Vision-Based Social Robot Navigation with 6D Head Pose Estimation and F-Formation Analysis

Georgina Woo
Department of Computer Science, Hunter College

March 28, 2025

Problem & Hypothesis

Traditional robot navigation algorithms primarily focus on obstacle avoidance and predefined path planning, often failing to account for the presence, movement, and social behavior of humans. This limitation makes it difficult for robots to navigate smoothly in human-populated environments, where movement patterns are dynamic and shaped by social context. Many existing approaches assume that a robot's perception is around human eye level, while robots with lower or higher vantage points must recognize and respond to human interactions from different perspectives while still navigating safely and effectively.

Methods

This research proposes a vision-based framework for socially-aware navigation, integrating pre-trained 6D head pose estimation to infer human attention and movement intentions, along with F-formation analysis to classify common patterns of social interaction. By identifying social formations—such as face-to-face conversations, side-by-side walking groups, and open clusters—the system enables robots to predict pedestrian movement and adapt their navigation strategy accordingly. The framework is being developed and tested on the Misty II robot, using single-stream visual data.

Results (Expected Findings)

Evaluations will assess the system's accuracy in detecting social formations, efficiency in navigation paths, and adherence to social norms, with a focus on the unique perception challenges posed by lower or higher vantage points.

Conclusion & Significance

By enabling robots to proactively adjust to human motion and social context, this work aims to advance human-robot interaction and socially-aware navigation, paving the way for safer, more intelligent robot behavior in shared spaces. Future research will explore multi-agent coordination, semantic scene understanding, and predictive motion planning based on group dynamics and individual attention shifts.