Robust Scuba Diver Tracking and Recovery in Open Water Using YOLOv7, SORT, and Spiral Search

Faraz Lotfi, Khalil Virji, and Gregory Dudek

Mobile Robotics Lab at McGill University {f.lotfi,kvirji,dudek}@cim.mcgill.ca

Introduction

Robust scuba diver tracking is essential for many underwater tasks.

- Navigating to workspaces.
- Staying close by for safety and communication.

We present a framework for **underwater scuba diver tracking and recovery** using AQUA, an autonomous underwater vehicle (AUV).

We modularize our approach into vision and control modules and perform experiments in open water using three PID controllers and a monocular camera.

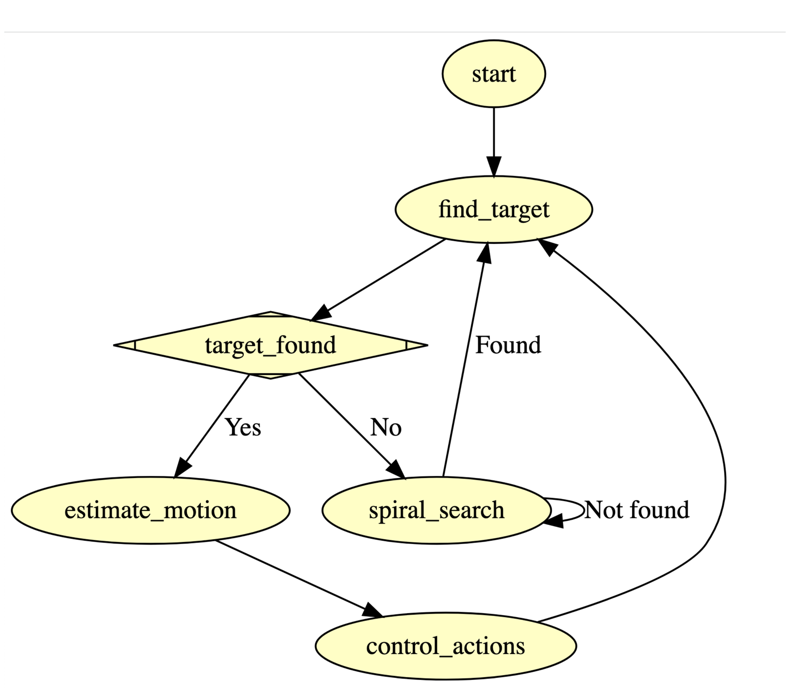

Fig 1: Block diagram of our proposed diver tracking system.

Vision Module

We employ a tracking-by-detection approach using:

- **YOLOv7** [1] for diver detection. We evaluated seven state-of-the-art object detectors and identified YOLOv7 as the highest-performing one, considering both speed and average precision (AP).
- **SORT** [2] for tracking detections between frames. We adapt SORT to provide temporally smooth inputs to the controllers, estimate missed detections, and remove false detections.

Our vision module is trained on the **VDD-C dataset** [3], which contains over 100,000 images of scuba divers in ocean and pool environments.

We introduce **frame-to-frame variance** as a metric to evaluate the temporal stability of object detectors. We show that SORT significantly reduces frame-to-frame variance during tracking sequences.

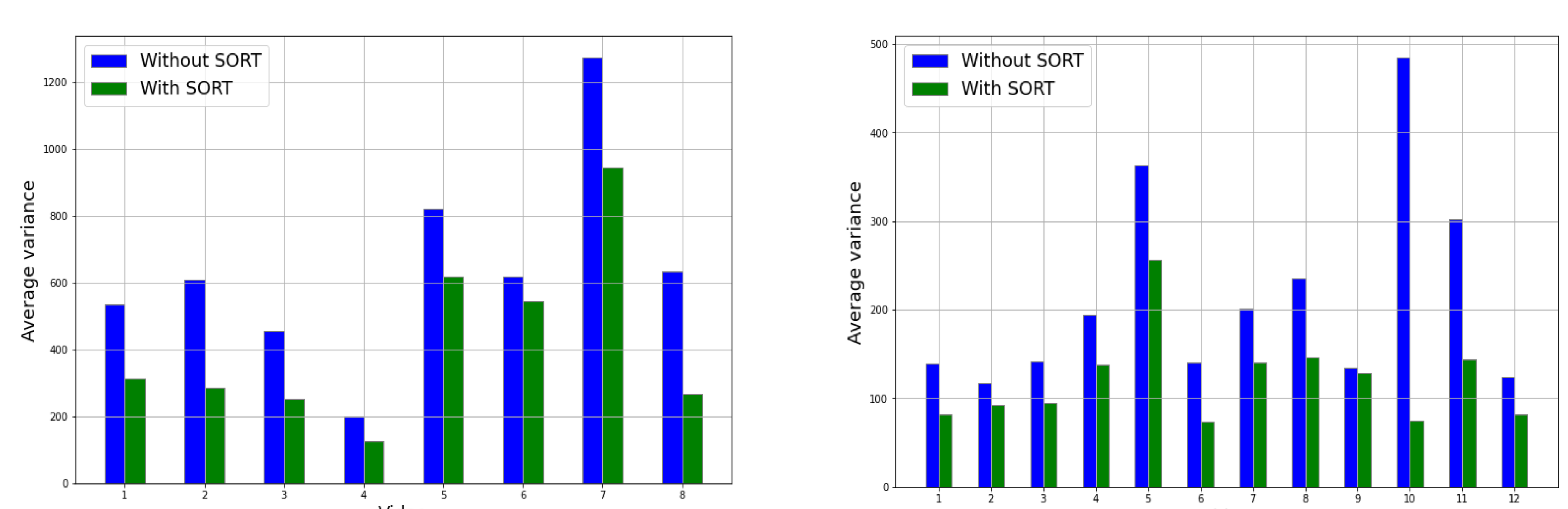

Fig 2: SORT reduces the frame-to-frame variance of videos from the VDD-C test set (left) and open water experiments (right). This increases the temporal stability of detections passed to the controllers enabling them to follow the diver successfully.

