

3D Perspective Effects on a Smart Phone

Jai Prakash

Carnegie Mellon University
Master of Science in Computer Vision

jprakash@andrew.cmu.edu

Jennifer Lake

Carnegie Mellon University
Master of Science in Computer Vision

jelake@andrew.cmu.edu

Abstract

In this project, we aim to create a 3D perspective transformation on a smart phone screen that will give the illusion of screen having depth. The user should feel as if they are looking into a long hallway, where objects can appear deeper and deeper into the hall. This is achieved by finding the three-dimensional vector between the users face and the center of the phone. This was achieved by tracking the users face using the front-facing camera. We also explored integrating this method with the smart phone's Internal Measurement Unit (IMU) in order to estimate the smart phone orientation. In both methods, Kalman filtering was used in order to create a more reliable vector prediction.

1. Introduction

1.1. Motivation

As mobile phones have evolved over the last twenty years, many of the major milestones have been in display improvements. Mobile phones have gone from very small, simple displays to larger and more colorful displays. At the forefront of this evolution, there is a growing trend of three-dimensional displays.

There are several benefits to having more 3D-like displays. 3D-like displays may help users

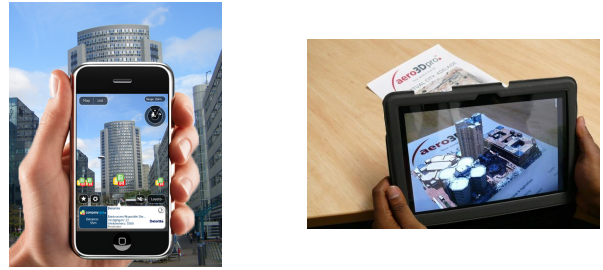


Figure 1: Augmented reality in present smart-phones

interact with User Interfaces in a more intuitive way, such as using the phone movement itself as a command. This can help when using the phones when wearing gloves, which do not allow the use of the touch screen. In addition, 3D-like displays may help users navigate using 3D maps, which are more representative of the surroundings in urban areas. Finally, 3D-like displays would be a major boost to the gaming industry, which constantly strives to make more and more realistic games. With perspective effects that mimic real-world perspective effects, the user would experience a more immersive gaming experience.

The holy grail of 3D perspective transforms is to be able to provide augmented reality to the user. This would allow the user to use their smart phone to interact with the world around them in an intuitive manner, such as in figure 1. We hope that some of the work in this project, may help in future projects striving towards this goal.

1.2. Background

The change in appearance of an object when viewed from multiple viewpoints is known as parallax [1]. Parallax occurs due to the shift in perspective. An example of the parallax effect is shown below [2].

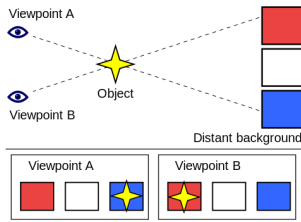


Figure 2: An example of the parallax effect. The top of the figure shows the scene from above. The bottom of the figure shows the views from each viewpoint.

There are three types of parallax effects: negative, zero, and positive parallax (figure 2) [3].

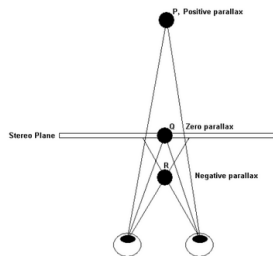


Figure 3: Example of positive, zero, and negative parallax

Positive parallax occurs when the user gets the impression of looking through the screen plane and this effect will be the focus of this project. Negative parallax occurs when the user get the impression that objects are floating above the user screen and this effect will not be addressed in this project. Zero parallax is the impression that the object lies on the screen and this conventional way of viewing objects on a screen.

1.3. Related Work

The trend of introducing of positive parallax effects into smart phones has been growing in the last few years. In 2013, Apple introduced the positive parallax effect on their line of iPhones, which allows for a 3D-like feeling, by changing the position of the background wallpaper according to the orientation of the smart phone [4]. A year later, the Amazon Fire Phone introduced a full positive parallax effect, which was branded as "Dynamic Perspective" [5]. It is this type of effect that we aim to create in this project however, unlike the Amazon Fire Phone, we will be implementing this effect on simpler, standard smart phone hardware.

Prior to the growing trend in parallax effects in smart phones, the Microsoft Kinect was to estimate a person's pose and using that information, render a parallax effect on a television screen. One interesting example of this is the Virtual Window, which has a television behind a false window and the landscape on the television changes based on the perspective of the user. One major constraint of this type of system is that the perspective change only works for one user at a time.

In the future, we expect parallax effects to lead to Augmented Reality for smart phones. This could allow users to seamlessly interact with the surrounding environment and could lead to new uses for smart phones.

