordinate system, but a grid is used to increase the probability of satisfying this requirement.

Continuous indirect translation with camera optical flow. The Sweep [11] interaction technique uses optical-flow image processing, which involves rapidly sampling successive images from a camera phone and sequentially comparing them to determine relative motion in the (x, y, θ) dimensions. No visual tags are required. The camera doesn't even need to be pointed at the display. To invoke the Sweep function, you rotate the joystick button downward, which acts as a clutch to indicate to the system that you are actively controlling the cursor. Pressing the joystick button inward indicates selection or dragging. The phone is waved in the air to control the cursor. You can release the clutch button to reposition your arm, which is similar to the way a mouse can be lifted to be repositioned on the desktop surface. In the Sweep mode, you can ignore the screen on the phone and focus your attention on the large display to observe cursor movement.

Figure 5. The Sweep technique uses camera input and optical flow image processing to control a cursor.

Orient

The orienting subtask involves specifying an orientation (not a position) in a coordinate system. Orientation is also not limited to graphics subtasks as it can relate to physical orientation in the real world. For example, the orientation subtask could be used to control the orientation of a security camera, a spot light, or a steerable projector.

As with position, the orientation techniques can be either discrete or continuous. Some of Foley et al.'s original graphics interactions carry over directly for smart phones including *indirect continuous orientation with velocity-controlled joystick*

9 [1] An alternative implementation of the Point & Shoot technique uses pen input instead of the cross-hair image so that the user repositions the cursor by selecting the desired position directly on the live camera image displayed on the phone screen.

and *discrete orientation with angle type-in*. The remaining techniques observed in our survey include:

Indirect continuous orientation with locator devices. The user can specify the angle of orientation by using a continuous quantifier or one axis of a physical locator. The Sweep technique supports detection of rotation around the Z-axis (coming out of the display) allowing interactions like rotating a puzzle piece in a jigsaw puzzle application, where the phone is used like a ratchet to adjust orientation. The optical-flow processing used by Sweep is also capable of detecting rotation around the X and Y-axis. However in Sweep, rotation around the Y-axis is mapped to translate position along the X-axis and rotation around the X-axis is mapped to translate position along the Y-axis.

Continuous direct orientation with camera tracking. Madhavapeddy's tagged GUI dials can be oriented using the phone camera to track movement.

Direct discrete orientation with camera image. The Point & Shoot technique supports discrete orientation along the Z-axis by rotating the camera orientation for the acquired image.

Select

In the selection subtask, the user makes a selection from a set of alternatives, for example a set of commands. One input option commonly used for this task is to make a menu selection using a cursor controlled through the positioning subtask. Instead of commands, the set of alternatives might be a collection of entities that can be operated upon. In graphics interactions these may be displayed entities, such as graphical icons that form part of the GUI composition. However, this subtask is also very relevant to ubiquitous computing interactions as a user may select a physical object to operate upon, such as selecting a lamp to adjust its setting. Many selection techniques carry over directly from Foley et al.'s earlier analysis such as *simulated pick with cursor match*, *character string name type-in* common for command prompts, or *button push–soft keys* where commands are presented as labeled buttons. The remaining selection techniques are as follows:

Direct pick of tagged objects. The E-tag project at Xerox PARC [13] investigates how tagged objects can be used to present information on a wireless mobile computer equipped with an electronic tag reader. For example, selecting a book by scanning its embedded RFID tag would bring up a web-reference to the book (perhaps allowing it to be purchased). Similar interactions have also been proposed for visual tags in the environment [12] and tagged GUI elements [7, 14] where a camera is used to acquire an image to decode the selected tag. Patel and Abowd [8] present a physical world selection method for mobile phones in which a modulated laser pointer signal triggers a photosensitive tag placed in the environment, allowing users to bring up a menu to control the object on their handheld device.

RFIG Lamps [15] allows a handheld projector to be used to select objects with photosensitive RFID tags in the physical world. The handheld projector emits a gray-code pattern that allows the tags to determine their relative position in the projected view. Waving the handheld projector around, you can navigate a cursor in the center of the projected view to select individual physical objects.

Direct pick with camera. The Point & Shoot technique can be used without a cursor to directly select items on a display, or items in the real world such as printed selections on a poster.

Direct pick by laser pointer. Myers et al. [16] proposed a multi-device technique called "semantic snarfing". This technique uses a laser pointer to make a coarse-grained selection of a screen region on a display in the environment monitored by a camera. The interaction then composes a GUI on the screen of the handheld device to make the fine-grained GUI selection.

