

Scalable Decision-Theoretic Coordination and Control for Real-time Active Multi-Camera Surveillance

Prabhu Natarajan[§], Trong Nghia Hoang[†], Yongkang Wong[§],
Kian Hsiang Low[†] and Mohan Kankanhalli[†]

Interactive Digital Media Institute [§], Department of Computer Science[†]
National University of Singapore, Republic of Singapore

{prabhu,yongkang.wong}@nus.edu.sg[§],{nghiaht,lowkh,mohan}@comp.nus.edu.sg[†]

ABSTRACT

This paper presents an overview of our novel decision-theoretic multi-agent approach for controlling and coordinating multiple active cameras in surveillance. In this approach, a surveillance task is modeled as a stochastic optimization problem, where the active cameras are controlled and coordinated to achieve the desired surveillance goal in presence of uncertainties. We enumerate the practical issues in active camera surveillance and discuss how these issues are addressed in our decision-theoretic approach. We focus on two novel surveillance tasks: maximize the number of targets observed in active cameras with guaranteed image resolution and to improve the fairness in observation of multiple targets. We discuss the overview of our novel decision-theoretic frameworks: Markov Decision Process and Partially Observable Markov Decision Process frameworks for coordinating active cameras in uncertain and partially occluded environments.

Categories and Subject Descriptors

I.4.8 [Scene Analysis]: Tracking; I.2.9 [Robotics]: Commercial robots and applications, Sensors

Keywords

Active camera networks, Smart camera networks, Multi-camera coordination and control, Surveillance and security

1. INTRODUCTION

Surveillance security is becoming a part of the building infrastructures due to recent security threats like the Mumbai terrorist attack and Boston bomb blasts. Central to the problem of surveillance is that of monitoring, tracking, and observing multiple mobile targets of interest distributed over a large-scale obstacle-ridden environment (e.g., airport terminals, railway and subway stations, bus depots, shopping malls, school campuses, military bases, etc.). It is often necessary to acquire high-resolution videos/images of these targets for supporting real-world surveillance applications like

activity/intention tracking and recognition, biometric analysis like target identification and face recognition, surveillance video mining, forensic video analysis/retrieval, among others. Traditional surveillance systems consists of large number of fixed/static CCTV (Closed Circuit Television) cameras that are placed to constantly focus at the selected important locations in the buildings like entrance/exit, lobby, etc. A relatively large network of such cameras has to be installed in order to observe the targets in any region of the environment at high-resolution, which is impractical in terms of equipment, installation, and maintenance costs.



Figure 1: Axis 214 PTZ cameras.

The use of active PTZ (Pan/Tilt/Zoom) cameras is becoming an increasingly popular alternative to that of fixed/static cameras for surveillance because the active cameras are endowed with pan-tilt-zoom capabilities that can be exploited to focus on and observe the targets at high image/video resolution. Hence, fewer active cameras need to be deployed to be able to capture high-resolution images/videos of the targets in any region of the environment. 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 targets and cameras increases. Figure 1 shows the images of Axis 214 PTZ cameras that are widely used in research and commercial purposes.

This paper aims to address the following central problem in surveillance: How can a network of active cameras be coordinated to monitor a set of moving targets with a guaranteed image resolution? Monitoring a set of targets 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 practical 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' stochastic motion, unknown targets'

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...\$15.00.

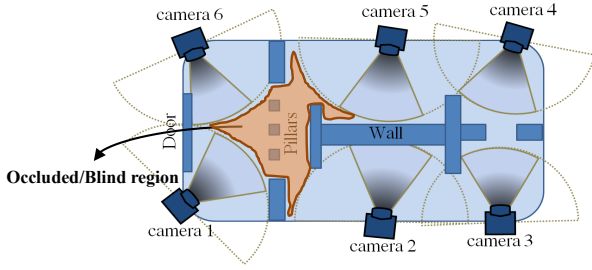


Figure 2: An example of *partially observable* surveillance setting with 6 PTZ cameras.

locations, noisy camera observations, etc. These uncertainties in the surveillance environment make it difficult for the active cameras to know where to observe in order to keep the targets within their fields of view (fov). Consequently, they may also lose track of the observed targets.