1.3.1. Virtual Window

The Virtual Window works by using a Microsoft Kinect sensor, which projects a light pattern onto the user and uses an RGB-D camera to find the pose and 3-dimensional pose of the user [6]. Using this information, the position of the user's head and eyes can be located using built-in tools in the Kinect API. A high-resolution, high-quality image of an outdoor scene is then moved in order to simulate that the screen is in fact a window. The image only needs to be shifted to the appropriate location because the images are taken

from far away (see figure 4).



Figure 4: Virtual Window from two different perspectives

1.3.2. Amazon Fire Phone

The Amazon Fire Phone uses four, front-facing camera's with wide 120 °viewing angles [7]. This allows for at least two cameras to be able to locate the user's face, regardless of how the phone is held. The face tracking works in low light as well, due to infrared light being used to illuminate the user's face. By tracking the user's face, the vector between the phone and the user can be approximated and thereby the correct perspective transform can be computed and displayed.

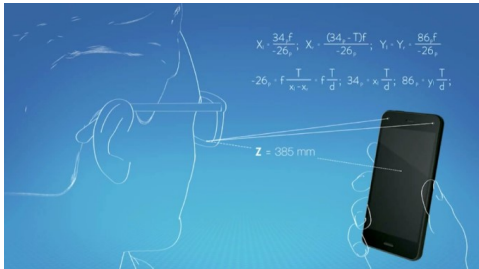


Figure 5: Example of Amazon Fire Phone's Dynamic Perspective

1.3.3. Augmented Reality

The ultimate goal of estimating a phone's orientation with respect to the user is to be able to implement Augmented Reality (AR) into smart phones. The smart phone Augmented Reality can broadly be classified into two types:

- **Indirect AR** Figure 1 shows augmented reality applications in current smart phones.

This gives the user a video see-through experience, as the view is from the camera's perspective.

- **Direct AR** These are the systems that give a perspective from user's point of view and hence give more immersive experience than indirect AR system. The Microsoft HoloLens and meta-glasses (Figure 6) are examples of direct AR systems.



Figure 6: Direct AR from Meta-glasses

The current direct AR systems are generally bulky and requires user to wear special glasses to use them. However, a direct AR system can be achieved in smart phones and tablets by using digital transparency [8]. Digital transparency system is shown in Figure 7. This system tracks the eyes of the user and finds the angle which the tablet subtends at user's eye and crops the rear camera preview to the same angle to create a virtually transparent interface. This system gives a method to achieve direct AR on smart phones and tablets, thereby giving a more immersive **glass-see-through** experience than the conventional **video-see-through** experience.

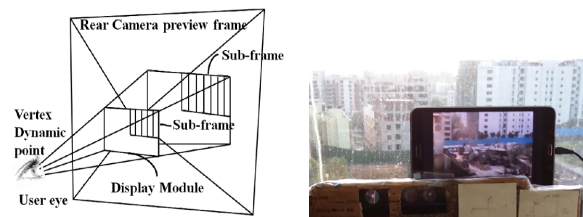


Figure 7: Digital transparency using Kinect on Android tablet

This method uses a Kinect sensor to track the accurate 3D position of the user with respect to tablet. The Kinect sensor makes the system bulkier and difficult to use in smart phones. So, in this project we are addressing another method to achieve digital transparency using just the front facing camera and attempting to integrate the inertial sensors. A complete hands-free digital transparency system is out of scope of this project. The main concentration is on creating a positive parallax effect which responds to the user's position with respect to the tablet.

2. Method

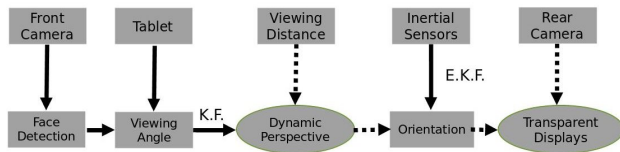


Figure 8: Block diagram of system

Creating a positive parallax effect involves finding the viewing angle of the user and rendering the graphical object depending on position of the user. This might involve one of three scenarios:

- Tablet is static and user moves
- Tablet moves and user is static
- Both tablet and the user can move

In order to limit the scope of the project, we have also placed a number of restrictions on how this systems is to be used. We also assume the following restrictions:

- There are no drastic lighting changes
- The user is not walking or running
- The user is standing or sitting on a stationary surface.

- The user and the tablet are moving smoothly and slowly
- The tablet is held at a constant distance from the user's face.
- There is only one user and the user is always in view of the front camera.
- There are no other faces in the view of the front camera
- The user is looking at the camera or very close to it
- The user is an arm's length or less from the camera
- The user's face is the primary object in the view of the camera

To find the motion and orientation of the phone, on-board inertial sensors can be deployed to find the orientation of the phone. A study on sensor fusion has been made to find the orientation of the phone.