Voice Recognition. The user speaks the name of the selected command, and a speech recognizer determines which command was spoken. The Personal Universal Controller [17] supports automatic generation of speech interfaces (as well as graphical interfaces) to issue commands to objects in the real world.

Gesture Recognition. The user makes a sequence of movements with a continuous positioning device such as the joystick, camera, trackpad, or accelerometers. For example, Patel et al. [18] used gesture recognition of accelerometer data from the handheld device to authenticate users that wanted to access data on their mobile phone through an untrusted public terminal. Using this technology, users could securely bring up data on the public terminal from their phone without removing it from their purse.

Path

The pathing task involves specifying a series of positions and orientations over time, and is thus always continuous and closed loop. Although the path task has different requirements from position and orientation, pathing adheres to the same taxonomy as the corresponding positioning and orienting techniques.

Quantify

The quantifying task involves specifying a value or number within a range of numbers. In our survey, the quantifying tasks using phone input were accomplished through a GUI using 1-D position or orientation subtasks. Another option for numeric input within a bounded range is to use tagged GUI elements [14].

Text

Text entry for mobile phones is a well-studied area as it is central to text-based mobile communications (like SMS) and personal information management functionality. We refer readers interested in this topic to previous work. [19]

Studies and evaluation of techniques

The use of mobile phones as input devices for their environment is still very new, and there have been only a handful of thorough evaluations of the different techniques [4, 11, 16]. These studies used different experimental parameters, making the results difficult to directly compare. In Figure 6, we present a comparison table for the best studied task: positioning. To create this comparison table, we combined our knowledge of the interaction techniques and the collective knowledge on their desktop computing counterparts to extrapolate rough estimates for a variety ergonomic measures. The ergonomic parameters are mostly borrowed from Foley et al.'s survey of interaction techniques.

Perceptual, Cognitive, and Motor load refer to basic psychological processes necessary for an interaction technique.

Sensitivity to distance. Since the amount or degree of separation between the display and the user is, in many cases, dynamic and unpredictable, the range of distances the interaction will support is an important design consideration. Perspective size is influenced by both display size and interaction distance. Sensitivity to distance plays a significant role when considering what types of interaction techniques suit a particular task. For example, interactions that require aiming, such as Point & Shoot, or laser pointers work best when users are closer to the target so that the target is perceptively larger. On the other hand, interactions that don't require aiming are typically not significantly affected by distance. For example, the Sweep technique is appropriate for interactions where users are further from the display so that the display is perceptively smaller.

The remaining measures are self explanatory, but further clarification can be found in [2].

SPATIAL LAYOUT OF DESIGN SPACE

The Position, Orient, and Select techniques are summarized in a Foley-style graph in Figure 7. Card, Mackinlay, and Robertson [20] point out that this format is somewhat ad hoc and that there is no attempt at defining a notion of completeness. Card then builds on the work of Buxton [21] to create a systematic spatial layout of the design space of input devices. Card's design space is well known and captures the physical properties of manual devices very well. However, it fails to capture many traits that are relevant to ubicomp interactions such as modality, or feedback as discussed in [22].

Using Foley et al.'s taxonomy, we propose a new 5-part spatial layout for mobile phone interaction tasks discussed in our survey including dimensionality, relative vs. absolute, interaction style (direct vs. indirect), supported subtasks (position, orient, and selection), and feedback from the environment (continuous vs. discrete). The resulting categorization is shown in Figure 8. Using this graphical layout, we are able to pinpoint gaps in the breadth of the interaction techniques surveyed. It should be noted that this layout is not comprehensive, and other layouts may pinpoint further gaps. An alternative layout might include phone output modalities used, which would demonstrate the sparse usage of auditory and haptic feedback in these techniques.

DESIGNING FOR SERENDIPITY

One key design consideration is the ease and speed of setting up a data connection between the phone and the environment or the device it is controlling. We note that for projects that use short range wireless communications models such as Bluetooth, visual or RFID tags can be used to encode the connection information for the environment. This creates a very low threshold of use and allows for highly serendipitous interactions.

SOCIAL ACCEPTANCE

Smart phones today are social devices. While smart phone ubiquity seems inevitable, social acceptance will influence the success of these new interactions. Remind yourself, for example, of the first time you came across a person using a wireless headset to communicate via their mobile phone. For many people, this communication technique is still awkward and strange, particularly in public places. Smart phone interaction will require users to perform particular actions and behaviors which might feel unintuitive and awkward to them. Furthermore they will perform these actions in the presence of passive or active others, both familiars and strangers. On one hand, outside observers might find these interactions disturbing or embarrassing, but on the other hand these kinds of interaction have the potential to raise your social status. One example is people who proudly flaunt their expensive smart phone in public.

