REU Project: Real-Time Multi-Camera Object Detection, Matching, and Depth Estimation for 3D Mapping

SenSIP Algorithms and Devices REU

Emilio Montoya, REU Student, Cochise College, Arizona State University

Graduate Mentor: David Ramirez, Faculty Advisor: Andreas Spanias

SenSIP Center, School of ECEE, Arizona State University

ABSTRACT

- Simultaneous live video feed from 2 cameras will run through an object detection CNN.
- The use of ORB points for feature matching across simultaneous live video feed from two cameras
- The use of a 360° camera for a broad POV for one camera
- Develop a room mapping algorithm which incorporates object detection, depth estimation, and feature point mapping.

EXPERIMENTAL METHODS: SENSORS

- **Camera**
  - 360 Fisheye Camera
  - Additional Simulated Robotic Agent
- **Detect**
  - Convolutional Neural Network
  - Object Detection
- **Feature Point Similarity**
  - Creation of ORB points
  - Matching ORB points between images
- **Depth Estimation**
  - Monocular Depth Estimation
  - 3D image generated using depth estimation
- **3D**
  - Structure from Motion
  - Real Time Position

PROBLEM STATEMENT

- We are trying to localize and detect alterations on the environment, caused by animal interaction and human interferences.

GOAL

- The goal is to develop a 3D digital model in which the environmental alterations can be located and be analyzed for possible causes.

PRELIMINARY RESULTS

- We found that interactions between neural networks vary from the CNN used
- SSD MobileNet v2 has a hard time detecting black cats with a black background
- Feature Point Matching performs efficiently on images and on video
- Depth estimation can be used to recreate the structure of an environment from an image

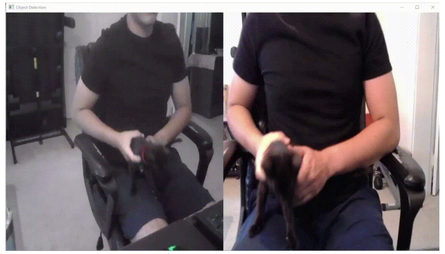

Figure 1. CNN module detecting person and cat with multiple cameras

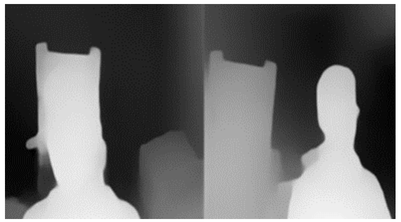

Figure 3. depth output image generated by a Depth Generator Network

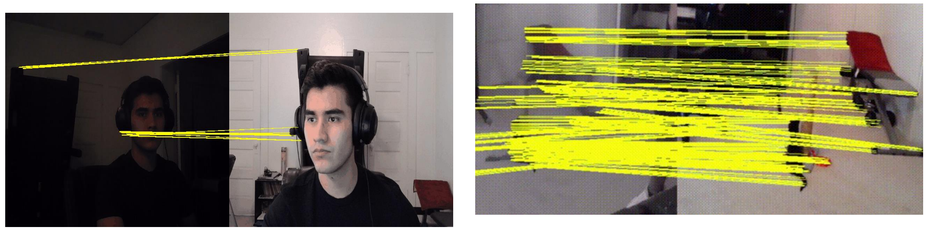

Figure 2. Feature Point Similarity Matching (yellow lines)

Figure 4. Combination of input RGB image and depth output image to develop a 3D model of the ground truth environment

REFERENCES

[1]Liu, Wei, et al. "SSD: Single Shot Multibox Detector." European Conference on Computer Vision. 2016.

[2] Sandler, Mark, et al. "Mobilenetv2: Inverted Residuals and Linear Bottlenecks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.

[3] Bian, JiaWang, et al. "GMS: Grid-based Motion Statistics for Fast, Ultra-robust Feature Correspondence." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.

[4] Alhashim, Ibraheem, and Peter Wonka. "High Quality Monocular Depth Estimation via Transfer Learning." arXiv preprint arXiv:1812.11941. 2018.

[5] Ranftl, René, et al. "Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-shot Cross-dataset Transfer." arXiv preprint arXiv:1907.01341. 2019.

ACKNOWLEDGEMENT

This project was funded in part by the National Science Foundation, award number 1659871.