CENTER ON Pervasive Personalized Intelligence

Together We Go Farther

Anomaly Detection and Explanation via Observation Aggregation and Visual-Language Models

Shivendra Agrawal (5th yr. Ph.D.), Ashutosh Naik (MS), Bradley Hayes (Advisor)
University of Colorado Boulder

Problem and Industrial Relevance

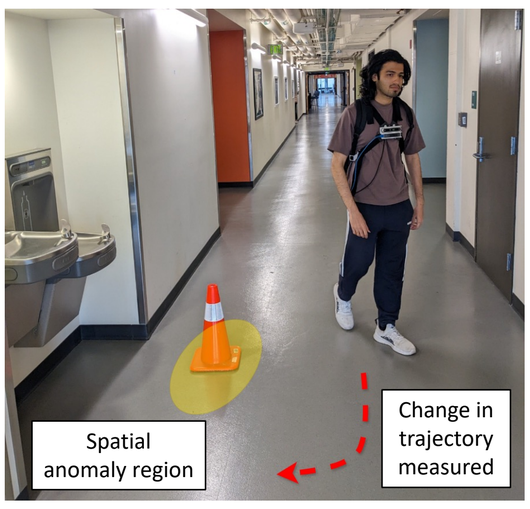
Crowd behavior changes can provide insight into anomalies

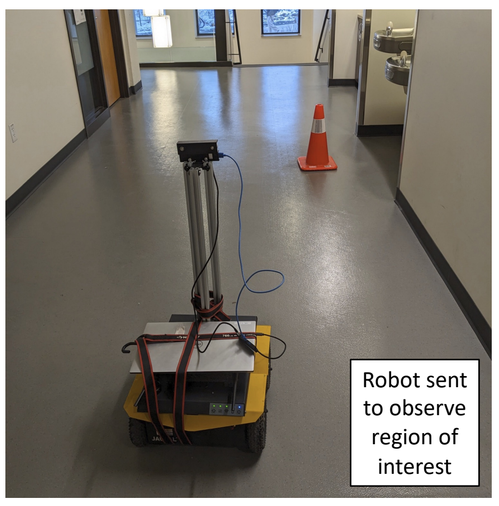
We can explain these anomalies by

- **Measuring change in people's behavior** - People's movement behavior can change because of obstacles, safety hazards, or change in goals resulting in anomalous trajectories.
- **Using VLM's to explain anomaly** - VLMs can leverage semantic knowledge of the world to explain these anomalous patterns.

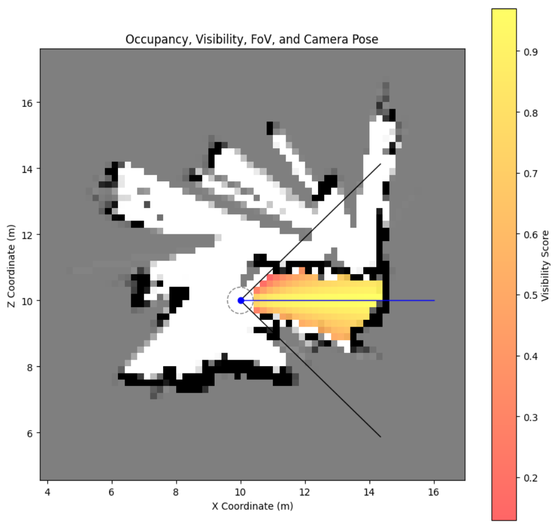
System Overview - Robot monitoring system

1. We created a system to measure changes in people's spatial movement patterns over time.
2. Our system locates the anomalous regions.
3. A robot is then sent to regions of interest for surveillance.
4. It finds the relevant observations corresponding to the anomaly.
5. A VLM is used to explain the anomaly.

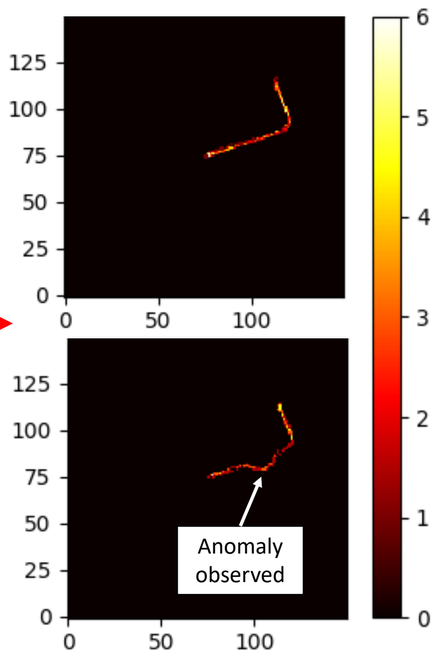




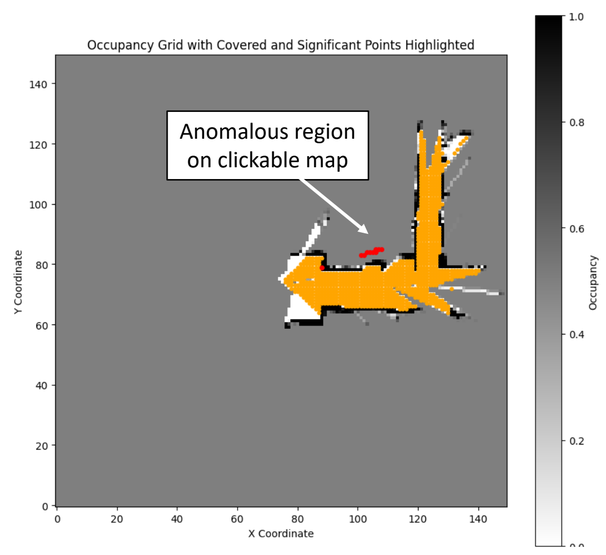
Approach and Case Studies



- Robot maps the surrounding using RealSense D455 & T265 cameras.
- It also computes the observability for each cell.