- **Camera - Target ratio:** In practice, the number of targets to be observed is usually 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.
- **Trade-off between maximizing the expected number of targets and the resolution of these observed targets:** Increasing the resolution of observing some targets through panning, tilting, or zooming may result in the loss of other targets being tracked; and increasing the number of targets to be observed by decreasing the zoom level, may result in the decrease of the resolution of these targets. 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. The computational time required for calculating optimal control decisions for the cameras should be made in polynomial time for increasing number of targets and cameras.
- **Real-time:** The control decisions for these cameras should be made proactively in real-time.
- **Occlusions:** Many real-world surveillance environments contain obstacles like pillars, walls and barriers that occlude the FoV of some or perhaps even all of the cameras. This can also be due to privacy issues in monitoring certain regions of the environment. Therefore, it is highly impractical for the cameras to persistently track the observed targets in such environments. The regions where the targets cannot be observed by any of the cameras due to obstacles or privacy issue are referred as *blind regions* and the surveillance environment with *blind regions* is called as *partially observable* environments (see Figure 2). Hence when the targets are in the *blind regions*, the camera coordination framework has no information about these targets which causes performance degradation of the surveillance system.
- **Fairness property:** Fairness is a vital property in active camera surveillance where the active cameras are coordinated to observe the targets in the surveillance environment such that no target is “starved” of observation by the cameras for a long time. When there is no fairness in coordinating active cameras, then one or more targets

may not be observed (i.e., starved) for long duration which may lead to a loophole in surveillance, i.e., the behavior of those targets will neither be monitored nor recorded in high-resolution. Hence it is necessary to incorporate fairness property in the coordination framework.

Therefore, the problem of controlling and coordinating a network of active cameras in order to monitor a set of targets is challenging and needs significant research attention.

In this paper, we discuss our novel decision-theoretic multi-agent approach for controlling and coordinating multiple active cameras in surveillance that addresses the above mentioned issues. Decision-theoretic approach provides formal and principled frameworks to coordinate the planning of active cameras’ control decisions under stochastic and 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 utility of the surveillance can be modeled as formal objective functions, such that the active cameras are coordinated to achieve this high-level surveillance goal. The first goal of the surveillance system is to maximize the number of targets observed with guaranteed image resolution. We refer to this surveillance task as **maximizing observations of multiple targets (MOMT)** problem. That is, a network of active cameras coordinating to obtain images/videos of the moving targets in the surveillance environment with guaranteed image resolution. A drawback of this surveillance goal is that one or more targets may not be observed by the active cameras for long duration as the cameras are coordinated to focus or observe locations where there are more number of targets in the environment. To overcome this limitation, we propose fairness in observation of targets as our second goal of surveillance. That is, no target is “starved” of observation by active cameras for long duration. We refer to this tasks as **fairness in observation of multiple targets (FOMT)** problem.

Specifically, we discuss the existing works in active camera coordination and control (Section 2) and provide an overview of our novel decision-theoretic frameworks to control and coordinate multiple active cameras: (a) Markov Decision Process (MDP) framework to solve the MOMT task in *fully observable* surveillance environment (Section 3), (b) Markov Decision Process (MDP) framework to solve the FOMT task in *fully observable* surveillance environment (Section 3), and (c) Partially Observable Markov Decision Process (POMDP) to solve MOMT task in *partially observable* surveillance environment (Section 4). Our decision-theoretic formulations exploits the inherent properties and structures that are present in our surveillance problems, in order to scale the framework for increasing number of targets. That is, the assumption that the motion of each target is conditionally independent of other targets and cameras in the environment given the current position, direction and speed of that target is exploited.

2. CAMERA CONTROL & COORDINATION APPROACHES

In early days, the PTZ cameras are controlled manually by humans using the joy-stick and other hardware devices. Based on the videos from static cameras, the user in the security office controls the PTZ cameras to focus on regions of

interests in high-resolution [3]. For example, [6] developed a surveillance system known as FLYSPEC system that converts the user selected region of interest in the wide-view cameras into the corresponding PTZ control signals and sends it to the PTZ cameras.

