Sweep and Point & Shoot: Phonecam-Based Interactions for Large Public Displays

Rafael Ballagas
Media Computing Group
RWTH Aachen University
Aachen, Germany
ballagas@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

ABSTRACT
This paper focuses on enabling interactions with large public displays using the most ubiquitous personal computing device, the mobile phone. Two new interaction techniques are introduced that use the embedded camera on mobile phones as an enabling technology. The "Point & Shoot" technique allows users to select objects using visual codes to set up an absolute coordinate system on the display surface instead of tagging individual objects on the screen. The "Sweep" technique enables users to use the phone like an optical mouse with multiple degrees of freedom and allows interaction without having to point the camera at the display. Prototypes of these interactions have been implemented and evaluated using modern mobile phone technologies. This proof of concept provides a performance baseline and gives valuable insights to guide future research and development. These techniques are intended to inspire and enable new classes of large public display applications.

Categories & Subject Descriptors: H.5.2 [Information Interfaces and Presentation]: User Interfaces – evaluation/methodology, haptic I/O, input devices and strategies, interaction styles

General Terms: Human Factors; Experimentation

Keywords: mobile phones; camera phones; visual codes; large displays

INTRODUCTION
Large-scale electronic displays are increasingly found in public places like airports, train stations, shopping malls, and museums. Unfortunately, most of today's public electronic displays are not interactive, making it difficult to capture interesting information or to influence the display's content. Large displays in public places are often inaccessible for direct touch-based interaction since they need to be protected from vandalism, and installing dedicated hardware for interaction can be prohibitively expensive.

Personal devices, like camera-equipped mobile phones, open up new possibilities in this domain. People are comfortable with using their own devices and they usually have their mobile phones with them. Cameras provide a powerful input channel and the phones can connect to situated displays via wireless technologies, such as Bluetooth. Potential applications include interactive art, games, bulletin boards, and advertising.

In this paper, we present *Sweep* and *Point & Shoot*, two phonecam-based interaction techniques for large public displays. We present the results of a user study which examines prototype implementations and then conclude with a discussion of future research and development.

RELATED WORK
Remote Commander [5] enables individuals to use pen input on a PDA to control a cursor, which requires two hands for operation. Since Remote Commander focuses on semi-public display environments, it does not provide any mechanisms for spontaneous interaction. The C-Blink [4] system uses the phone screen as an input device. The user runs a program on the phone to rapidly change the hue of the phone screen and then waves the phone in front of a camera mounted on the large display. This tracks the position of the phone to control a cursor on the large display. Madhavapeddy et al. [3] introduce techniques that use visual tags known as SpotCodes. Interaction involves using a phonecam to scan tags or to manipulate tagged GUI widgets. The main distinction of our design is that it can be used to select any arbitrary pixel, where Madhavapeddy's work only allows the user to select or manipulate tagged objects.

INTERACTION TECHNIQUES
We developed two complementary interaction techniques using the phonecam: *Sweep* based on optical movement detection and *Point & Shoot* based on visual tag sensing.

Sweep
The *sweep* technique allows the phonecam to be used like an optical mouse. Using optical flow image processing, the phonecam samples successive images and then sequentially compares them to determine relative motion in the (x, y, θ) dimensions thus allowing the camera to be used as a three degrees of freedom input device. Since optical flow processing is performed directly on the phone rather than on the computer driving the display, *sweep* can scale to a high num-

CHI 2005, April 2–7, 2005, Portland, Oregon, USA.
ACM 1-59593-002-7/05/0004.

Figure 1. The sweep technique can be used to control a cursor on the large display like an optical mouse.

ber of users. In our prototype implementation, a high latency occurs (200ms) when calculating movement parameters from successive images. However, mobile computing trends indicate that future phones will have the processing power to create more fluid interactions. To invoke the *sweep* function, users vertically push and hold the phonecam joystick which acts as a clutch to activate movement detection. To control the display's cursor, they wave the phone in the air. Users release the joystick to reposition their arm, similar to repositioning a mouse on a desktop. Motion can be recognized with any unobstructed camera image, meaning that users can point the camera **anywhere**, even at the floor, allowing for a more comfortable arm posture. In the *sweep* mode, users can ignore the phone's screen and focus their attention on the large display to observe the cursor movement.

