Direct Pointer: Direct Manipulation for Large-Display Interaction using Handheld Cameras

Hao Jiang
Tsinghua University
Beijing, China
h-jiang@mails.
tsinghua.edu.cn

Eyal Ofek
Microsoft Research Asia
Beijing, China
eyalofek@microsoft.com

Neema Moraveji
Microsoft Research Asia
Beijing, China
neemam@microsoft.com

Yuanchun Shi
Tsinghua University
Beijing, China
shiyc@tsinghua.edu.cn

ABSTRACT

This paper describes the design and evaluation of a technique, Direct Pointer, that enables users to interact intuitively with large displays using cameras equipped on handheld devices, such as mobile phones and personal digital assistant (PDA). In contrast to many existing interaction methods that attempt to address the same problem, ours offers direct manipulation of the pointer position with continuous visual feedback. The primary advantage of this technique is that it only requires equipment that is readily available: an electronic display, a handheld digital camera, and a connection between the two. No special visual markers in the display content are needed, nor are fixed cameras pointing at the display. We evaluated the performance of Direct Pointer as an interaction product, showing that it performs as well as comparable techniques that require more sophisticated equipment.

Author Keywords

Large display, shared display, remote pointing, interaction.

ACM Classification Keywords

H5.2. Information interfaces and presentation (e.g., HCI): User Interfaces.

INTRODUCTION

Large displays, from projection-screens used in meeting rooms to train station billboards, are effective at displaying information but difficult to interact with. Displays that are protected from physical interaction (e.g., to protect from vandalism) have the same problem. In situations like these, some type of remote interaction technique is desired. Interacting with, and not merely pointing to, the content on the display is the problem addressed in this paper.

Camera-phones are perhaps the best candidate for remote interaction devices in such situations. They are near-ubiquitous and contain the needed components to become digital pointing devices: designed to be comfortable to hold, and equipped with wireless connections.

Rather than present a new interaction technique that requires a high learning effort, we attempt to enable an intuitive pointing technique: directly pointing to the desired target. The result, Direct Pointer, is a system that controls the cursor on the display by analyzing the view of the camera. A description of this system is followed by a performance evaluation of this technique in a controlled environment using a standard test, showing that ours is comparable with other interaction methods including laser pointers. This paper concludes with a discussion of some possible applications that emerge as result of this research.

RELATED WORK

Earlier works tried to extend the desktop interaction scheme to large displays using peripherals such as a remote mouse with an isometric joystick [3]. Though it does work, this method is not ideal for ergonomic reasons: it is difficult to use with only one thumb.

Sweep [1] is a rather natural interaction technique that lets users move a camera-phone along the desired direction of the cursor motion. It mimics the operation of a desk mouse in midair. However, by comparing consecutive frames of the camera, it offers only indirect control of the cursor position. Further, depending on the depth of objects in the camera images, same-distance camera motions may generate different distances for the cursor to move, making control difficult.

Point & Shoot [1] deduces the user's desired object of selection by displaying visual codes. By recognizing those visual codes in the camera image, the position of the cursor relative to those landmarks is calculated. This technique allows pointing directly at desired targets. However, the stated implementation does not support real-time visual feedback of the cursor, forcing users to concentrate their attention on the display of the camera (to align the codes) rather than the large display. Further, the visual codes on the display, even if they appear only for a brief time during the selection process, do contaminate the display for other viewers, making it inappropriate for shared displays.

Leveraging the intuitiveness of a laser pointer, some systems use a pre-calibrated, fixed camera to visually track the bright dot on the display, creating a remote pointer [6, 7]. These systems have the advantages of natural interaction and immediate feedback, but a fixed camera must be installed for each display, which is not always feasible. What's more, security is not guaranteed in usages where user authentication is needed because anonymous users could also point to the display.

Other works include C-Blink [5] and VisionWand [10]. C-Blink uses a hue-difference-based method to track the position of the cell phone display and receives data from it. However, due to current performance of TFT screens and cameras, its throughput in data transmission is much smaller than traditional wireless techniques. VisionWand uses two cameras to build 3D position and direction of the wand by tracking its colorful endings. This setup requires even more effort than the regular laser pointer tracker.

