Image Based Smartphone Interaction with Large High Resolution Displays

Lynn Nguyen, Jürgen P. Schulze*
University of California at San Diego, La Jolla, CA

1 Introduction

We investigate the practicality of using smartphones to interact with large high resolution displays. We do not intend to find the spatial location of the phone relative to the display, but we want to identify the object a user wants to interact with through image recognition. The interaction with the object itself is done with the smartphone.

Large high-resolution displays, often implemented by tiling a wall with large LCD displays, are getting increasingly popular. While they are great for viewing high resolution data, they can be difficult to interact with because of their size. Common methods of interaction include using a desktop control station with a standard mouse, a gyroscopic presentation mouse, or even a 3D tracking system. Unfortunately, these methods are not without their shortcomings. Using a standard or gyromouse is bothersome because the pixel resolution of the mouse is lower than that of the display wall. And because the cursor position of the mouse is independent of the location of the user it is easy to lose track of the cursor. The tracking system is able to address these issues but can be quite expensive. Another issue with many of the current navigation methods is that they are designed for a single user.

Many previously developed smartphone-wall interaction methods assume the user wants a precise location on the display [McCallum09; Jeon10]. However, this level of granularity is not required of all applications. Often, the data on the display consists of multiple objects. We can think of interaction with the display in two parts: at a coarse grain level where the user needs to determine the object to interact with, and at fine grain level where the interaction is localized to the object, and executed on the smartphone. Our prototype implements a scenario in which a set of images is being viewed on a display wall, and users can annotate the images with their phones. The coarse grain interaction is determining which image to focus on while the fine grain interaction is the note taking on the selected image.

2 System Description

Our prototype uses an HTC Aria Android phone to select one of multiple photographs on Calit2's AESOP wall, a 4 by 4 array of 46" monitors. We use a computer vision algorithm to match what the phone sees to one of the displayed photographs to determine which image the user selected. The system uses a client-server model communicating via TCP/IP. The server is outfitted with a database to store user comments on images.

The smartphone continuously sends 768×432 pixel image frames (query images) of what its camera sees at a rate of two frames per second to the server. The phone does not do feature extraction since that is more efficient on the server side. The only processing the client does is converting the query image to JPEG and scaling it to a size the server requests. The server receives query images from the client and extracts features with the SURF algorithm from the OpenCV library. Once these features are extracted, a matching image is determined using a nearest neighbor algorithm provided by the FLANN library.

When running the client application on the phone, the user initially

*e-mail: lkn001@ucsd.edu, jschulze@ucsd.edu

Figure 1: *User selecting an image on a large display with a phone.*

sees the image from the camera, superimposed with some of the EXIF metadata of the selected photograph. Additionally, the image on the wall will be highlighted, to give the user verification that the correct image has been recognized, see Figure 1. In our prototype images turn from grayscale to color when highlighted (but the image recognition works just as well with color images). On a finger tap, a menu will appear at the bottom of the screen, which offers further information about the image: view a description of the image, see on a map the location where the image was taken, as well as view or make comments on the image selected.

We extracted the features from the photographs using the SURF algorithm, whose rotation- and scale-invariance proved to be crucial for this application in which the phone is hand held at odd angles. In order to obtain the actual matched image, we compare the query image against every displayed photograph. The comparison is a nearest neighbor algorithm provided by FLANN. For each photograph, a match percentage is determined by the number of matched descriptors in the query image over the total number of descriptors in the query image. The highest match percentage obtained determines the selected photograph. In order to reduce the probability of false matches, this percentage must be higher than an empirically determined threshold.

There are many benefits to using a smartphone as an interaction device for a large display. The main advantage of our system over previous work is that it does not rely on additional infrastructure such as a tracking system or visual markers to operate. The many features of the phone allow multiple users to select objects on the display in an intuitive manner while offering many more functions than just selecting. One thing to note is that in our current implementation, duplicate or very similar images that the user wants to treat as distinct entities will not be treated correctly by the system. A possible solution would be to use neighboring image information, or motion of the device.

References

S. Jeon, J. Hwang, G. J. Kim, and M. Billinghurst. Interaction with large ubiquitous displays using camera-equipped mobile phones. Personal Ubiquitous Computing, 14(2):83-94, 2010

D. C. McCallum, P. Irani. ARC-Pad: Absolute+Relative Cursor Positioning for Large Displays with a Mobile Touchscreen. In UIST'09, pp. 153-156, 2009