NO REVISIONS

Project Proposal "Self-Guiding Sentinels: an Accurate Physical Attack Surveillance System"

Sarah Mundy, *srm2238*

Summary

This proposal outlines the development of Self-Guiding Sentinels, an innovative computer vision system designed to detect and respond to physical tampering of surveillance equipment. Unlike existing self-monitoring systems that focus primarily on network infrastructure, Self-Guiding Sentinels extends protection to physical surveillance assets by differentiating between legitimate maintenance and malicious tampering. When attacks are detected, the system triggers contextually appropriate audio responses, including anthropomorphic expressions of "pain", to deter attackers and alert security personnel. The project leverages cutting edge deep learning architectures optimized for real-time surveillance video processing based on diffusion techniques, combined with sophisticated event classification to minimize false alarms while maximizing accuracy and minimizing the costs associated with running such a system.

Project Objectives

I plan to develop a robust computer vision model capable of distinguishing between deliberate physical attacks on surveillance equipment (specifically cameras), accidental interference or environmental factors, legitimate maintenance activities, and sophisticated evasion attempts designed to mimic maintenance. To facilitate this, I will augment a publicly available dataset by creating a synthetic dataset representing diverse tampering scenarios to strengthen classification capabilities and record a small sample of actual scenarios.

Technical Implementation

For this project I will be using the PyTorch framework as it has a strong community for video processing, is the current SOTA for research experimentation, and allows for flexible custom dataset implementations.

I plan to use a hybrid approach for the architecture. The backbone will be a Vision Transformer as it is better at capturing long-range dependencies, and has a strong performance on video understanding tasks. I will add a lightweight diffusion component specifically for modeling uncertainty in predicted trajectories. Diffusion models also excel at generating potential future frames for improved predictions and filling in occluded views. This gives us better real-time performance than a fully fledged diffusion model, and enhances the ability to handle ambiguous situations.

Dataset

There will be three data sources used for this project:

1. The publicly available UHCTD: A comprehensive dataset for camera tampering erection by the Quantitative Imaging Laboratory. (http://qil.uh.edu/main/datasets/)
2. Building on the synthetic dataset generated in the midterm paper experiment with better examples for model training
3. Staged scenarios filmed specifically for this project using volunteers

This small dataset will include 10-20 hours of footage, 5 attack scenarios, 5 false attack scenarios. The data will be manually annotated for pre-attack indicators.

Evaluation Metrics

The performance of Self-Guided Sentinels will be evaluated based on the following metrics:

1. Detection accuracy using an F1 score for tampering detection
2. False positive rate
3. False negative rate
4. Response latency
5. Evasion resistance
6. Computational efficiency based on CPU/GPU resource utilization

Deliverables

- A proof of concept hybrid model using the above hybrid computer vision architecture building on the my previous proof of concept diffusion model in python
- A Minimal Viable Dataset, which contains both synthetic and staged scenarios in video or frames.
- Prototype Response System (without speaker connection)
- Comprehensive Analysis Report

As this is a high-level perspective, my strategy could change based on new knowledge gained during this process.