The block diagram of the system is shown in the figure 8. Solid arrow shows the algorithms implemented for this project. The dotted arrow parts are yet to be figured out to build a full-fledged digital transparency system. However, the main focus is only on creating the positive parallax effect.

To create this effect, we need to first track the position of user with respect to the tablet. In this project we assume that the distance of the eye from tablet is fixed (approximately 45-50 cm). This assumption fails to address the scale factor in dynamic perspective. Therefore we don't get zoom-in and zoom-out effect as we move forward and backward. Since the distance from the tablet is assumed to be constant, we use the viewing angle to render the user interface (UI) in 3D. The following sections describe the exact methods used in each step.

2.1. 3D Geometry of Problem

In order to render a positive parallax effect, we must estimate the vector between the user and the phone, as shown in figure 9.

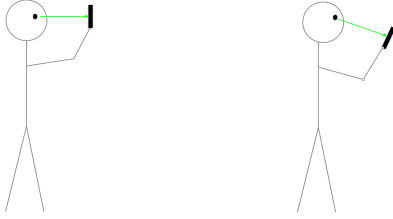


Figure 9: Example of vectors to be found

Normally, a user will hold a smart phone orthogonally to the vector between the user's eye and the principle point of the camera (figure 10). This is a natural position of the screen, since it is easiest to view from this angle and does not strain the eyes.

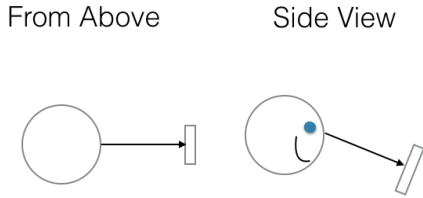


Figure 10: Conventional way of holding a smart phone

However, it is possible for the user to hold the smart phone in an unconventional matter, such as in figure 11.

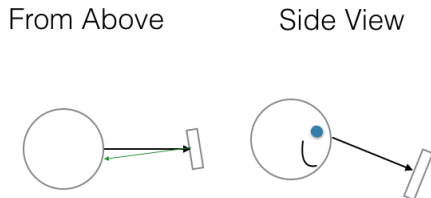


Figure 11: Unconventional way of holding a smart phone.

-Jenna: finish this

2.2. Finding viewing angle of user

Since all the coordinates are in image frame of reference, the angle can be found by using the following equations.

$$\theta_x = \tan^{-1} \left(\frac{x_f - \frac{W}{2}}{f_x} \right) \quad (1)$$

$$\theta_y = \tan^{-1} \left(\frac{y_f - \frac{H}{2}}{f_y} \right) \quad (2)$$

In equation 1 and 2, (x_f, y_f) is the location of the centroid of the user's face. Image size is $W \times H$. f_x and f_y are the focal length of the camera in pixels.

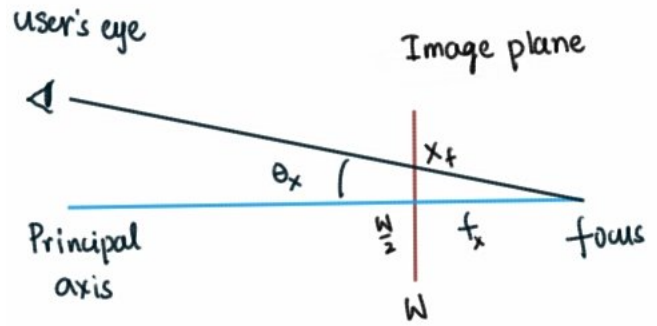


Figure 12: The viewing angle of the user with respect to front camera on smart phone

Figure 12 shows the viewing angle of the user. The view angle is the angle subtended by the user's position with the principal axis of the camera. The focal length of the camera is found by using camera calibration in OpenCV [12].

-Jai: Include equations from reference

2.3. Face Detection

The Viola-Jones algorithm was used for face detection because it is robust and fast [11]. The Viola-Jones algorithm has four main steps:

1. Compute Haar Features
2. Compute Integral Image

3. Use AdaBoost to Train Classifier

4. Cascade Classifiers

In the first step, Haar features are computed. Haar features represent simple structure present in human faces, such as darker eyes and lighter cheeks or a lighter nose and darker eyes. These features are calculated at different orientations. An example of Haar features can be found in figure 13.