CONCLUSION

Project / Author Name	Technique	Ergonomic Measures									
		Cognitive Load	Perceptual Load	Motor Load	Visual Acqui-sition	Motor Acqui-sition	Ease of Learning	Fatigue	Error Proneness	Sensitivity to Distance	Feedback
Sweep [11]	Indirect, Relative Camera	M	M	H	M	H	M	H	M	L	C
Silfverberg et al. [4]	Indirect, Relative Velocity-controlled Joystick	H	M	L	M	M	M	L	M	L	C
Remote Commander [3]	Indirect, Relative Trackpad	M	M	M	M	H	M	M	M	L	C
Rock 'n' Scroll [5]	Indirect, Relative Accelerometer	H	M	L	M	M	M	L	M	L	C
Blinkenlights [6]	Indirect, Relative Directional step keys	H	H	L	H	M	M	L	M	L	D
Point & Shoot [11]	Direct, Absolute Camera	M	H	M	H	H	M	M	H	H	D
C-blink [10]	Indirect, Absolute Camera	M	M	H	M	M	M	H	M	M	C
Madhavapeddy et al. [7]	Direct, Relative Camera	L	M	M	M	M	M	M	L	H	C
Olsen et al. [9]	Direct, Absolute Laser Pointer	L	M	M	M	M	M	M	H	H	D

Figure 6. Comparison of mobile phone-based positioning techniques (L = low, M = medium, H = high, C = continuous, D = discrete).

Smart phones provide a rich set of tools enabling us to control and interact with our environments. We have taken you through a brief structured tour illustrating the state of the art in using the smart phone as an input device. Our overview and in-depth discussion structures existing interaction techniques in a preliminary design space that helps identify the range of existing techniques. We also identify key design considerations for deploying these interaction techniques in real-world applications. This analysis is intended to help inspire applications that begin to use these technologies as well as to inform the design of future smart phone interaction techniques.

REFERENCES

1. Mark Weiser. The computer for the 21st century. *Scientific American*, 265:94–104, September 1991.
2. James D. Foley, Victor L. Wallace, and Peggy Chan. The human factors of computer graphics interaction techniques. *IEEE Comput. Graph. Appl.*, 4(11):13–48, 1984.
3. Brad A. Myers et al. Collaboration using multiple PDAs connected to a PC. In *Proc. CSCW*, pages 285–294. ACM, 1998.
4. Miika Silfverberg, I. Scott MacKenzie, and Tatu Kauppinen. An isometric joystick as a pointing device for handheld information terminals. In *Proc. GI*, pages 119–126, 2001.
5. Joel F. Bartlett. Rock 'n' scroll is here to stay. *IEEE Comput. Graph. Appl.*, 20(3):40–45, 2000.
6. http://www.blinkenlights.de.
7. Anil Madhavapeddy, David Scott, Richard Sharp, and Eben Upton. Using camera-phones to enhance human-computer interaction. In *Adjunct Proc. of Ubicomp*. Springer-Verlag, September 2004.
8. Shwetak N. Patel and Gregory D. Abowd. A 2-way laser-assisted selection scheme for handhelds in a physical environment. In *Proc. Ubicomp*, pages 200–207. LNCS 2864, Springer, 2003.
9. Dan R. Olsen and Travis Nielsen. Laser pointer interaction. In *Proc. CHI*, pages 17–22. ACM, 2001.
10. Kento Miyaoku et. al. C-blink: a hue-difference-based light signal marker for large screen interaction via any mobile terminal. In *Proc. UIST*, pages 147–156. ACM, 2004.
11. Rafael Ballagas, Michael Rohs, Jennifer G. Sheridan, and Jan Borchers. Sweep and point & shoot: Phonecam-based interactions for large public displays. In *Extended Abstracts CHI*. ACM Press, April 2005.
12. Michael Rohs. Real-world interaction with camera phones. In *2nd International Symposium on Ubiquitous Computing Systems (UCS 2004)*, pages 74–89, Tokyo, Japan, 2005. LNCS 3598, Springer.
13. Roy Want et al. Bridging physical and virtual worlds with electronic tags. In *Proc. CHI*, pages 370–377. ACM, 1999.