The above setup was improved by controlling the PTZ cameras to automatically gaze the targets that are observed in the static cameras [10] and omni-directional cameras [5]. These static wide-view cameras and omni-directional cameras are calibrated with the PTZ cameras on a common ground plane coordinates. Based on this simple setup (i.e., coupling static wide-view cameras and PTZ cameras), there has been many works that have been explored in multi-camera control and coordination. From the survey, we can see that researchers have adopted different approaches to control and coordinate multiple active cameras such as control theory [16, 21], game theory [4, 18], state machine [15, 20], multi-agent system [9], probabilistic approach [17], and many other ad-hoc approaches [14] as well. In this paper, we discuss a novel decision-theoretic approach to control and coordinate multiple active cameras in surveillance [11, 12, 13]. The nature of the surveillance problem (i.e., choosing optimal actions of active cameras in presence of uncertainties) makes decision-theoretic approach an appropriate choice to control and coordinate active cameras in stochastic surveillance environment. Spaan et al. [19] have proposed a decision-theoretic approach to select camera views in the surveillance system. Their work is purely based on selecting one or more static cameras which is different from our work on controlling active cameras to accomplish a desired surveillance task. Our decision-theoretic approach offers some of the following advantages in surveillance:

- Decision-theoretic approach models the interaction between active camera network and the surveillance environment effectively. Specifically, 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.
- It provides formal, principled and rich mathematical models like Markov Decision Process (MDP), Partially Observable Markov Decision Process (POMDP), etc. for planning optimal control actions for cameras in presence of uncertainties like targets' motion and location, noisy camera observations, occlusions, etc.
- Multiple high-level surveillance goals can be defined formally as mathematical objective functions.

When decision-theoretic models are poorly designed or used naively for a surveillance problem, the state space explodes and hence computing optimal actions for these cameras becomes intractable. For example, [19] has serious limitation in terms of scalability in number of targets and cameras. Whereas in our work, we exploit the structures and properties of the underlying surveillance problem to improve the scalability issue.

3. MDP FRAMEWORK FOR COORDINATING CAMERAS

A novel Markov Decision Process (MDP) framework has been proposed to control active cameras in a *fully observable* surveillance environment, i.e., the locations, directions and speeds 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 proposed 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, greedy solution (i.e., one step look-ahead of target's motion) has been proposed 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 between maximizing the expected number of targets and the resolution of these observed targets 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.

Formally, the MDP framework is defined as a tuple $(\mathcal{S}, \mathcal{A}, T_f, R)$ consisting of:

- a set \mathcal{S} of joint states of active cameras (i.e., discrete configurations of pan/tilt/zoom value) and targets (i.e., discrete location, direction and speed) in the surveillance environment,
- a set \mathcal{A} of joint actions (i.e., commands to move the active cameras from one PTZ configuration to another) of active cameras,
- a transition function $T_f : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$ denoting the probability $P(S'|S, A)$ of switching from the current state $S \in \mathcal{S}$ to the next state $S' \in \mathcal{S}$ using the joint action $A \in \mathcal{A}$,
- a real-valued reward function $R : \mathcal{S} \rightarrow \mathbb{R}$ representing the high-level surveillance goal.

In the MDP framework, the policy function $\pi : \mathcal{S} \rightarrow \mathcal{A}$ maps from each state to a joint action of the cameras. Solving the MDP involves choosing the policy that maximizes the expected reward for any given state. The optimal greedy policy, denoted by π^* , maximizing the expected utility of the system in the next time step is given by

$$\pi^*(S) = \arg \max_{A \in \mathcal{A}} \sum_{S' \in \mathcal{S}} R(S') P(S'|S, A) .$$

For MOMT task, the reward function R is defined as the total number of targets that are observed by any of the active camera with guaranteed resolution. As mentioned previously, the drawback of this task is the lack of fairness in the observation of targets, i.e., targets are "starved" of observation from active cameras for long duration. Therefore in [13], we define FOMT task and formally realize a popular fairness metric in resource allocation problems known as max-min fairness, for achieving fairness in active camera surveillance. We formulate this max-min metric as a reward function and optimize it, such that no target is "starved" of observation by active cameras for long time.

The main challenge in the MDP is managing the state space \mathcal{S} . This is because the state space grows exponentially in the number of targets and active cameras. Hence, the policy computation time for both the surveillance tasks becomes exponential. In practice, the structure of the problem and environment can usually be exploited to reduce the number of states and the time required to compute the op-

timal policy. We have shown in [12] and [13] how the state space can be managed for our MOMT and FOMT surveillance tasks, respectively, and thus allowing the MDPs to be solved more efficiently.