Missing from earlier research is a system that allows for all the advantages of a common laser pointer along with the advantages of modern mouse cursors.

DESIGN

Direct Pointer allows direct manipulation of the cursor with continuous visual feedback, closely resembling the laser pointer. This is possible with the use of common components such as a display and a handheld camera, and without a fixed camera as in the laser pointer system. Our method uses the position of the cursor on the display as the sole source of input. There is no need for special tags on the display or around it and, it works independent on the display content. What's more, since the interaction is performed via the wireless channel of the camera, security of interaction is also guaranteed.

The present system is based on a closed-loop feedback between the handheld device and the display. The loop is constructed by capturing a view of the screen with the handheld camera and sending the data via wireless channel to the server. The server then updates the cursor position (see Figure 1).
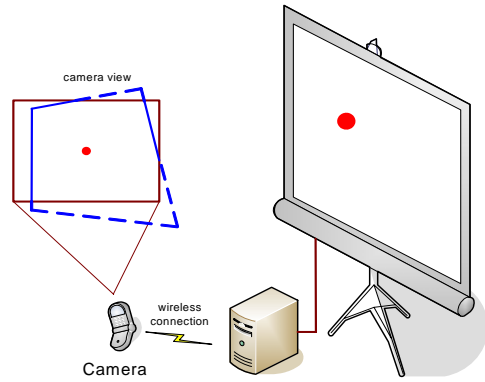

Figure 1. Deployment of Direct Pointer.

When a cursor is displayed in sight of the handheld camera, it is viewed by the camera at the center of the frame. If the cursor is identified at a different location in the frame, its position should be updated so that it will move back to the center of the camera frame.

The system runs in the following stages:

(1) The display shows a cursor.

(2) The handheld camera captures a video frame, and locates the cursor position in this frame (see below for implementation). This position is sent back to the display.

(3) The server computer gets the feedback from the camera. According to the cursor offset from the center of the camera frame, a new corrected cursor position is calculated.

(4) Return to stage 1.

Identifying the cursor in the camera image

Cursor identification is calculated on board the handheld device by its processing unit. Different features of the cursor can be used during the identification, such as color, shape, unique motion or a combination of those.

When using motion as a feature, we assume that the motion of the cursor is different from the rest part of the display.
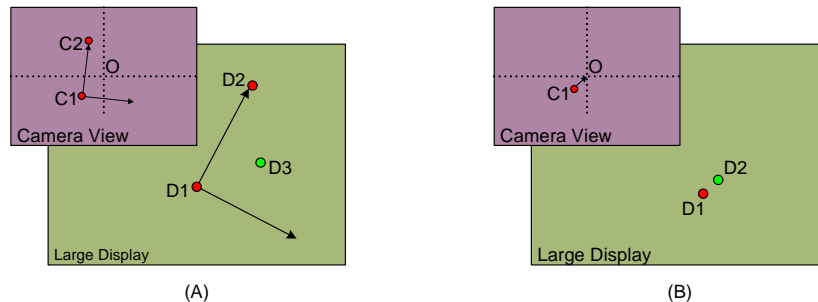

Figure 2. Estimating the new position of cursor. (A): The last two positions of cursor in both planes (C1 and C2, D1 and D2) define a 2-D coordinate system each. Using the translation, rotation and scale transformation between them, the center (O) of the camera view is mapped to the new cursor position (D3) in the display. (B): If the last two positions of cursor are too close to generate stable scale, only the last one is used to build a translation mapping. The displayed cursor is shifted (from D1 to D2), by its offset from the center (O) in the camera view.
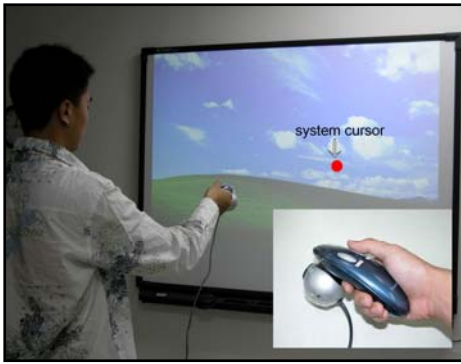
Figure 3. Prototype of Direct Pointer.