14. Michael Rohs. Visual code widgets for marker-based interaction. In *IWSAWC'05: Proc. ICDCS Workshops*, pages 506–513, June 2005.

15. Ramesh Raskar et. al. Rfig lamps: interacting with a self-describing world via photosensing wireless tags and projectors. *Trans. Graph.*, 23(3):406–415, 2004.

16. Brad A. Myers et al. Interacting at a distance: measuring the performance of laser pointers and other devices. In *Proc. CHI*, pages 33–40. ACM, 2002.

17. Jeffrey Nichols et. al. Personal universal controllers: controlling complex appliances with guis and speech. In *Proc. CHI*, pages 624–625. ACM, 2003.

18. Shwetak N. Patel et al. A gesture-based authentication scheme for untrusted public terminals. In *Proc. UIST*, pages 157–160. ACM, 2004.

19. Lee Butts and Andy Cockburn. An evaluation of mobile phone text input methods. In *Proc. CRPITS*, pages 55–59. Australian Computer Society, Inc., 2002.

20. Stuart K. Card, Jock D. Mackinlay, and George G. Robertson. A morphological analysis of the design space of input devices. *Proc. Trans. Inf. Syst.*, 9(2):99–122, 1991.

21. William Buxton. Lexical and pragmatic considerations of input structures. *SIGGRAPH Comput. Graph.*, 17(1):31–37, 1983.

22. R. Ballagas, M. Ringel, M. Stone, and J. Borchers. iStuff: A Physical User Interface Toolkit for Ubiquitous Computing Environments. In *Proc. CHI*, pages 537–544. ACM, April 2003.