4. POMDP FRAMEWORK FOR COORDINATING CAMERAS

A novel Partially Observable Markov Decision Process (POMDP) framework has been proposed to control active cameras in a *partially observable* surveillance environment, i.e., the case where we do not have static cameras that can observe the entire surveillance environment at a low-resolution. Hence, the targets' information are observed only through the active cameras. In such *partially observable* environment, the targets may not be continuously observed in any of the active cameras due to *blind regions* in the surveillance environment. This setup is more realistic because, many real world environments (like airports, railway and subway stations, schools and university campuses, etc.) have occlusions due to physical structures like walls and pillars, and also restricted regions where the cameras cannot be placed. Figure 2 shows an example of overhead view of a *partially observable* surveillance setup with 6 active cameras and the occlusion caused by pillars and non-overlapping active cameras.

This framework resolves some of the above-mentioned surveillance issues in the following ways: (a) the targets are tracked even when they are in blind spots by modeling the belief over the targets' states (i.e., locations, directions and speeds) and updating the belief using the Bayesian paradigm based on the probabilistic models for targets' motion and active cameras' observations; (b) the actions of the active cameras are coordinated to simultaneously improve the belief over the targets' states and maximize the expected number of targets observed with a guaranteed resolution; (c) the targets' motion uncertainty is modeled by a probabilistic motion model; (d) the noisy camera observation is modeled by having a probabilistic observation model; (e) the non-trivial trade-off between maximizing number of targets and the image resolution of observing these targets is addressed by coordinating the cameras' action such that the expected number of targets is maximized while maintaining a guaranteed image/video resolution; (f) the scalability in number of targets is improved by exploiting the properties in our MOMT task and (g) the optimal cameras' actions are computed in real-time by using sparse data structures to store and manipulate the probabilities.

Formally, the POMDP framework is defined as a tuple $(\mathcal{S}, \mathcal{A}, \mathcal{Z}, T_f, O_f, R)$ consisting of

- a set \mathcal{S} of joint states of active cameras and targets in the surveillance environment,
- a set \mathcal{A} of joint actions of active cameras,
- a set \mathcal{Z} of joint observations of the targets taken by the cameras,
- a transition function $T_f : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$ denoting the probability $P(S'|S, A)$ of going from the current joint state $S \in \mathcal{S}$ to the next joint state $S' \in \mathcal{S}$ using the joint action $A \in \mathcal{A}$,
- an observation function $O_f : \mathcal{S} \rightarrow [0, 1]$ denoting the probability $P(Z|S)$ of observing the joint observation $Z \in \mathcal{Z}$ given the joint state $S \in \mathcal{S}$,
- a real-valued objective/reward function $R : \mathcal{S} \rightarrow \mathbb{R}$ representing a high-level surveillance goal.

At any given time, the exact state of the environment is not fully observable to the POMDP controller. Instead, it maintains a belief B over the set \mathcal{S} of all possible states, that is, $B(S)$ is the probability that the environment is in the state $S \in \mathcal{S}$ such that $\sum_{S \in \mathcal{S}} B(S) = 1$. At every time step, the POMDP controller issues an action $A \in \mathcal{A}$ and makes an observation $Z \in \mathcal{Z}$ from the environment. Based on the action A and observation Z , the prior belief B is updated by Bayes' rule to the posterior belief B' as follows:

$$B'(S') = \eta P(Z|S') \sum_{S \in \mathcal{S}} P(S'|S, A)B(S)$$

where $\eta \triangleq 1/P(Z|B, A)$ is a normalizing constant. A policy π for the POMDP controller is defined as a mapping from each belief B to an action A . Solving a POMDP involves choosing the optimal policy π^* that maximizes the expected reward for any given belief B :

$$\pi^*(B) = \arg \max_{A \in \mathcal{A}} \sum_{Z' \in \mathcal{Z}} R(B')P(Z'|B, A).$$

When the number of targets and active cameras increases, the state space and hence the belief space of the POMDP grow exponentially. Therefore, computing the optimal policy incurs exponential time. Fortunately, by exploiting the structure of our surveillance problem, the optimal policy for a given belief B can be computed efficiently.

In this framework, we proposed a greedy solution for MOMT problem in *partially observable* surveillance environment that is scalable in number of targets. As shown in simulations in [11], our POMDP framework can scale up to 20 targets in real-time. The real camera experiments of our MDP and POMDP frameworks, show the feasibility of our decision-theoretic approach in real world surveillance.

