Indoor 3D Computer Vision System Using Depth Cameras

Advisor: Professor Murali Subbarao

Aaron Knoll and Jing Huang

Stony Brook University
Electrical and Computer Engineering

Visionary Systems

Overview

The goal of our project was to utilize the cutting edge in affordable multi-sensor RGB-D technology to identify objects and faces using 2D techniques, and verify results using depth data. This technology can be adapted for use in a variety of applications, including automated quality control for factory assembly lines, autonomous robot navigation, and authentication applications, such as unlocking a safe, home, or electronic device.

Methods

Object Detection and Recognition

- Employed deep-learning based object detection and recognition using the CAFFE model.

Distance to Object

- Computed mean of scaled-depth data within bounded region found when performing object detection.

Face Detection and Midface Isolation

- Employed Viola-Jones frontal face detection using Haar-like features for initial face region bounding and cropping.
- Further isolated midface using depth data and perspective projection computations (to determine bounding region in terms of real-world measurements).

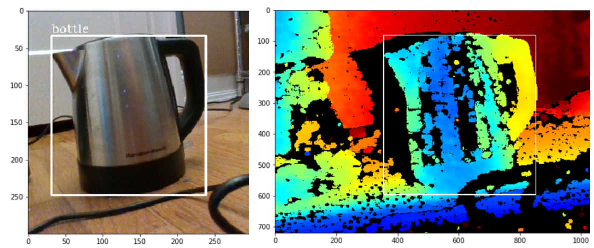
Face Recognition

- Employed a highly accurate, open-source deep learning-based 2D face recognition system.

Face recognition Results Verification

- Compared computed coefficients of variation for face depth data.
- Computed the closeness of face depth data keypoint matches (using ORB).
- Computed the minimum Euclidean distance between face depth datasets.

Results

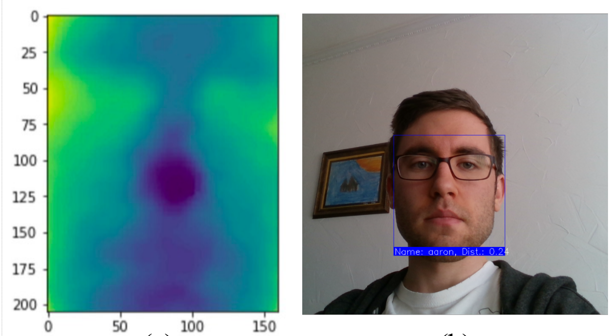


(a) (b)

```
print("Detected a {0} {1:.3} meters away.".format(className, dist))
Detected a bottle 0.233 meters away.
```

(c)

Figure 1. Sample object detection and recognition results: (a) object detected, bounded, recognized, and labeled, (b) Object bounded (in colorized depth image), (c) computed distance to object.



(a) (b)

```
print(minEuclideanDist(faceRegionDepth, faceRegionDepth2))
0.621751556172943
```

(c)

Figure 2. Sample face detection and recognition results: (a) normalized and interpolated midface depth image (colorized), (b) face detected, bounded, recognized, and labeled, (c) computed minimum Euclidean distance of depths.

Conclusion

Despite various challenges, we managed not only to build an indoor computer vision system capable of both object and face recognition with verification through depth data analysis, but also to come up with a custom interpolation method for depth data with clusters of missing data and a custom depth data comparator that considers various possible translations using a common region of closest depth to reduce computational complexity, among other things.

What is most important, however, is that we expanded our horizons and were empowered to do more and think bigger with an exposure to the technologies and techniques that form the foundation of so many of the modern vision-based wonders we take for granted.

Glossary

CAFFE: A deep learning framework developed by Berkeley AI Research; an acronym for Convolutional Architecture for Fast Feature Embedding.

Computer vision: An interdisciplinary scientific field concerned with how computers can gain a high-level understanding from digital images or videos.

Interpolation: A method for estimating values within the range of a discrete set of known data points.

ORB: An open-source algorithm used for keypoint feature detection and description; an acronym for Oriented FAST and Rotated BRIEF.

RGB-D: Four-channel data; red, green, blue, and depth.

Acknowledgements

We would like to express our deepest gratitude to Professor Subbarao for his wisdom, guidance, and unyielding support. We would also like to thank Professor Dhadwal for the high standards he set for us and the confidence he instilled to meet or exceed them.

Visionary Systems