



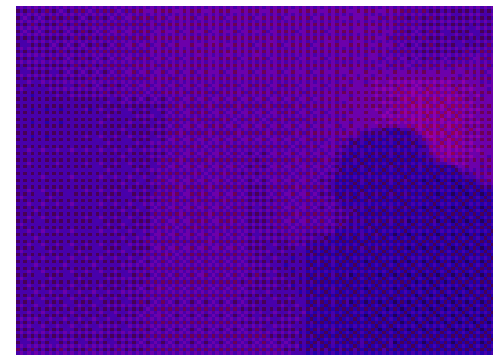
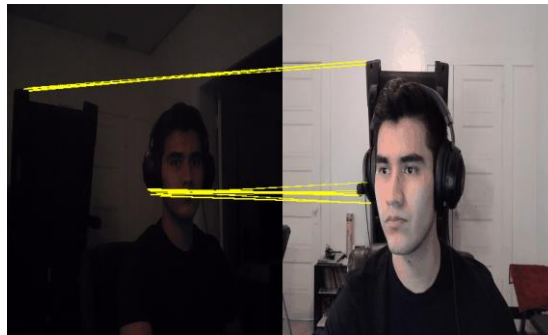
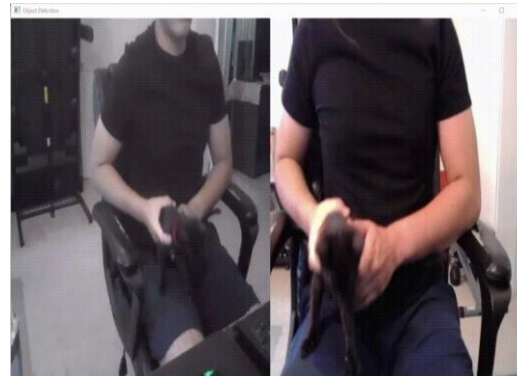
REU Project: Real-Time Multi-Camera Object Detection, Matching, and Depth Estimation for 3D Mapping