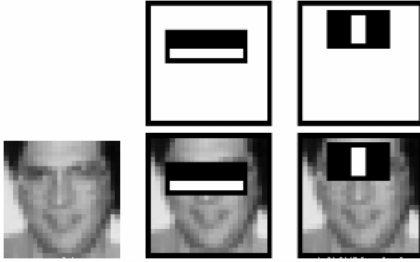


Figure 13: An example of Haar features

The integral image is then computed by summing the value of a pixel and all pixels to the left and above it. This can be done in one pass if the integral image is calculated from the origin outward using equation 3. This is done so that the Haar calculations, which are defined in equation 4 can be completed quickly.

$$I(x, y) = i(x, y) + I(x - 1, y) + I(x, y - 1) - I(x - 1, y - 1) \quad (3)$$

$$score = \sum blackpixels - \sum whitepixels \quad (4)$$

-Jenna : finish this

2.4. Inertial Measurement Unit

An Inertial Measurement Unit (IMU), is a sensor that features a triad of accelerometers and a triad of gyroscopes [13]. In cell phones, the IMU also typically features a magnetometer.

-Jenna : finish this

2.5. Kalman Filter

-Jai : Include equations

2.6. Extended Kalman Filter

-Jenna : Include Equations

3. Experiments

3.1. Image Geometry Experiments

- Jai: put results here / how it was done

3.2. Camera Calibration

The camera calibration matrix is returned as

$$K = \begin{bmatrix} f_x & \alpha_x & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 145 & 0 & 930 \\ 0 & 145 & 601 \\ 0 & 0 & 1 \end{bmatrix} \quad (5)$$

The value of the focal length is in terms of the pixels. The resolution of the camera is 1920x1280, and the principal axis seems to be almost at the center of the image.

3.3. IMU Experiments

-Jenna: put results here and how much it sucked

4. Conclusions

-Both: wrap it up, future work if room is needed

References

- [1] Szeliski, Richard. *Computer Vision: Algorithms and Applications*. 1st ed. London: Springer-Verlag, 2010. Print.
- [2] *Parallax Example*. Digital image. *Wikipedia*. Wikimedia, 31 May 2006. Web. 15 Dec. 2015.
- [3] Positive vs. Negative Parallax. Digital image. *CSE 621 Contemporary Computer Graphics*.

- California State University, San Bernadina, n.d. Web. 15 Dec. 2015.
- [4] Yarow, Jay. "Here's Why Apple Made That Motion-Effect For The Background Of The New iPhone Software." *Business Insider*. Business Insider, Inc, 25 Sept. 2013. Web. 11 Dec. 2015.
- [5] Pelegrin, Williams. "Why Isn't the Fire Phone Truly 3D? Amazons Dynamic Perspective Tech Explained." *Digital Trends*. Digital Trends, 18 June 2014. Web. 11 Dec. 2015.
- [6] Sevilla, Beth. "Winscape." *Winscape*. Rational Craft, n.d. Web. 14 Dec. 2015.
- [7] Crook, Jordan. "Amazons Fire Phone Uses Depth And 3D Effects To Stand Out." *TechCrunch*. TechCrunch, 18 June 2014. Web. 14 Dec. 2015.
- [8] Prakash, Jai, and Ajay Vijayvargiya. A Method for Obtaining Digital Transparency in Electronic Devices. Samsung, assignee. Patent 4506/CHE/2014. 16 Sept. 2012. Print.
- [9] Alvarez, Edgar. Dynamic Perspective. Digital image. *Engaget*. AOL, 8 June 2014. Web. 15 Dec. 2015.
- [10] Kalman, R. E. "A New Approach to Linear Filtering and Prediction Problems". *Journal of Basic Engineering*, 1960
- [11] Viola, Paul, and Michael J. Jones. "Robust Real-Time Face Detection." *International Journal of Computer Vision* 57.2 (2004): 137-54. *ProQuest*. Web. 15 Dec. 2015.
- [12] "Camera Calibration With OpenCV." *OpenCV 3.0.0 Documentation*. Itseez, 10 Nov. 2014. Web. 15 Dec. 2015. http://docs.opencv.org/3.0-beta/doc/tutorials/calib3d/camera_calibration/camera_calibration.html.
- [13] Graves, Joe. "Design And Control Of A Vehicle For Neutral Buoyancy Simulation Of Space Operations." Thesis. University of Maryland, 1997. Print.
- [14] Cunningham, Andrew. "iPhone 6 and 6 Plus: In Deep with Apples Thinnest Phones." *Ars Technica*. Conde Nast, 22 Sept. 2014. Web. 14 Dec. 2015.
- [15] Aguilar, Mario. "Here's Why the iPhone 5S Accelerometer Is So Screwed Up." *Gizmodo*. Gawker Media, 16 Oct. 2013. Web. 14 Dec. 2015.
- [16] Quan, Wei. *INS/CNS/GNSS Integrated Navigation Technology*. Heidelberg: Springer, 2015. *Springer Link*. Springer. Web. 14 Dec. 2015.