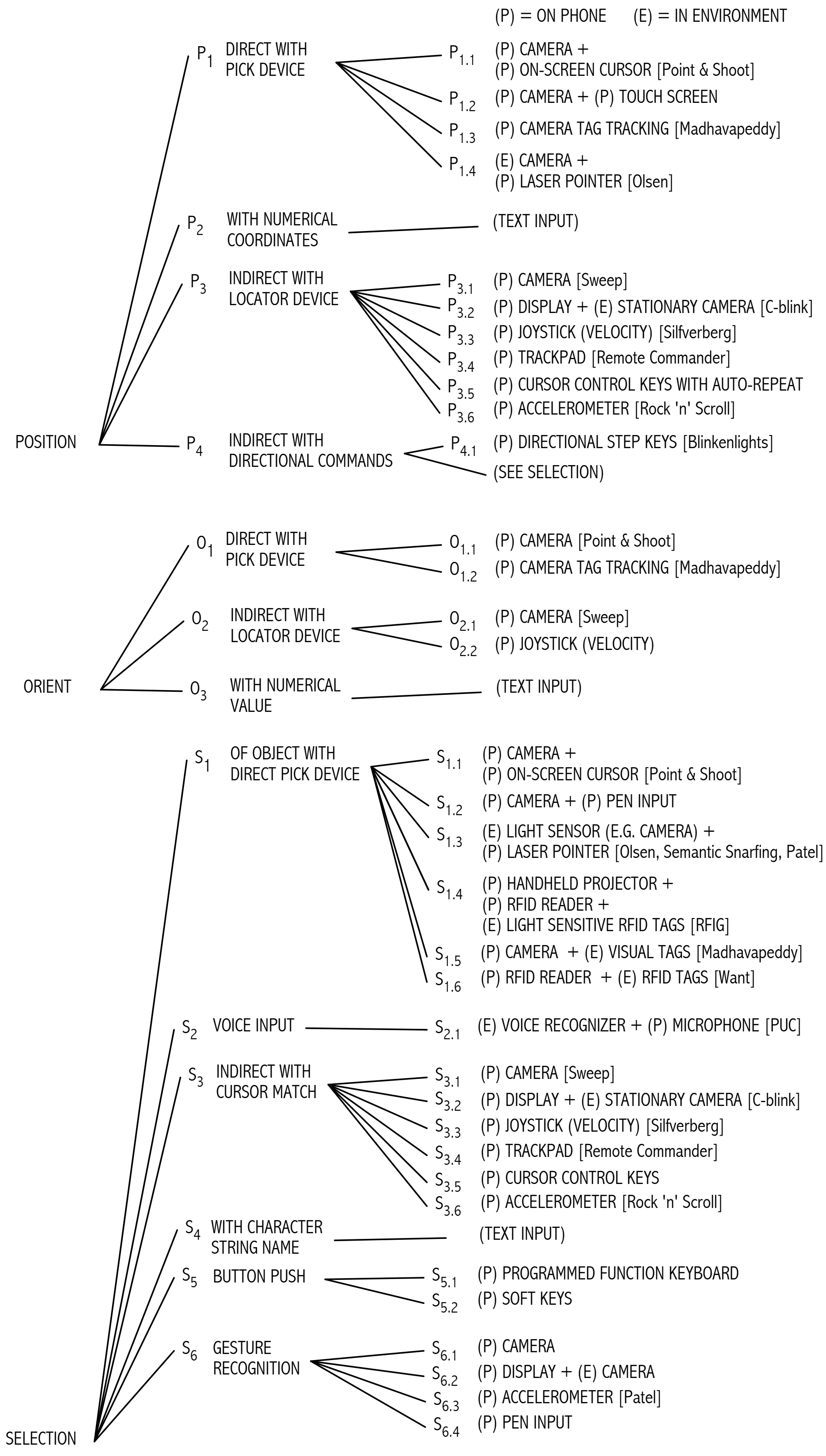

Figure 7. Summary of position, orientation, and selection techniques using a smart phone as an input device.

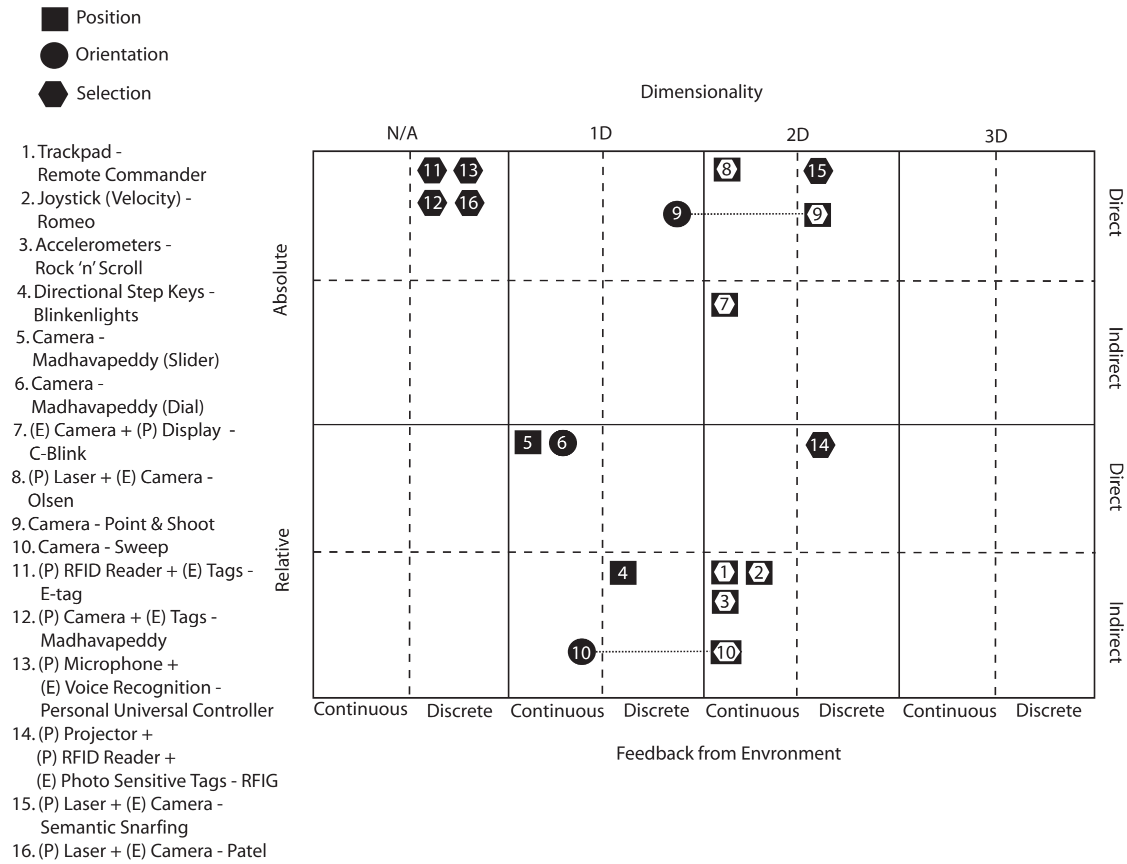

Figure 8. Classification of different mobile phone interactions that have been implemented in the projects surveyed. Inspection of the diagram reveals opportunities for future work – for instance, developing interaction techniques that support 3-D orientation, or 3-D positioning.