Control Module

PID controllers were used to control the pitch, yaw, and linear velocity of AQUA.

After a tracking failure, we implement **spiral search** [4] for target recovery. This algorithm searches left and right for the diver using exponentially increasing excursions away from the starting point.

For efficient recovery, we initialize the search with the last known location of the diver.

Model	AP$_{0.5}$	AP$_{0.75}$	AP$_{0.5:0.95}$	FPS
DETR-R50	0.904	0.507	0.508	40
DETR-DC5-R50	0.887	0.476	0.487	38
DETR-R101	0.903	0.487	0.496	27
DETR-DC5-R101	0.911	0.499	0.505	26
YOLOv4	0.898	0.431	0.468	29
YOLOv5	0.894	0.578	0.538	**107**
YOLOv7	**0.931**	**0.615**	**0.566**	70

Table 1: State-of-the-art object detectors evaluated on the VDD-C test set.

Fig 3: Our vision module identifying a diver in harsh conditions.

Open Water Experiments

We conduct our experiments in the ocean off the west coast of Barbados.

- 68 continuous tracking sequences recorded across 14 videos.
- Average continuous tracking length was 267 frames (33 seconds) with a maximum of 783 frames (97 seconds).
- Spiral search was effective in quickly recovering the diver between sequences.

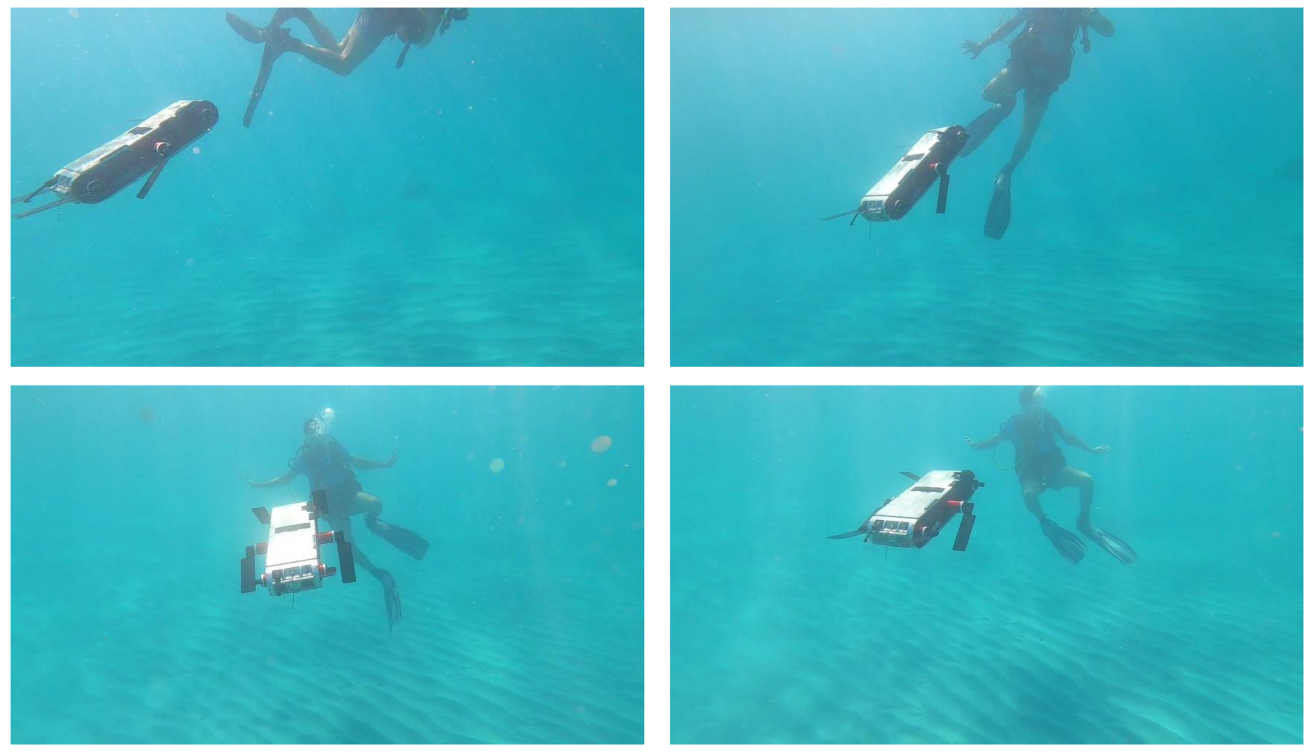

Fig 4: AQUA successfully tracking a diver in different directions.

Selected References

[1] Wang, Chien-Yao, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023.
[2] Bewley, Alex, et al. "Simple online and realtime tracking." *2016 IEEE international conference on image processing (ICIP)*. IEEE, 2016.
[3] de Langis, Karin, Michael Fulton, and Junaed Sattar. "Video Diver Dataset (VDD-C) 100,000 annotated images of divers underwater." (2021).
[4] Burlington, Scott, and Gregory Dudek. "Spiral search as an efficient mobile robotic search technique." *Proceedings of the 16th National Conf. on AI, Orlando Fl.* 1999.

McGill

MRL Mobile Robotics Lab at McGill University