PhD Forum: Decision-Theoretic Coordination and Control for Active Multi-Camera Surveillance

Prabhu Natarajan
Department of Computer Science, National University of Singapore, Republic of Singapore
prabhu@comp.nus.edu.sg

***Abstract*—In this thesis, we present novel decision-theoretic multi-agent approaches for controlling and coordinating multiple active cameras in surveillance. Decision-theoretic approaches models the interaction between active camera network and the uncertain surveillance environment effectively. The goal of the surveillance is to maximize the number of targets observed in active cameras with guaranteed image resolution. We enumerate the practical issues in active camera surveillance and discuss how these issues are addressed in our decision-theoretic approaches. The existing camera control approaches have serious limitations in terms of scalability in number of targets. Where as in our approaches, the scalability in number of targets has been improved by exploiting the structure and properties that are present in our surveillance problem. We proposed two novel decision-theoretic frameworks: Markov Decision Process (MDP) and Partially Observable Markov Decision Process (POMDP) frameworks for coordinating active cameras in *fully observable* and *partially observable* surveillance settings.**

I. Introduction

Active cameras are becoming popular in today's surveillance systems. These active cameras are endowed with pan-tilt-zoom capabilities, which can be exploited to provide high-quality surveillance of targets in many real-world applications: activity/intention tracking and recognition, biometric analysis like face recognition, surveillance video mining, forensic video analysis/retrieval, sports video, and many more. Most of the activities of interests are sporadic in nature and are scattered across the surveillance environment. Therefore, the active cameras can be steered and zoomed to focus on these activities at a high-resolution. Manual control of these cameras in the above applications becomes difficult, especially when the number of cameras and targets increases. In order to achieve effective real-time surveillance, an efficient collaborative mechanism is required to control and coordinate these cameras' actions, which is the main focus of this thesis.

This thesis aims to address the following issue in surveillance: *How can a network of active cameras be coordinated and controlled to maximize the number of targets observed with a guaranteed image resolution?* Monitoring a set of targets in an environment with a guaranteed resolution is an important surveillance task. Coordinating active cameras in order to observe these targets with a guaranteed image resolution is challenging and non-trivial. This is due to the following issues in designing a coordination framework for active cameras in surveillance:

Multiple sources of uncertainties: The surveillance environment is fraught with multiple sources of uncertainties such as targets' motion and location, camera sensor noise, and many more. These uncertainties in the surveillance environment make it hard for the active cameras to know where to observe in order to keep these targets within their fields of view (fov) and they may consequently lose track of the observed targets.

Camera - Target ratio: In practice, the number of targets to be observed is much greater than the number of available active cameras. When the number of targets increases, the camera coordination framework, if poorly designed, tends to incur exponentially increasing computational time, which degrades the performance of the entire surveillance system.

Non-trivial trade-off: There exists a non-trivial trade-off between maximizing the expected number of targets and the resolution of these observed targets. That is, increasing the resolution of observing some targets through panning, tilting, or zooming may result in the loss of other targets being tracked. Therefore it is necessary to address this trade-off in the underlying camera coordination framework.

Scalability: The camera coordination framework should be scalable with an increasing number of targets and cameras. That is, the computational time required for calculating optimal control decision for the cameras should be made in polynomial time for increasing number of targets and cameras.

Surveillance environment: Most of the real-world surveillance environment contains obstacles like pillars, walls, barriers, etc. that occludes the fov of the cameras. Therefore it is highly impractical for the cameras to keep the targets always in its fov. This can also be due to privacy issues in monitoring certain regions of the environment. Such regions in the surveillance environment where the targets cannot be observed in any of the cameras due to obstacles or privacy issue is termed as *blind regions* and the surveillance environment with *blind regions* is called as *partially observable* environment. Hence when the targets are in the *blind regions*, the camera coordination framework has no clue on these targets and results in performance degradation of surveillance system.

Therefore controlling and coordinating network of active cameras in order to maximize the number of observed targets is challenging and needs significant attention in research.