Emilio Montoya, REU Student, Arizona State University

Graduate Mentor: David Ramirez Faculty Advisor: Andreas Spanias

SenSIP Center, School of ECEE, Arizona State University



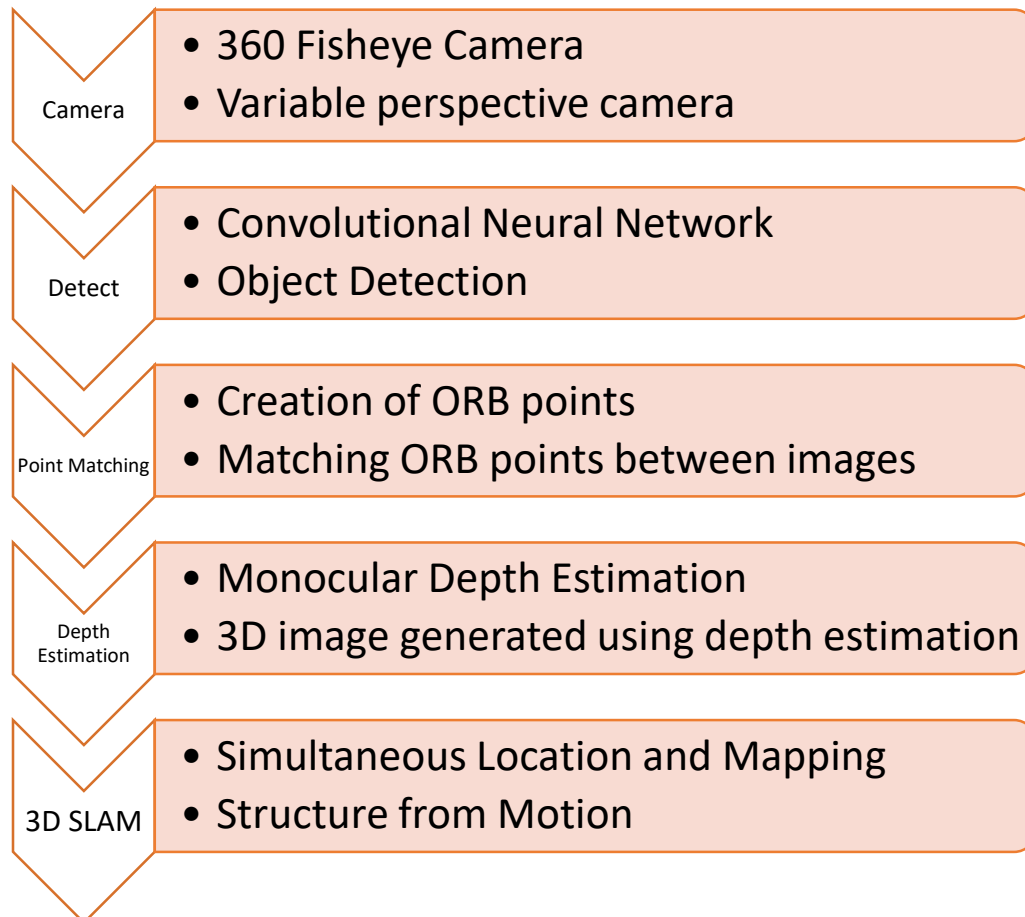


PROBLEM STATEMENT AND GOAL



- Problem
 - When away from the home pets can be destructive
 - Need to monitor pet behavior and check for significant changes
- Goal
 - Simultaneous live video from two cameras
 - Cat tracking object detection
 - Match overlap and objects and objects across live videos
 - Calculate distance of objects and room from cameras
 - Develop a room mapping algorithm





Garmin VIRB 360

- Omnidirectional Camera
- Video Resolution
 - 5.7K/30FPS



Razer Kiyoh

- Video Resolution
 - 720 at 60 FPS

OBJECT DETECTION

- Modules Tested

- CenterNet HourGlass
- SSD MobileNet
- Faster R-CNN
- SSD Resnet
- EfficientDet D0

- Performance Variation

- Trade off between speed and precision
- Results at the end



FEATURE POINT MATCHING

- Computer vision algorithm
 - Similarity matching across images
- Multi-Camera Point of Views
 - Main Camera
 - 360 omnidirectional camera
 - Variable perspective camera

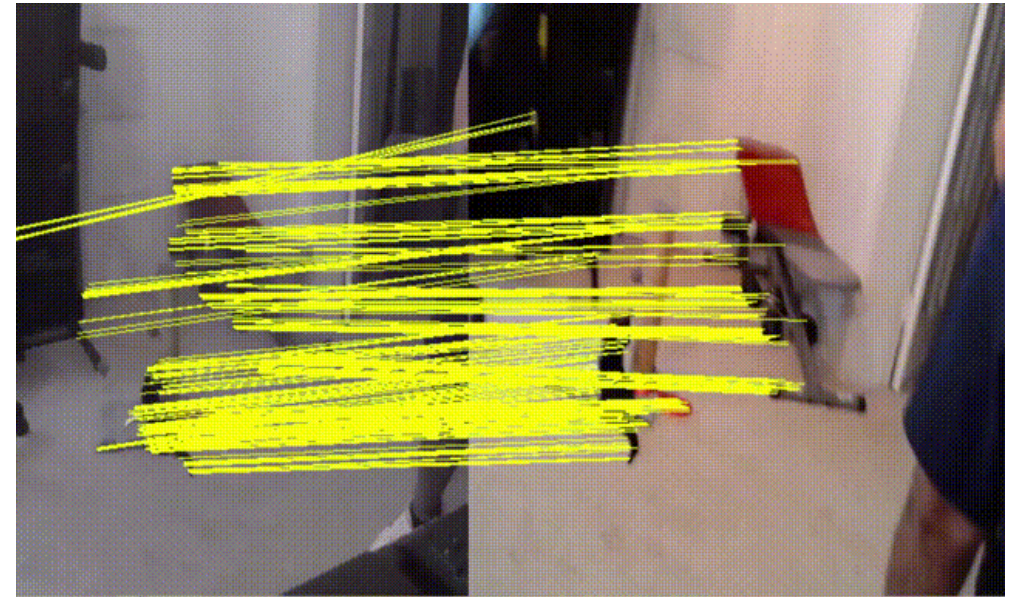


Figure 1. Feature Point Similarity Matching (yellow lines)