5. EXPERIMENTS AND DISCUSSIONS

In this section, we present empirical evaluation of our decision-theoretic frameworks for MOMT and FOMT surveillance tasks. Our proposed frameworks are simulated in Player/Stage simulator to perform extensive experimentations and implemented using real Axis 214 PTZ cameras to demonstrate its feasibility in real surveillance system. In all our frameworks, we assume that the delay in moving the camera to a specified state is negligible as the state-of-the-art cameras are capable of panning/tilting at a speed of 360°/sec [1]. Also in our MDP frameworks, we assume that the static cameras are calibrated accurately such that the 3D position of the targets are minimal. This is achieved by placing the cameras at high altitude. In order to maintain the stability of our policies in all our frameworks, we compute the policies for every 3 to 5 seconds. It is important to point out that there is no standard benchmark surveillance environments and datasets for active camera networks to compare our proposed approach with the other systems in the literature. We measure the average of percentage of targets being observed (denoted by *PercentObs* in the graphs) by the active cameras over finite number of iterations for MOMT tasks and average minimum observation time (denoted as *Fairness* in graphs) of all the targets over finite number of iterations for FOMT tasks.

5.1 Simulation Results

Figure 3 shows the results of MDP framework for MOMT task in a corridor setup of 40×5 discrete target locations, 4 active cameras each with 3 discrete PTZ positions and 50 targets. It has been observed that our MDP framework

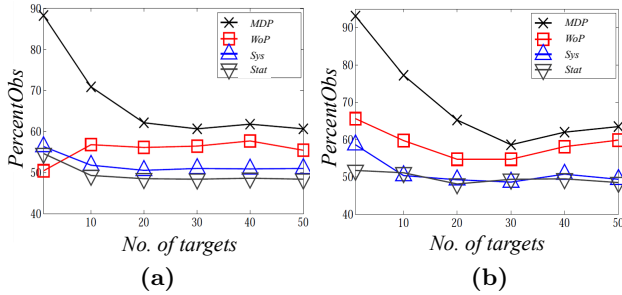


Figure 3: Performance evaluation of MDP framework for MOMT task in corridor setup: (a) non-clustered targets and (b) clustered targets. In the graph, *MDP* denotes our MDP framework for MOMT tasks, *WoP* denotes the baseline approach in which cameras are controlled without predicting the targets’ next locations; *Sys* denotes systematic panning of active cameras to all its states and *Stat* denotes static placement of active cameras.

outperforms other approaches in maximizing the number of targets because, the transition model in MDP helps to predict the next locations of the targets and the cameras are controlled to look at the predicted locations of the targets. The performance is much better when the targets move in clusters.

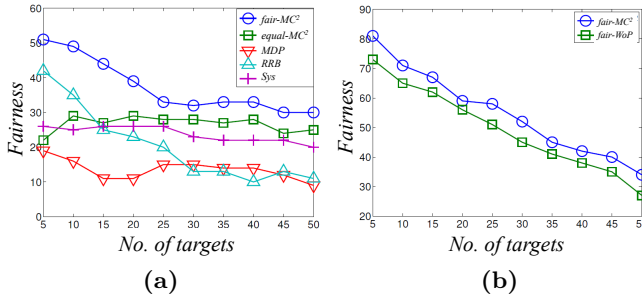


Figure 4: Performance evaluation of MDP framework for FOMT task in hallway setup. In the graphs, *fair-MC²* achieves our FOMT task, *equal-MC²* achieves equal observation times of all targets, *MDP* achieves our MOMT task, *RRB* observes targets based on round-robin fashion, *Sys* denotes systematic panning of active cameras to all its states, *WoP* denotes the baseline approach in which cameras are controlled without predicting the targets’ next locations.

Figure 4 shows the results of MDP framework for FOMT task in a hallway setup of 30×20 discrete target locations, 5 active cameras each with 3 discrete PTZ positions and 50 targets. It has been observed that our fairness metric achieves better fairness by maximizing the minimum observation times of the targets when compared to other approaches. Also the performance is much better when the cameras are controlled to focus on the predicted locations of the targets when compared to the *WoP* method where the cameras are controlled to observe the current locations of the targets.