II. Decision-Theoretic Approach

We propose novel decision-theoretic multi-agent approaches for controlling multiple active cameras in surveillance. Decision-theoretic approaches provides formal, principled frameworks to coordinate the planning of active cameras control decisions under stochastic, partially observable environments (e.g., uncertainty in targets motion and locations) in achieving the desired surveillance objective/task. It models the surveillance task as a stochastic optimization problem in which optimal actions of the cameras are determined such that the utility of the surveillance is increased. The goal of the surveillance system is to *maximize the number of targets observed with guaranteed image resolution*. That is, the network of active cameras coordinates to obtain images/videos of the moving targets in the surveillance environment with guaranteed image resolution. This formalization exploits the inherent properties and structures that are present in our surveillance problem, in order to scale the framework for increasing number of targets - i.e., we exploit the conditional independence property between individual targets and active cameras in our surveillance problem. Specifically we propose two novel decision-theoretic frameworks to control active cameras in surveillance: (a) Markov Decision Process (MDP) for *fully observable* surveillance environment (i.e., the active cameras are supported by static wide-view cameras that can observe the surveillance environment completely and track the targets at low resolution at every instance of time); and (b) Partially Observable Markov Decision Process (POMDP) for *partially observable* surveillance environment (i.e., we do not have static cameras that can observe the surveillance environment completely and these environments have *blind regions* where the camera network has no clue on the targets when they are in these regions). In both MDP and POMDP frameworks, computing control actions for active cameras incurs time that is linear in the number of targets observed during the surveillance.

A. *Markov Decision Process framework*

A novel Markov Decision Process (MDP) framework [1] has been proposed to control active cameras in a *fully observable* surveillance environment, i.e., the location, direction and velocity of the moving targets are estimated from a set of wide-view static cameras that are calibrated site wide. In this environment, the targets are assumed to be visible to the static cameras at every instance of time and based on the observations from the static cameras, the MDP framework directs the active cameras to observe the targets in high-resolution. In order to direct the active cameras to the predicted locations of the target, we proposed a greedy solution (i.e., one step look-ahead of target's motion) to solve the underlying MDP. Specifically the MDP framework resolves some of the above mentioned issues in the following ways: (a) the motion of the targets are modeled probabilistically; (b) the non-trivial trade-off has been addressed by controlling the active cameras to maximize the number of targets by guaranteeing the predefined image/video resolution; (c) the scalability in number of targets has been improved by exploiting the properties that are present in our surveillance problem; and (d) in order to compute optimal control decisions for cameras in real-time, we pre-compute the solutions off-line and do a look-up operation on our stored solutions during the surveillance. As shown in simulation, our MDP framework can achieve high-quality surveillance of up to 50 targets in real time and the computational time increases linearly with number of targets.

B. *Partially Observable Markov Decision Process framework*

A novel Partially Observable Markov Decision Process (POMDP) framework [2] has been proposed to control active cameras in a *partially observable* surveillance environment, i.e., we do not have static cameras that can completely observe the surveillance environment. Hence, the targets' informations are observed directly from the active cameras. Also in this setup, the targets may not be continuously observed in any of the active cameras due to the *blind regions* in the surveillance environment. This setup is more realistic because, many real world environments (like airports, railway stations, schools and university campuses, etc.) have occlusions due physical structures like walls and pillars, and also restricted regions where the cameras cannot be placed. This framework resolves some of the surveillance issues as follows: (a) the targets motion uncertainty is modeled by probabilistic motion model; (b) the targets location and direction uncertainty are modeled by having a probabilistic distribution over the targets location and direction (known as *belief* of targets); (c) camera sensor noise is modeled by having a probabilistic observation model; (d) the non-trivial trade-off is modeled by coordinating the cameras action such the expected number of targets are maximized by maintaining a guaranteed image/video resolution; (e) scalability in number of targets is improved by exploiting the conditional independence properties in our surveillance problem and (f) the optimal cameras' actions are computed in real-time by using sparse data structures to store and manipulate only the non-zero probability values. As shown in simulation, our POMDP framework can achieve high-quality surveillance of up to 20 targets in real time and the computational time increases linearly in number of targets.

III. Conclusion and Future works

In this thesis, we propose novel decision-theoretic approaches to control and coordinate multiple active cameras in surveillance. We try to maximize the number of targets observed in active cameras by guaranteeing the predefined image resolution. We proposed two novel decision-theoretic frameworks: MDP and POMDP frameworks to control multiple active cameras in *fully observable* and *partially observable* surveillance environments. We exploit the conditional independence property between individual targets and active cameras to improve the scalability in number of targets. In future we would like to improve on scalability in number of cameras and show the generality of our approaches for different multiple goals of the surveillance.

References

[1] P. Natarajan, T. N. Hoang, K. H. Low, and M. Kankanhalli. Decision-theoretic approach to maximizing observation of multiple targets in multi-camera surveillance. In *Proc. AAMAS*, pages 155–162, 2012.

[2] P. Natarajan, T. N. Hoang, K. H. Low, and M. Kankanhalli. Decision-theoretic coordination and control for active multi-camera surveillance in uncertain, partially observable environments. In *Proc. ICDSC*, 2012.