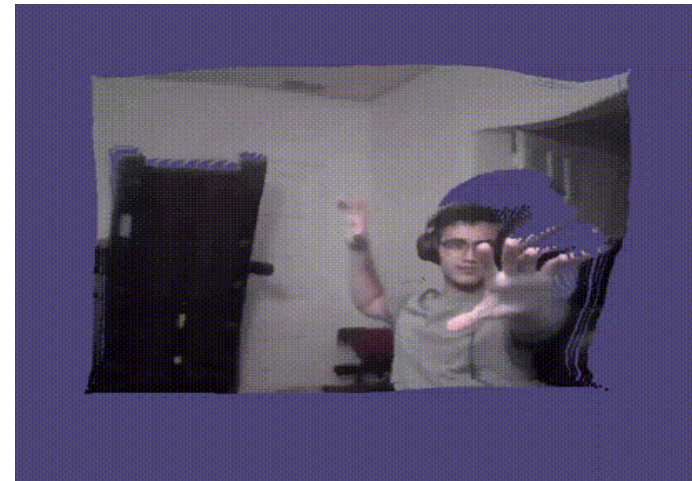
MiDas v2

- Depth Generator Network
 - RGB input image
 - Depth map output



DenseDepth

- Point Cloud Reconstruction
 - Combination of input output data





RESULTS



Object Detection CNN Module Type	Average Inference Time from Video
SSD MobileNet v2	0.1129 seconds
EfficientDet D0 512x512	0.8080 seconds
CenterNet HourGlass 104 KeyPoints 512x512	5.6910 seconds
SSD ResNet152 FPN 640x640 (RetinaNet152)	14.8599 seconds
Faster R-CNN ResNet152 v1 640x640	22.9677 seconds

Depth Generator Network Module Type	Average Inference Time from Video
MiDas v2	0.8561 seconds
Feature Point Matching	
Grid-based Motion Statistics	0.3378 seconds
Object Detection, Depth Estimation, and Feature Point Matching	
SSD MobileNetv2 MiDas v2 Grid-based Motion Statistics	0.82364 seconds



FINDINGS



- Interactions between Neural Networks
 - We found inconsistencies
 - When SSD MobileNet v2, MiDas v2, and Grid-based Motion Statistics ran together
 - Time by frame: 0.82364 seconds
 - Running Faster than anticipated
 - When SSD MobileNet v1 FPN 640x640, MiDas v2, and Grid-based Motion Statistics ran together
 - Time by frame: 3.8675 seconds
 - Running slower than anticipated

Modules Used	Average Time from Video
SSD MobileNet v2	0.1129 seconds
MiDas v2	0.8561 seconds
Grid-based Motion Statistics	0.3378 seconds
Combined Time	1.3068 seconds

Modules Used	Average Time from Video
SSD MobileNet v1 FPN 640x640	1.6190 seconds
MiDas v2	0.8561 seconds
Grid-based Motion Statistics	0.3378 seconds
Combined Time	2.8129 seconds



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] Howard, Andrew G., et al. "Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications." arXiv preprint arXiv:1704.04861 (2017).
- [4] Lin, Tsung-Yi, et al. "Focal Loss for Dense Object Detection." Proceedings of the IEEE International Conference on Computer Vision. 2017.
- [5] Ren, Shaoqing, et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." Advances in Neural Information Processing Systems 28. 2015.
- [6] He, Kaiming, et al. "Deep Residual Learning for Image Recognition." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.
- [7] Lin, Tsung-Yi, et al. "Feature Pyramid Networks for Object Detection." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.
- [8] Duan, Kaiwen, et al. "Centernet: Keypoint Triplets for Object Detection." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.
- [9] Melekhov, Iaroslav, et al. "Image-based Localization Using Hourglass Networks." Proceedings of the IEEE International Conference on Computer Vision Workshops. 2017.
- [10] Tan, Mingxing, Ruoming Pang, and Quoc V. Le. "Efficientdet: Scalable and Efficient Object Detection." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.