Figure 5 shows the results of our POMDP framework for MOMT task in a corridor setup of 40×5 discrete target locations and hall setup of 40×5 discrete target locations with 4 active cameras each with 3 discrete PTZ positions and 20 targets. Our POMDP-based approach performs better than the *MP* approach due to its ability to keep track of the

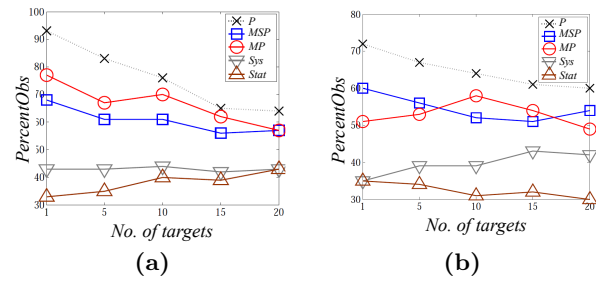


Figure 5: Performance evaluation of POMDP framework for MOMT task in (a) corridor and (b) hall setup. In the graphs, *P* denotes our POMDP framework for MOMT task; *MP* denotes our MDP framework for MOMT tasks without static cameras; *MSP* denotes our MDP framework for MOMT task where Gaussian noise is added to static camera observation; *Sys* denotes systematic panning of active cameras to all its states and *Stat* denotes static placement of active cameras.

targets’ locations and directions through its Bayesian belief update process. It outperforms the *MSP* approach because the observations (i.e., target’s location) taken by the active cameras in our POMDP controller are more accurate as compared to the noisy observations taken by the static cameras in the *MSP* approach. The *Sys* and *Stat* approaches suffer from performance degradation because the cameras are controlled independently of the targets’ information.

5.2 Real-camera Results

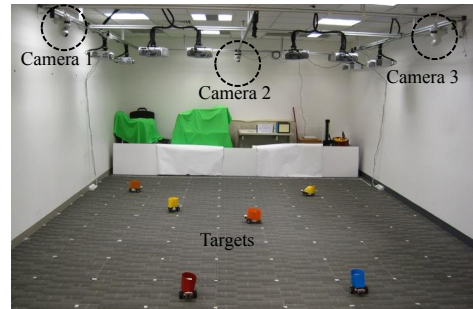


Figure 6: Real-camera experiment setups.

We have conducted real-camera experiments with 3 Axis 214 PTZ cameras to monitor up to 6 Lego robots (targets) in an indoor lab environment as shown in Figure 6. Each camera has 3 states. The states of the cameras are determined such that all the target’s locations of the environment can be observed at high-resolution by at least one camera. We have a static camera that can track these robots based on OpenCV Camshift tracker. For MDP framework we had a static camera that is calibrated to obtain the approximate locations of the robots at every time step. The results of our MDP and POMDP frameworks are shown in our demo video¹

We have also conducted preliminary study on tracking multi-targets using real-world surveillance footages. We employed a correlation filters based tracker [2] which reported to operate at 669 frames per second on single target. Our preliminary study shows that the correlation filters based tracker can achieve real-time tracking on 8 targets with image resolution of 1280×960 pixels at 25 fps. Our next work

¹<http://www.comp.nus.edu.sg/~lowkh/camera.html>

include the implementation of correlation filter based person detector and evaluation of person tracking for PTZ cameras.

6. CONCLUSIONS AND FUTURE WORKS

In this paper, we have given an overview of our novel decision-theoretic paradigm to control and coordinate multiple active cameras in surveillance. We focus on two surveillance goals: (a) maximize the number of targets observed in active cameras by guaranteeing the predefined image resolution and (b) achieve fairness in observation of multiple targets. 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 our works in the following ways: Firstly, we like to improve on the scalability in number of cameras and show the generality of our approaches for different goals of the surveillance. One way to do achieve this is to distribute the policy computation to individual cameras and compute the approximate global policy by passing messages between cameras. Secondly, our POMDP framework has been designed to solve only MOMT task. So we would like to extend our MDP framework for FOMT task to POMDP framework. Thirdly, we have tested our frameworks with robots in indoor setup. We would like to port all our frameworks and evaluate them in real-world setup. Lastly, we would like to consider consumer grade wearable visual sensors and mobile robots [7, 8] as part of the surveillance environment. While CCTV camera has been used as a primary surveillance platform, most of the surveillance environment are generally partially observable as stated in Section 1. The wearable visual sensors and mobile robots can provide addition information on the unobserved area. It is also more suitable to conduct high level visual analysis, such as person-of-interest detection, which can be used to prioritize the task for active multi-camera surveillance.