Figure 4. Multi-directional tapping test.

First, compensation for the motion of the background in the camera frame is needed. To do so, we identify several interest points between consecutive frames, and estimate the affine transformation between them [8]. Then, after warping one frame to the other according to the transformation, the cursor can be easily detected, as an area of difference between the frames.

Estimating the new position of cursor

Earlier works use visual tags positioned at known points on the screen boundary [9] or in the display content [1] to generate a global mapping between the screen and the camera image. Such methods rely on additional data that might either interfere with the display content or be hard to get (when the camera is seeing only part of the screen, the boundaries cannot be identified).

The present system maintains an approximation of the mapping between the screen and the camera image at real time, using only the location of the cursor. Special markers on the display or on the screen boundaries are not required (in fact, we may not see the boundaries in the view of the camera). Instead of estimating the exact projective transformation, we use the last two positions of the cursor to generate a translation, rotation and scale transformation between them (see Figure 2). By mapping the center of the camera frame to the screen, using this transformation, we find the new position of the cursor.

Although the method above is only an approximation, and not the full description of the transformation between the camera and the screen, when applying it to the close-loop feedback, the cursor would gradually converge towards the correct position. When the motion between consecutive frames is very small, which means the distance between the last two positions of the cursor is very small, the estimation of that transformation may be very noisy. To handle this problem, we simplify the mapping even further to translation only, and shift the displayed cursor according to its offset from the center in the camera view. Figure 2 illustrates this method.

EVALUATION

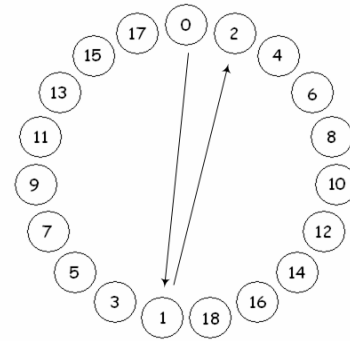To evaluate the performance of Direct Pointer, a Pentium 4 3.2GHz PC was used to control the display content, and we connected, via wire, to it a Logitech QuickCam Pro 4000 webcam. The camera grabs 30 frames with the resolution 320x240 per second. We combined a Logitech Cordless Presenter with the webcam to perform the basic clicking actions (see Figure 3). For this experiment, the handheld processing, was actually done on the display computer.

A multi-directional tapping test based on ISO 9241-9 [2] was used in our experiment. The numbers marked in Figure 4 show the order of the targets. A block ends when the target returns to 0, so each block consists of 19 trails.

Fourteen subjects (9 male), aged 19-40, participated in the experiment. All participants sat 3.6 meters away in front of a 70-inch 1280x1024 display. A red circle with the diameter 48 pixels was used as the cursor for detection. The target width was 72 pixels and the distance between targets was 140, 340, 540, 740, and 940 pixels, so the index of difficulty was 1.56, 2.52, 3.09, 3.50 and 3.81 respectively. Each subject tested 4 blocks for each index of difficulty, so this experiment was a (19 trails) x (4 blocks) x (5 indices of difficulty) x (14 participants) factorial design. The total number of clicks was 5320.

Before the experiment started, all subjects were asked to practice with the Direct Pointer until they felt comfortable. A subjective questionnaire based on [2] was given to the participants after the experiment was done to gather the qualitative feedback.

RESULT & DISCUSSION

Since no other interaction methods were involved in this experiment, we are more concerned about the throughput of our Direct Pointer. Direct Pointer's throughput along different indices of difficulty is 3.88, 3.52, 2.72, 3.01 and 2.92 respectively (see Figure 5). The average is 3.21. This throughput is compared with other interaction techniques in the literature we found [4, 6], and those techniques were also evaluated based on ISO 9241-9.

