A Comparison of Input Methods for Zoom: Touch-based, Accelerometer-based and Facial Tracking

Vincent Chu
Dept. of Electrical Engineering and Computer Science
York University
Toronto, Ontario, Canada M3J 1P3
cse13261@cse.yorku.ca

ABSTRACT
To be added later.

Keywords
Android, lean, facial-tracking, accelerometer, zoom

INTRODUCTION
Mobile devices such as smartphones and tablets commonly have smaller displays than traditional desktop computers. This constraint often results in content that is either too large to be viewed fully on-screen at any given time or has details that are too indistinguishable when scaled to fit. In order to overcome these challenges, implementers conventionally utilize two touch gestures to manipulate the content displayed. Users can drag their fingers across the screen to move the content up, down, left and right. By pinching two fingers together or moving them apart, users can increase or decrease the scaling of the content revealing more details in a particular region or more of the content as a whole while trading off finer details.

It is commonplace for mobile devices to also be equipped with cameras and a multitude of sensors. These cameras and sensors enable possibilities for novel, new solutions to these problems. This paper will focus on the front-facing camera and the accelerometer of the devices.

An accelerometer is a sensor that measures the proper acceleration or g-force exerted on a device. It provides data about the motion of the device. Tilting or lateral moving the device in a vertical or horizontal direction parallel to the screen, pitching and rolling respectively, are often mapping to horizontal and vertical movement on the plane of the screen. A more rarely used movement is to move the device upward and downward perpendicular to the display, in the z-axis. See Figure 1. This movement could be used to change the height or zoom of the display. A downward movement could decrease the zoom level, while an upward motion could increase the amount of zoom.

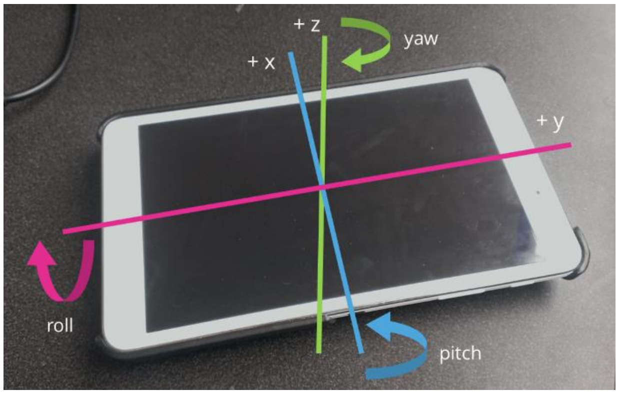

Figure 1 The axes of the *Samsung Galaxy Tab 4* (pitch, roll, and yaw)

When a user has difficulty seeing an object, he or she naturally leans in towards the object to better inspect it [3]. Using the front-facing camera to track the user's face, a computation of the distance between the user's eyes can be used to infer the user's distance from the screen. As the user moves closer to the camera, the display can increase the magnification and conversely, as the user moves away the magnification is decreased.

This paper will describe the methodology and findings of an experiment that will be conducted to compare the user performance of conventional touch-based zoom implementations with the accelerometer and facial-tracking approaches as presented earlier.

Related Work
Joshi et al. [4] evaluated touch-free position and zooming control for viewing large imagery on mobile devices. Their experiment compared touch-based position control to various combinations of touch-free position, zoom and rate controls. They used face-tracking to compute the face's position and estimate the distance from the camera. Using these values they determined the viewing angle of the user. The z-axis was linearly mapped to zoom level. They found that zooming interfaces required approximately twice the amount of time compared to non-zooming interfaces. The experimenters attribute the increased time to the additional degree-of-freedom resulting in an increased precision required to acquire the target. Additionally, latency of the zoom and oscillation resulted in user frustration at times.

Harrison and Dey [3] similarly studied the lean and posture of users in correlation with the magnification of the displays. They implemented a system that calculated the user's amount of lean and increased the magnification proportionally. The calculation for the lean was computed based on the distance between the eyes when the user is in a "nominal posture". When the user leans towards the screen, the measured distance between the eyes increases, and is used to determine the amount of magnification to apply. The participants of the study had generally positive impressions of the system. However, the study also found a degradation in performance when seeking items in the display.

Francone and Nigay [2] studied the camera-based head tracking to calculate the position of the device relative to the user's head. This allowed them to display a 3D scene according to the vantage point of the user, enhancing the depth perception and creating a virtual 3D space within (or beyond) the screen. Participants in the study found the interaction with the system "natural and immersive". However, the authors noted that there were limitations in the camera's field of view, where at certain angles the face would be out of the range of the camera.

In a study comparing facial-tracking and accelerometer-based input, Cuaresma and MacKenzie [1] used a game to evaluate user performance when controlling a cursor-like object on the screen. The evaluation found that facial-tracking was inferior to tilt-based input. The authors proposed that facial-tracking was inferior due to its heavy-resource requirements that result in a noticeable system degradation and the lack of user experience with facial-tracking compared with the conventional tilt-input.