7. ACKNOWLEDGMENTS

This research was carried out at the NUS-ZJU Sensor-Enhanced Social Media (SeSaMe) Centre. It is supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the Interactive Digital Media Programme Office.

8. REFERENCES

- [1] AXIS 232D+ Network Dome Camera datasheet (<http://www.axis.com>).
- [2] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui. Visual object tracking using adaptive correlation filters. In *Proc. CVPR*, pages 2544–2550, 2010.
- [3] R. T. Collins, A. J. Lipton, H. Fujiyoshi, and T. Kanade. Algorithms for cooperative multisensor surveillance. *Proc. IEEE*, 89(10):1456–1477, 2001.
- [4] C. Ding, A. A. Morje, J. A. Farrell, and A. K. Roy-Chowdhury. Opportunistic sensing in a distributed ptz camera network. In *Proc. ICDCS*, pages 1–6, 2012.
- [5] R. Khoshabeh, T. Gandhi, and M. Trivedi. Multi-camera based traffic flow characterization & classification. In *Proc. Intelligent Transp. Sys. Conf.*, pages 259–264, 2007.
- [6] Q. Liu, D. Kimber, J. Foote, L. Wilcox, and J. Boreczky. FlySPEC: a multi-user video camera system with hybrid human and automatic control. In *Proc. Int. Conf. on Multimedia*, pages 484–492, 2002.
- [7] K. H. Low, W. K. Leow, and M. H. Ang, Jr. Task allocation via self-organizing swarm coalitions in distributed mobile sensor network. In *Proc. AAAI*, pages 28–33, 2004.
- [8] K. H. Low, W. K. Leow, and M. H. Ang, Jr. Autonomic mobile sensor network with self-coordinated task allocation and execution. *IEEE Trans. Syst., Man, Cybern. C*, 36(3):315–327, 2006.
- [9] T. Matsuyama and N. Ukita. Real-time multitarget tracking by a cooperative distributed vision system. *Proc. of IEEE*, 90(7):1136–1150, 2002.
- [10] C. Micheloni, E. Salvador, F. Bigaran, and G. Foresti. An integrated surveillance system for outdoor security. In *Proc. AVSS*, pages 480–485, 2005.
- [11] P. Natarajan, T. Hoang, K. Low, and M. Kankanhalli. Decision-theoretic coordination and control for active multi-camera surveillance in uncertain, partially observable environments. In *Proc. ICDCS*, pages 1–6, 2012.
- [12] 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.
- [13] P. Natarajan, K. H. Low, and M. Kankanhalli. No one is left "unwatched": Fairness in observation of crowds of mobile targets in active camera surveillance. In *Proc. ECAI(PAIS)*, 2014.
- [14] C. Piciarelli, C. Micheloni, and G. L. Foresti. PTZ camera network reconfiguration. In *Proc. ICDCS*, pages 1–7, 2009.
- [15] F. Z. Qureshi. Collaborative sensing via local negotiations in ad hoc networks of smart cameras. In *Proc. ICDCS*, pages 190–197, 2010.
- [16] V. K. Singh, P. K. Atrey, and M. Kankanhalli. Cooperative multi-camera surveillance using model predictive control. *Machion Vision Application*, 19(5-6):375–393, 2008.
- [17] E. Sommerlade and I. Reid. Probabilistic surveillance with multiple active cameras. In *Proc. ICRA*, pages 440–445, 2010.
- [18] B. Song, C. Soto, A. K. Roy-Chowdhury, and J. A. Farrell. Decentralized camera network control using game theory. In *Proc. ICDCS*, pages 1–8, 2008.
- [19] M. T. J. Spaan and P. U. Lima. A decision-theoretic approach to dynamic sensor selection in camera networks. In *Proc. ICAPS*, pages 279–304, 2009.
- [20] W. Starzyk and F. Z. Qureshi. Multi-tasking smart cameras for intelligent video surveillance systems. In *Proc. AVSS*, pages 154–159, 2011.
- [21] Y. Wang, P. Natarajan, and M. Kankanhalli. Multi-camera skype: Enhancing the quality of experience of video conferencing. In *The Era of Interactive Media*, pages 243–253, 2013.