TRACE (ECE-05)

Target Recognition Autonomous Control Engine for Search and Rescue

Elias Braddock, Drew Dolaway, Kim Huy Heng, Kaleb Lefkowitz and Liam Ostrander

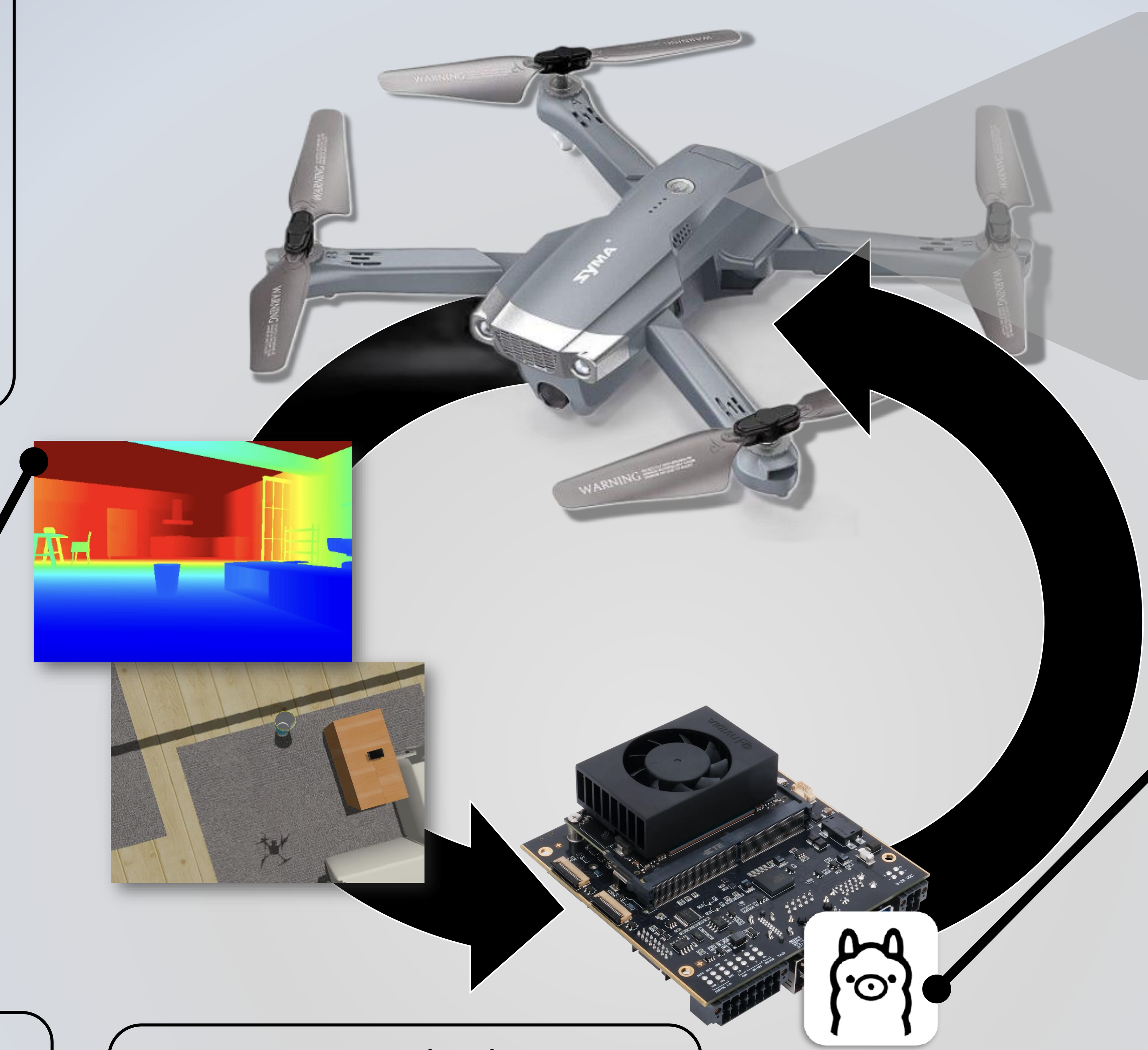
Advisor: Dr. Lifeng Zhou, Siwei Cai

Department of Electrical and Computer Engineering

Overview

Search and rescue (SAR) missions are time-critical and operate in hazardous environments. Although drones enhance situational awareness, current systems rely on human decision-making. High acquisition and maintenance costs of commercial platforms further restrict widespread deployment.

TRACE addresses these challenges by providing an open-source autonomous drone framework that integrates a **vision-language model (VLM)** with real-time navigation and control. By combing semantic visual reasoning with established autonomy techniques, TRACE enables scales, low-cost autonomous operations.



Gazebo

Gazebo is a physics-based virtual environment for **robotics simulation**. It can model terrain obstacles, lighting and gravity. Gazebo provides us with rigid-body **flight dynamics and collisions** as well as generates realistic sensor data **(IMU, camera, depth, GPS)**.

Depth Camera

The main sensor that our system relies on is the **Intel RealSense depth camera**, which provides **high-resolution depth measurements** and **RGB imagery** for robust environmental perception. The camera enables accurate **distance estimation** and **scene understanding**, supplying critical visual data to the downstream perception pipeline. This depth-aware sensing improves **target localization accuracy** and overall system reliability in real-world conditions.

Vision Language Model (VLM)

The backbone of our engine relies on **Vision Language Models (VLMs)**. VLMs extend **Large Language Models (LLMs)** with the ability to interpret visual inputs and generate natural-language responses conditioned on both images and text prompts. In our system, images from the depth camera and the downward-facing camera are provided to the VLM, which determines whether the **target object** has been detected, serving as the **primary inference mechanism**. To implement this, we use **OLLama** to serve models locally via a **RESTful API**, allowing us to select from a library of models; we chose **Qwen 3** for its lightweight size and high performance.

Applications

- In **search and rescue operations**, first responders and SAR teams can use autonomous drones to quickly locate victims in hazardous or hard-to-reach areas.
- For **disaster response and damage assessment**, emergency management agencies can efficiently evaluate affected infrastructure following earthquakes, floods, or fires.
- **Public safety and emergency support** benefits law enforcement and fire departments by providing risk-free aerial reconnaissance in unsafe environments.
- **Wilderness and maritime search** operations, such as those conducted by coast guard units or park services, can leverage TRACE to cover large, remote areas more effectively.

Future Plans

- **Real-World Test Platform**: Transition from simulation to physical drone testing to validate system performance
- **Expand Autonomous Capabilities**: Incorporate higher-level mission planning, adaptive search strategies, and dynamic re-tasking without human input
- **Fine-Tune VLM Understanding**: Improve VLM performance through task-specific fine-tuning for reliable target recognition and semantic reasoning
- **Real-Time Mapping & Localization**: Integrate online SLAM to enable accurate mapping and localization in unknown and GPS-denied environments.

Auto Navigation

The system begins by collecting **depth and RGB data** from the onboard forward-facing Intel RealSense D435i and downward-facing Arducam OV9281 cameras. This sensor data is processed to extract depth information and perform **semantic target identification** using a **vision-language model**. The processed depth data is used for object avoidance and visual-inertial SLAM, allowing the drone to **autonomously roam** and **construct a map of the surrounding environment**. The decision layer confirms detected targets via the downward facing cam, forcing the drone to **hover in place upon target detection**. The SLAM map stores environmental context, giving guidance for detecting future targets. These high-level navigation commands are executed through PX4 SITL, which handles low-level flight control, resulting in **autonomous navigation and real-time 2D occupancy grid mapping of the environment.**