Point & Shoot

The *point & shoot* interaction technique is illustrated in Figure 2. This technique needs visual codes [6] to establish an absolute coordinate system on the display surface. The detection of codes is invariant to perspective distortion arising from the mobility of the phone camera (as shown in Figure 3). The precise pixel targeted on the large display is calculated on the phone before a "selection" is wirelessly issued to the display. The minimum requirement is that one visual code must be in the camera image during selection. Integrating visual codes directly into the application layout works, but diminishes overall visual appeal. Alternatively, as in Figure 2, the codes can be temporarily flashed on the display during selection, but this may disturb others in multiuser scenarios. Future display technologies will allow the codes to be displayed in infrared so that they are recognizable by the camera, but invisible to humans.

To make a selection, users aim the phonecam at the desired target on the large display. The large display contents appear on the phonecam screen, which acts as a view finder. As users move the phone, the screen is continuously updated with a live camera image. A cross-hair in the center of the image allows for aiming. The user presses the phonecam joystick to "shoot" a picture and issue the resulting selection

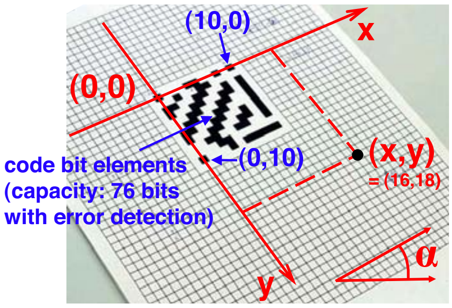

Figure 3. Each visual code has its own local coordinate system that is invariant to perspective distortion.

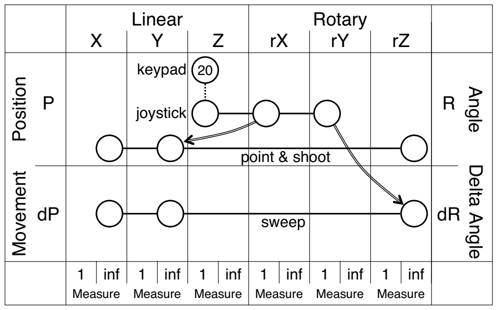

Figure 4. Card design space classification [1] of our camera-based interaction techniques.

to the display. For *point & shoot*, the user's locus of attention switches between the phone screen and the large display.

We also use visual codes to encode the public display's Bluetooth address. Users merely take a picture of the display (and it's visual code) to automatically establish a wireless connection. This results in a very low threshold of use and allows for highly serendipitous interactions.

Input Device Classification

As shown in the classification in Figure 4, the interaction properties of the device become richer by adding the camera as a relative and absolute movement sensor. The three horizontally connected circles labeled *sweep* correspond to the (x, y, θ) dimensions. In our implementation, relative rotation around the X axis (dR:rX) is equivalent to linear Y motion and relative rotation around the Y axis (dR:rY) is equivalent to linear X motion. This means that for *sweep*, bending the wrist is equivalent to moving the whole arm. The three horizontally connected circles labeled *point & shoot* represent absolute position sensing, which provides the X and Y position and the state of rotation around the Z axis.

STUDY

Using a within-subjects design, we asked users to complete a multidirectional tapping test based on ISO 9241-9 [2] using *point & shoot* and *sweep* as well as the simple joystick found on the Nokia 6600 mobile phone used in this study. Before each test, users were allowed to practice with the

Figure 2. Point & shoot technique: (Left) The phone display is used to aim at a puzzle piece on a large display. (Middle) Pressing the joystick indicates *selection* and a visual code grid flashes on the large display to compute the target coordinates. (Right) The grid disappears and the targeted piece highlights to indicate successful selection.

technique until they felt comfortable. We used Panasonic 60-inch plasma displays for our study and users performed the tests while standing one meter away from the display. For each test, users were asked to select targets as quickly as possible as they highlighted around a circle. Targets were equally spaced and users performed tests with a single index of difficulty of 3 (distance of 612 pixels and width of 87 pixels). We measured task completion times as the time between successful selection of sequential targets. Any selection outside of the target region was interpreted as an error. We alternated the order of tests for each participant to minimize learning and fatigue effects.

