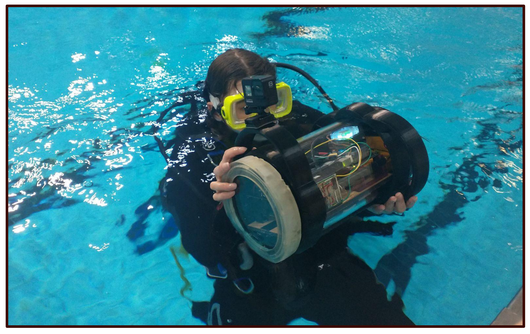
Pointing Underwater Isn't that Simple:

Improving Diver & Robot Interactions in a 3D Underwater Environment

Megdalia Bromhal[1], Junaed Sattar[2], Mentors Demetri Kutzke[2] and Chelsey Edge[2]

[1]University of North Carolina Wilmington, [2]University of Minnesota Interactive Robotics and Vision Lab



Hydro Eye Stereo Camera Used

Problem

- Underwater visibility is poor compared to terrain pointing
- Robot perceives world in 2D, while human perceives world in 3D

Robot doesn't understand what the diver points to

Objective

Enable robot to find item of interest from diver pointing

Applications

- Diver pointing to coral for robot to map or photograph
- Diver pointing to trash for robot to pick up or remove

Previous Work:

- Deep Lab Cut[5] pose estimator to find coordinates of diver
- Calculate disparity between stereo images to determine 3-dimensional coordinates for diver's elbow and wrist

Project Innovations:

- Use pose estimator (MediaPipe[4]) to find (x, y) coordinates for diver's wrist and elbow, since wrist and elbow primarily used in pointing
- Combine MediaPipe & mentor's work to calculate (x, y, z) for elbow and wrist on diver
- Compare MediaPipe with Deep Lab Cut localization accuracy to determine quality of pose estimation performance

Analysis

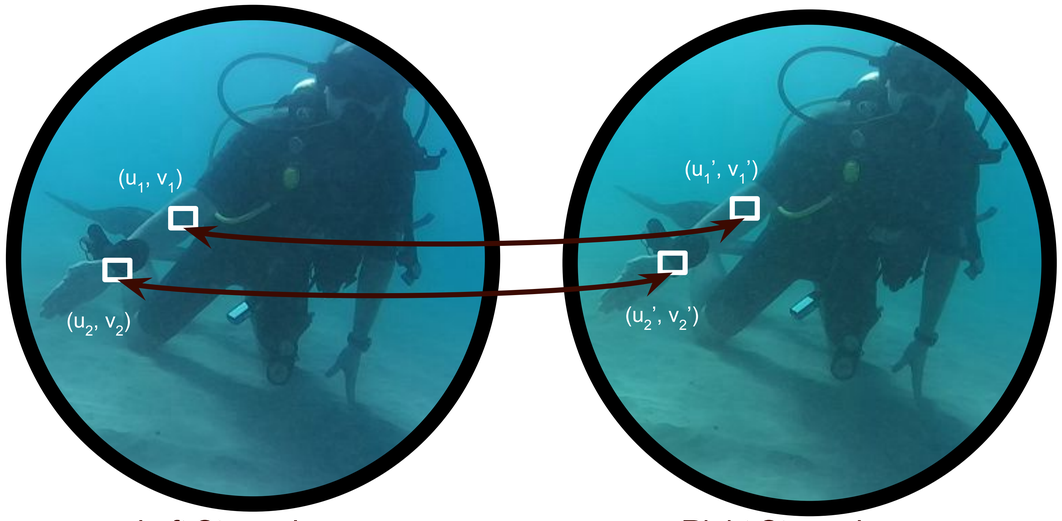
- Over 90 stereo images (left and right), MediaPipe performed on average 1.65 times more accurate at predicting a diver's pose than Deep Lab Cut
- However, MediaPipe detected a pose in only 8 out of those 90 image pairs (8.89%)

Next Steps

- Define 3-dimensional location of interest for robot to inspect from diver pointing

Finding Z

Using MediaPipe Pose Estimator to Calculate Z Coordinate



Left Stereo Image — Right Stereo Image

When rectified, $v=v'$ and disparity = $(u-u')$

Elbow = $(u_1-u_1', v_1-v_1') = (u_1-u_1', 0)$

Wrist = $(u_2-u_2', v_2-v_2') = (u_2-u_2', 0)$

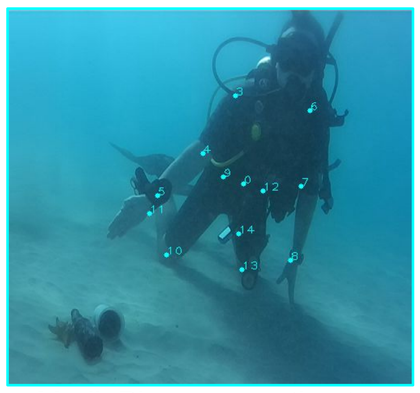
(x, y, z)

(x_1, y_1, z_1)

(x_2, y_2, z_2)

Checking the Accuracy of Z

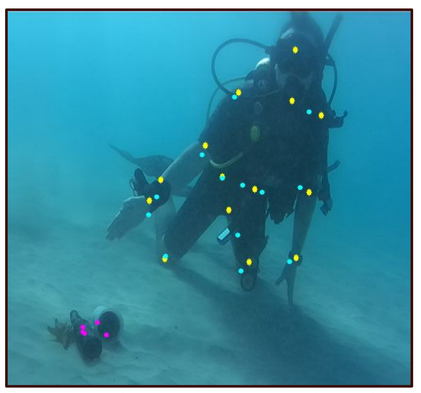
Comparing Location Accuracy of MediaPipe and Deep Lab Cut Pose Estimators



MediaPipe pose estimator

Deep Lab Cut pose estimator

Ground Truth (manually labeling)



Combination of pose estimators

Contact:

Megdalia Bromhal
University of North Carolina Wilmington
mb2710@uncw.edu

Interactive Robotics and Vision Lab
Department of Computer Science and Engineering
Lab Website: https://irvlab.cs.umn.edu/

References:

[3]Edge, Chelsey, and Junaed Sattar. "Diver Interest via Pointing: Human-Directed Object Inspection for AUVs." In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3146-3153. IEEE, 2023.

[4]Lugaresi, Camillo, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang et al. "Mediapipe: A framework for perceiving and processing reality." In *Third workshop on computer vision for AR/VR at IEEE computer vision and pattern recognition (CVPR)*, vol. 2019. 2019.

[5]Mathis, Alexander, Pranav Mamidanna, Kevin M. Cury, Taiga Abe, Venkatesh N. Murthy, Mackenzie Weygandt Mathis, and Matthias Bethge. "DeepLabCut: markerless pose estimation of user-defined body parts with deep learning." *Nature neuroscience* 21, no. 9 (2018): 1281-1289.