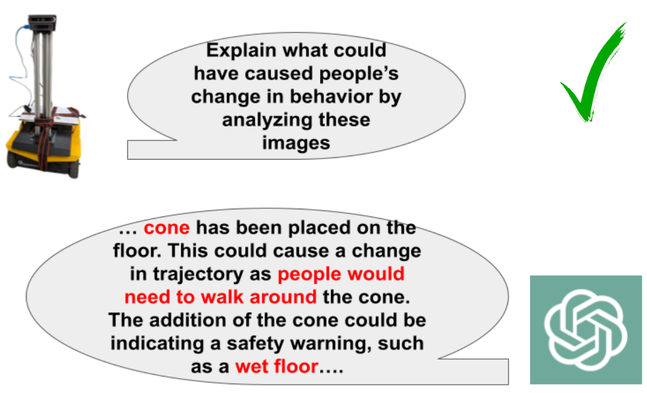


Change in trajectories observed

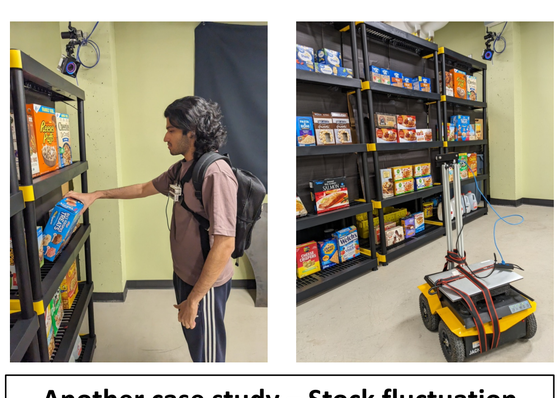


- Anomaly is overlaid on an interactive map.
- We observe a drift in localization.
- Our system optimizes by searching for the nearest observable cell.

System queries GPT-4V with the corresponding relevant images



GPT-4V explains the potential cause for the anomaly by comparing 'regular' and after monitoring images

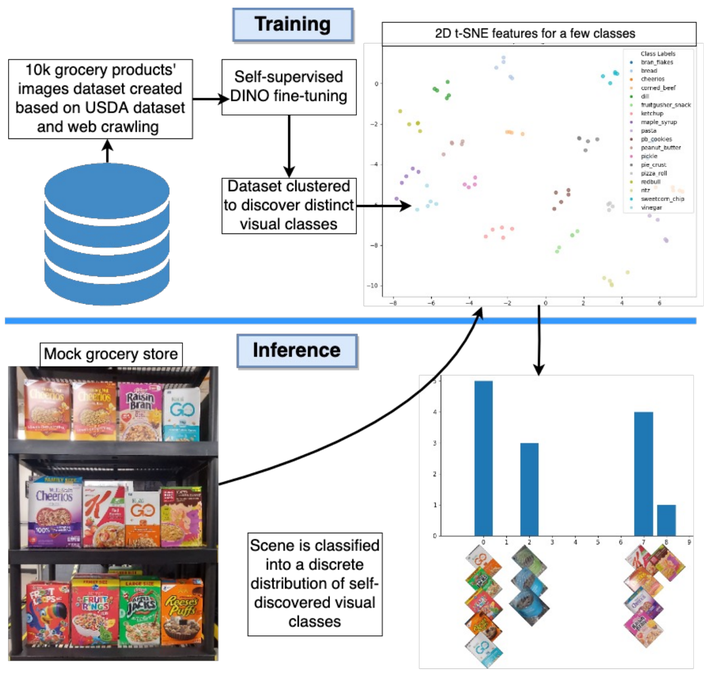


Another case study – Stock fluctuation
System observes areas with most time spent and flags those as anomaly.

System queries GPT-4V with the corresponding relevant images

GPT-4V cannot reason about stock reliably

Work in Progress



- Stock fluctuations cannot be reliably estimated by VLMs.
- We develop a novel way of quantifying distributions of products in a scene.
- We then compare the two distributions above using Jenson-Shannon Divergence.
- For example – The above two scenes result in JSD value of **0.06** (0 = totally similar, 1 = totally different).

Executive Summary

1. We present a robot monitoring system that reasons over people's trajectory and observation data aggregation to infer anomalous behaviors.
2. It then calculates the most relevant images and queries a pre-trained VLM (GPT-4V in this case) to explain the potential cause of the anomaly.
3. Through a preliminary testing, we show that the system is able to explain anomalies concerning a common scenario.
4. GPT-4V falls short on quantifying stock fluctuations.
5. We develop a novel way to quantify and compare stocks.
6. This work also shows that a more robust localization is also needed for practical applications.