Before the tests, users filled out a background questionnaire. After testing each input technique, users filled out a subjective questionnaire concerning the device comfort and performance based upon [2]. At the end of the three tests, users completed a questionnaire to express their opinions about each technique. Our group of 10 participants was diverse: 6 United Kingdom, 3 non-UK Europe, and one non-European. Most were 26-35 years old (5) or 17-25 (4), and one was over 45. 7 were right handed. 6 were men. The participants were equally split in pursuing technical and non-technical fields of study. Most had no experience in using camera phones.

RESULTS

The task completion time results are summarized in Figure 5. All three input techniques exhibit a large positive skew, meriting a log transformation of the data to more closely resemble a normal distribution for subsequent analysis. ANOVA showed that the results disprove the null hypothesis ($p < .01$). Tukey's post-hoc analysis showed that *sweep* is significantly slower than both *point & shoot* ($p < .05$) and the joystick ($p < .05$). However, there is no significant difference between the joystick and *point & shoot* for task completion times.

The results for error rates are summarized in Figure 6. Again the data exhibits a large positive skew, and a log transformation was employed before analysis. ANOVA showed that the results disprove the null hypothesis ($p < .01$). Tukey's post-hoc analysis showed that the *point & shoot* error rate was significantly different from both *sweep* ($p < .05$) and the joystick ($p < .01$). No significant difference appeared

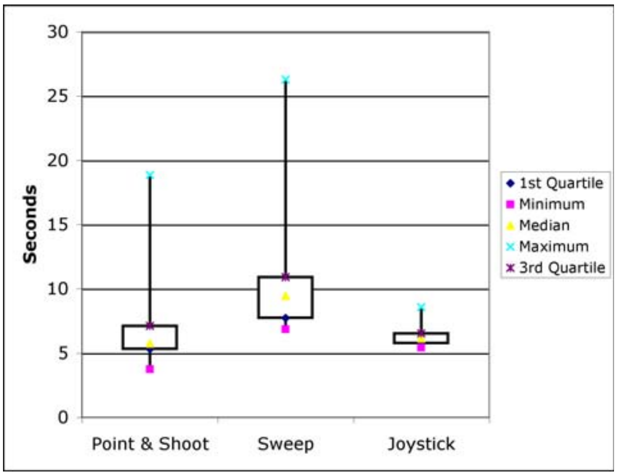

Figure 5. Task completion time results for multidirectional tapping test grouped by input technique.

between *sweep* and joystick for error rates according to Tukey's analysis, but the more powerful Student Newman-Keuls analysis showed that the difference was in fact significant ($p < .05$).

The results of the subjective questionnaire were analyzed using ANOVA as suggested by [2]. The significant differences between the input devices for these categories is summarized in Figure 7.

FINDINGS

One older user was unable to use *point & shoot* for two reasons. First, the user was unable to distinguish the small white cursor on the phone display. Secondly, trembling hands made aiming difficult. Another user had difficulty keeping the hand steady, but was able to overcome this by adopting a two-handed grasping technique.

Point & shoot performs poorly for both error rates and reliability, partly because of a quirk in our implementation. When the user issues a *point & shoot* command, the camera display freezes as the camera switches to high-resolution

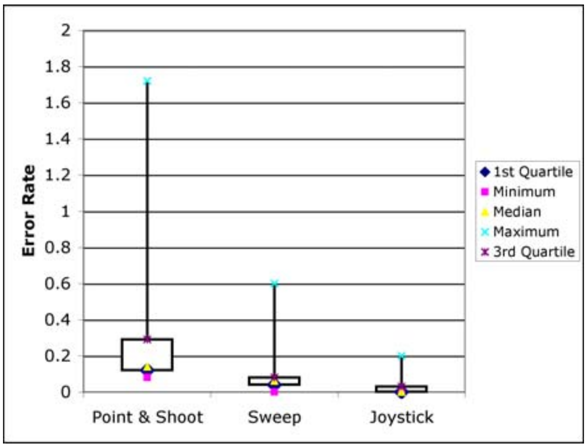

Figure 6. Error rates for multidirectional tapping test grouped by input technique.