METHOD

A user study will be performed to compare three input methods for zooming: accelerometer input, touch input, and facial tracking. Each method will be evaluated for speed and accuracy to determine its effectiveness and efficiency. Touch input is the more common input method and will be used as a reference point for the comparison.

Participants

This experiment will involve recruiting twelve participants from the local university campus.

Apparatus

This experiment will use a *Samsung Galaxy Tab 4* tablet running *Google Android 4.4.2 Kit-Kat* operating system. See Figure 2. The device has an 8″ (203.1 mm) display with a resolution of 1280 × 800 pixels and a pixel density of 188.68 pixels/inch.

Figure 2 *Samsung Galaxy Tab 4*

The software will be developed in Java using the Google Android SDK and *Qualcomm Snapdragon Facial Recognition API* [5], an API only available on certain devices running on a *Qualcomm Snapdragon* processor.

The experiment application will be a map-like application developed specifically for this research. It will implement three modes of operation to accommodate for the three input methods being investigated.

Each participant will be given five trials for each of the three input modes, fifteen trials in total. For each trial, a marker will be placed randomly on the map with a number indicating a zoom level difference. The difference is the amount of zoom required to reach the desired zoom. A positive difference indicates that the user must zoom in, and a negative difference indicates that the user must zoom out. A crosshair will be provided as a visual indicator for where to center the marker. A marker will always be placed in the field of view at the start of each trial.

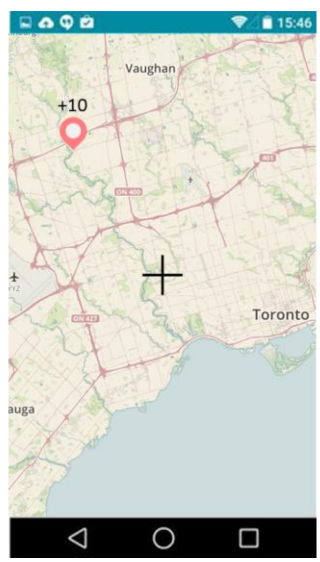

Figure 3 Experiment app mock-up user interface

The goal is for the user to center the marker and zoom until the level difference is zero. See Figure 3. A measure of the time duration required to complete each trial and the difference in zoom level and coordinate distance from the marker to the focal point for every 100 milliseconds will be recorded.

In touch mode, the user will use swiping to pan and two-fingered pinching to zoom in and out. In accelerometer mode, the user will move the device in the Z-axis, perpendicular to the display, to zoom in and out while tilting the device to center the focal point. In facial-tracking mode, the user will hold the device in a fixed position and move their head from side to side to center the focal point and lean their head towards and away from the screen to zoom in and out.

Each trial will be initiated with the press of a start button at the bottom of the screen. Additionally in accelerometer and facial tracking modes, each trial will start after the device has been held still for ten seconds at which time an audio tone will indicate the start of the trial.

Procedure

Participants will be briefed on the purpose and procedures of the experiment, and the experimenter will briefly demonstrate each of the three modes of the application. The test will occur in a quiet room with favourable lighting conditions where the participants will be seated at a table. For the facial tracking mode, participants will be instructed to keep their hands and forearms on the table to ensure that the device does not move. The device will be tilted back slightly such that the participant’s face is centered in the camera’s field of view. For the accelerometer and touch modes, participants will free to pick up the device. Participants will not be given practice trials. Following the completion of all fifteen trials, participants will complete a questionnaire soliciting demographic information and impressions about each of the input methods.

Design

A 3 × 5 within-subject design will be used. There will be two independent variables: input methods (touch input, accelerometer input, and facial tracking) and the trials (one through five). The dependent variables are the time of completion for each trial, and the difference of the zoom level and coordinate distance between the marker and focal point recorded over time. Participants are grouped into three groups, where each group is assigned a different ordering of the input methods in the trials. In total, there will be 180 trials for twelve participants each with five trials for each of the three input methods.

RESULTS AND DISCUSSION

To be added later.

CONCLUSION

To be added later.

ACKNOWLEDGMENT

To be added later.

REFERENCES

1. Cuaresma, J., & MacKenzie, I. S. (2014). A comparison between tilt-input and facial tracking as input methods for mobile games. Proceedings of the 6th IEEE Consumer Electronics Society Games, Entertainment, Media Conference - IEEE-GEM 2014, pp. 70-76. New York: IEEE. doi: 10.1109/GEM.2014.7048080
2. Francone, J., & Nigay, L. (2011, October). Using the user's point of view for interaction on mobile devices. In 23rd *French Speaking Conference on Human-Computer Interaction* (p. 4). ACM.
3. Harrison, C., & Dey, A. K. (2008, April). Lean and zoom: proximity-aware user interface and content magnification. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 507-510). ACM.
4. Joshi, N., Kar, A., & Cohen, M. (2012, May). Looking at you: fused gyro and face tracking for viewing large imagery on mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2211-2220). ACM.
5. Qualcomm. Snapdragon SDK for Android, https://developer.qualcomm.com/mobile-development/add-advancedfeatures/snapdragon-sdk-android, 2014.