Real-time Robust Person Detection and Tracking

Junle Lu, Mentor: Dr. Scott Craver
State University of New York at Binghamton, NY

Abstract

The goal of this project is to fuse camera and depth sensor data into a robust interface for a projector-based display table. Our primary goal is identifying the focus of a user's gesture amid a group of people, by triangulating face and gesture data. We employ Haar Cascades to detect faces and other objects, but this method alone is unreliable and computationally intensive. The depth sensor from an Xbox 360 Kinect is used to filter and reduce possible facial feature locations. By reducing candidate features with depth information and fusing that information in detection, we improve detection and speed.

Objectives

- *Integrate Haar Cascades method with libfreenect for real-time person detection on Xbox 360 Kinect.*
- *Reduce Haar Cascade computational burden while maintaining system speed and accuracy.*
- *Utilize depth sensor information to identify possible facial feature locations and discard regions unlikely to contain a face.*
- *Reliably locate users as part of a larger gesture recognition user interface for an interactive display.*
- *Optimize the system for speed.*

Methods

The Haar Cascade algorithm can efficiently and accurately detect the faces on a single image by computing a single integral image per frame, and sliding a window over the image at multiple scales[1]. The integral image is used to quickly compute Haar features (rectangles) that are matched to patterns in a decision tree.

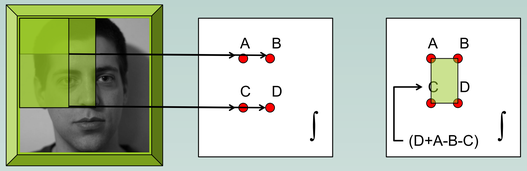

Figure 1. Each point in an integral image is the luminance sum of all image pixels from the upper left corner to that location. This allows rapid computation of the pixel sum in any window.

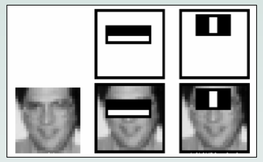

Figure 2. Haar Cascade example: a window is correlated against patterns of light and dark rectangles, by computing the pixel sum in these regions[1].

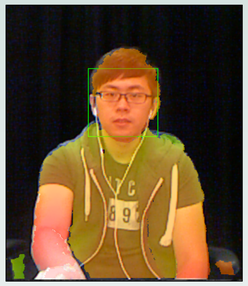

Figure 3. overlayed camera and depth data from Kinect

The algorithm was modified and adapted for real-time facial detection with an Xbox 360 Kinect driven by Libfreenect software. We compute an integral image of both image and depth data, and only subject a window to a Haar cascade face test if the average depth in the window is close to the value expected for a face of that size.

The Haar cascade algorithm was run at varying scales to measure average depth for found faces, shown in figure 5.

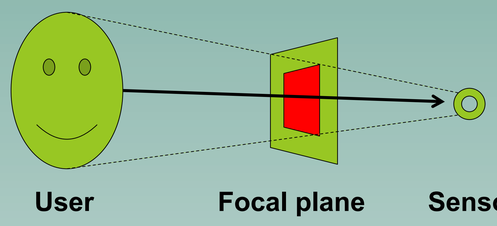

Figure 4. For a fixed head size, observed face depth (distance to sensor) is expected to be inversely proportional to window size in the camera frame.

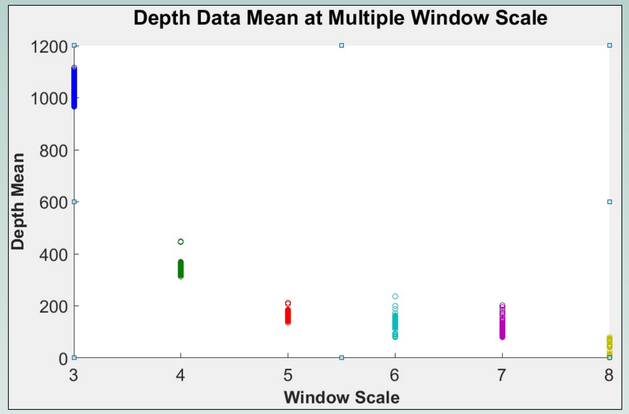

Figure 5. Observed face depth at varying window scales. For a scale of N, window size is 24N pixels in a 640x480 frame.

A linear relationship was observed between window scale and inverse of average depth as in figure 6. This is used to compute an upper and lower depth threshold for sliding windows.

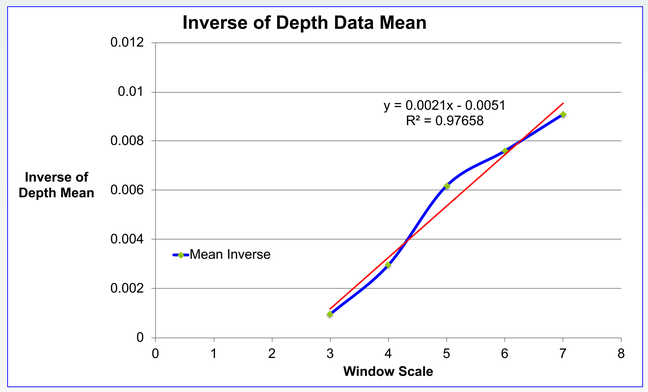

Figure 6. The Inverse of Average Value of Depth Sensor Data

Results

By incorporating depth information, most windows can be excluded from a Haar cascade search. We observe greater than 300-fold reduction in the number of Haar cascade tests, resulting in an approximate 40-fold reduction in computation time.

The disparity between test reduction and time reduction is possibly due to the Haar cascade spending more time on likely faces than on unlikely ones that are now excluded.

Scale 3 to 9, += 1.0	xy_incre += 1	Depth threshold xy_incre += 1	xy_incre += width/30	Depth threshold xy_incre += width/30
Funcation Call	1070784	3434	112933	305
Computation Time	475 ms	12.5 ms	54 ms	1.5 ms

Table 1. The Performance of Face Detection

Conclusion

The face detection performance was significantly improved with the fuse of depth sensor from Xbox Kinect 360. The program was able to run roughly 40 times faster without sacrificing detection accuracy.

High-speed face tracking is critical both for real-time face tracking, and its application as a user-interface in a machine with other computational obligations.

References

1. Paul Viola and Michael Jones. "Rapid Object Detection using a Boosted Cascade of Simple Features". 2001
2. Tripathy, Rajashree and Daschoudhury, R N. "Real-time Face Detection and Tracking Using Haar Classifier on SoC". *International Journal of Electronics and Computer Science Engineering* P175
3. Hossny, M and Nahavandi, S. "Low Cost Multimodal Facial Recognition via Kinect Sensors"
4. Yan, Chao and Zhang, Zhaoyang. "Robust real-time Multi-user Pupil Detection and Tracking Under Various Illumination and Large-scale Head Motion" *Computer Vison and Image Understanding* P1223-1238, 2011