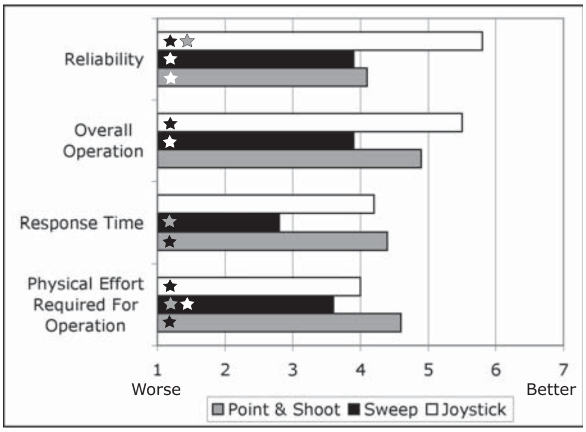

Figure 7. Significant results from the the subjective ratings of the various input mechanisms. The color of the stars indicates the technique for which a statistically significant relationship exists.

mode. This falsely indicates to users that they have taken a picture. Concurrently, a the "flash grid" command is sent to the public display. The phone waits for a confirmation that the grid is visible before acquiring the high-res image. In low-res mode, the phone's cursor may be centered on the target indicating a successful selection of the frozen on-screen image, but the actual center of the image may drift off the target by the time the high-res image is acquired.

One user likened one-handed operation to "trying to hold onto a bar of soap". Clearly, an alternative form factor would improve user experience. Although fatigue is not statistically significant, we note that at the end of the final test several participants complained that their thumbs hurt. We expect then, that there is a physical limit to the length of time users can interact with a large display using a phonecam. This matches our envisioned usage model for large public displays where shorter-term interactions prevail.

Screen size and distance play a significant role when considering whether *point & shoot* or *sweep* suit a particular task. Given that a target area remains the same, *point & shoot* seems particularly well suited for interactions where users are close to the display so that the display is perceptively larger. *Sweep* seems better suited for interactions where users are further from the display so that the display is perceptively smaller.

CONCLUSIONS

Prototyping *sweep* with today's phones resulted in high task completion times and consistently low scores in the qualitative evaluation. However, mobile phone trends indicate that phones will soon be technically capable of supporting fluid interactions. By developing this proof of concept now, we hope to establish a performance baseline for future incremental improvements and to drive technology to meet the demands of this interaction. Also, we hope to inspire new applications and metaphors that use *sweep* as a foundation.

Point & shoot was as good as the joystick for task completion times. An additional advantage of *point & shoot* is that it can be used to select items in the real world, such as physical objects and printed text, because the cursor is on the phone screen. Conversely, the joystick relies on the environment for cursor feedback. Also, we can use *point & shoot* when no joystick is available. Lastly, visual codes can be used to acquire information from the world [6].

ACKNOWLEDGEMENTS

The research is supported in part by the UK EPSRC (GR-S08848/01) and the Equator IRC (GR/N15986/01).

ADDITIONAL AUTHORS

Jan Borchers, RWTH Aachen University, Media Computing Group, Aachen, Germany, borchers@cs.rwth-aachen.de.

REFERENCES

1. Card, S. K., Mackinlay, J. D., and Robertson, G. G. A morphological analysis of the design space of input devices. *ACM Trans. Inf. Syst. 9*, 2 (1991), 99–122.
2. ISO. Ergonomic requirements for office work with visual display terminals (VDTs) - Requirements for non-keyboard input devices. *ISO 9241-9* (2000).
3. Madhavapeddy, A., Scott, D., Sharp, R., and Upton, E. Using camera-phones to enhance human-computer interaction. *Adjuct Proc. of Ubicomp*. Springer-Verlag (2004).
4. Miyaoku, K., Higashino, S., and Tonomura, Y. C-blink: A hue-difference-based light signal marker for large screen interaction via any mobile terminal. *Proc. UIST*. ACM (2004), 147–156.
5. Myers, B. A., Stiel, H., and Gargiulo, R. Collaboration using multiple PDAs connected to a PC. *Proc. CSCW*. ACM Press (1998), 285–294.
6. Rohs, M. Real-world interaction with camera-phones. *Proc. UCS*. IPSJ Press (2004).