ICS 483: Computer Vision (Fall 2025)

Final Project Proposal

Christian Moore

November 3, 2025

1 Motivation and Overview

In field robotics, effective human–robot interaction (HRI) is often limited by the need for physical controllers or laptop-based interfaces, which can be cumbersome during mobile operation. This project proposes a vision-based hand gesture recognition system to enable natural, hands-free control of mobile robots without requiring a tethered computer. Using an iPhone camera or similar monocular input, the system will interpret hand gestures in real time to issue control commands to the robot. This approach allows operators to interrupt, redirect, or command the robot directly through visual cues, ideal for dynamic or unstructured environments where mobility and adaptability are key.

2 Objectives

- Develop a computer vision pipeline for real-time hand gesture detection and classification using hand landmark estimation.
- Implement a gesture-to-command mapping module to trigger specific robot actions (e.g., stop, turn, follow).
- Integrate the vision system with a mobile robot control stack, either through:
  - (a) Hard-coded interrupts (gesture maps directly to an action), or
  - (b) A Vision-Language-Action (VLA) interface, where recognized gestures are semantically interpreted by an action expert model.
- Evaluate system performance on recognition accuracy, latency, and usability in outdoor or semi-structured field environments.

3 Methods

The system will use MediaPipe Hands for 2D/3D hand keypoint detection from live or recorded video captured by the iPhone camera. Extracted landmarks will be used to compute relative distances and angles between fingers to classify gestures. A rule-based approach will first be used for quick prototyping (e.g., open palm = "stop," fist = "start," pointing = "turn"). To reduce noise, temporal smoothing and confidence-based filtering will be applied across consecutive frames. There may be an optional ML classifier to improve generalization.

4 Expected Outcomes

By the end of the project, the system will demonstrate hands-free, real-time gesture control of a mobile robot, showing how intuitive visual communication can replace traditional hardware interfaces. The prototype will provide a framework for future HRI systems capable of integrating multimodal input (gestures, language, gaze) for adaptive robot behavior.

5 Impact

This project advances the intersection of computational imaging and human-robot interaction by enabling flexible, nonverbal robot control through vision alone. Such systems could be invaluable in search-and-rescue, environmental monitoring, or agricultural robotics, where users need fast, reliable, and portable interaction without carrying additional equipment.