The result of our first implementation of the Direct Pointer is promising. Compared with other interaction systems evaluated with this standard test, we get similar performance (see Figure 6).
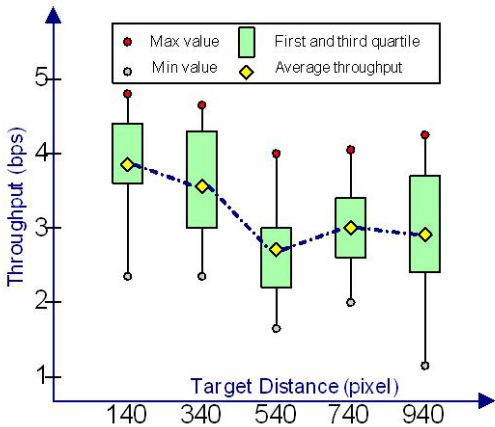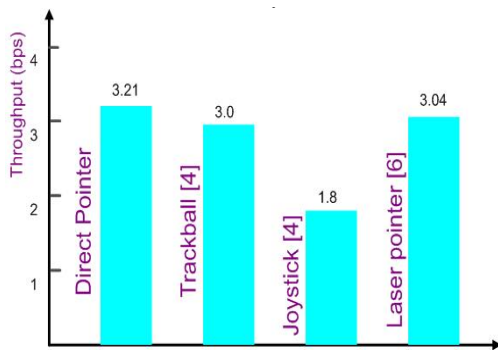
Figure 5. Throughput of Direct Pointer.

Figure 6. Comparison of throughput within devices.

In Figure 5, a decrease of throughput was observed when the distance between targets became wider. Several participants also complained that they had to delicately hold the camera to "feel safe" when performing long-distance tasks. This is perhaps due to the form-factor of our prototype because we also find that for long-distance targets, participants made a fast motion of the camera near the target followed by minor adjustments. Subsequent work on this system should address this.

From the questionnaire, most questions about operation scored in the middle of the 1-to-7 scale. A poor score in the "smoothness during operation" question (mean=2.78, SD=0.8) is the result of the occasional "fly away" bug of the cursor in our prototype. This bug is caused by the incorrect calculation of mapping due to the inherent delay of the camera. Using a camera with a higher frame rate and less delay would alleviate this problem and is perfectly feasible.

Average physical fatigue assessments appeared moderate for arm fatigue (mean=3.8, SD=0.8) and finger fatigue (mean=5.6, SD=1.3), but rather negative for wrist fatigue

(mean=2.3, SD=0.7). We attribute this to our prototype's weight (204g), something alleviated easily as it would be combined into a camera-phone or mobile device in future.

CONCLUSION & FUTURE WORK

This work shows the potential of a remote pointing device that requires no fixed camera or changes in the display to achieve cursor control with instant feedback and guaranteed security. This technique opens up new application and interaction opportunities with public and private displays such as billboards, home theaters, presentations, and more.

Future work includes the improvement in constructing the transformation where fast camera motion is more considered to make the mapping more accurate. A multi-user interaction scheme with Direct Pointer is also an interesting topic.

REFERENCES

1. Ballagas, R., Rohs, M. and Sheridan, J.G. Sweep and Point & Shoot: Phonecam-Based Interactions for Large Public Displays. Proc CHI'05, (2005).

2. ISO. Ergonomic requirements for office work with visual display terminals (VDTs) - Requirements for non-keyboard input devices. ISO 9241-9, (2000).

3. MacKenzie, I.S. and Jusoh, S. An Evaluation of Two InputDevices for Remote Pointing. Proc EHCI'01, (2001).

4. MacKenzie, I.S., Kauppinen, T. and Silfverberg, M. Accuracy Measures for Evaluating Computer Pointing Devices. Proc CHI'01, (2001).

5. 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'04, (2004).

6. Oh, J.-Y. and Stuerzlinger, W. Laser Pointers as Collaborative Pointing Devices. Proc Graphics Interface'02, (2002).

7. Olsen, D.R.J. and Nielsen, T. Laser Pointer Interaction. Proc CHI'01, (2001).

8. Shi, J. and Tomasi, C. Good Features to Track. Proc CVPR'94, pages 593-600, (1994).

9. Sukthankar, R., Stockton, R. and Mullin, P. Smarter Presentations: Exploiting Homography in Camera-Projector Systems. Proc ICCV'01, (2001).

10. Xiang, C. and Ravin, B. VisionWand: Interaction Techniques for Large Displays using a Passive Wand Tracked in 3D. Proc UIST'03, (2003).