No One is Left "Unwatched": Fairness in Observation of Crowds of Mobile Targets in Active Camera Surveillance
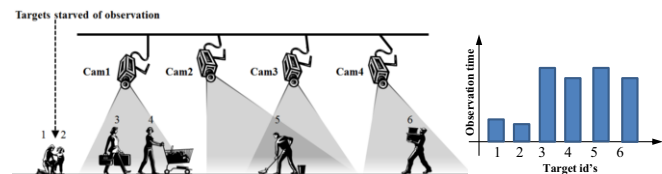
Prabhu Natarajan[1]   and   **Kian Hsiang Low**[1]   and   **Mohan Kankanhalli**[1]

Abstract.   Central to the problem of active multi-camera surveillance is the fundamental issue of fairness in the observation of crowds of targets such that no target is "starved" of observation by the cameras for a long time. This paper presents a principled decision-theoretic *multi-camera coordination and control* (MC^2) algorithm called fair-MC^2 that can coordinate and control the active cameras to achieve max-min fairness in the observation of crowds of targets moving stochastically. Our fair-MC^2 algorithm is novel in demonstrating how (a) the uncertainty in the locations, directions, speeds, and observation times of the targets arising from the stochasticity of their motion can be modeled probabilistically, (b) the notion of fairness in observing targets can be formally realized in the domain of multi-camera surveillance for the first time by exploiting the max-min fairness metric to formalize our surveillance objective, that is, to maximize the expected minimum observation time over all targets while guaranteeing a predefined image resolution of observing them, and (c) a structural assumption in the state transition dynamics of a surveillance environment can be exploited to improve its scalability to linear time in the number of targets to be observed during surveillance. Empirical evaluation through extensive simulations in realistic surveillance environments shows that fair-MC^2 outperforms the state-of-the-art and baseline MC^2 algorithms. We have also demonstrated the feasibility of deploying our fair-MC^2 algorithm on real AXIS 214 PTZ cameras.

1   INTRODUCTION

The problem of surveillance has emerged as a critical concern in many urban cities worldwide following a recent series of security threats like the Boston bomb blasts and Mumbai terrorist attacks. Central to the problem of surveillance is that of tracking and observing crowds of mobile targets spatiotemporally distributed over a large environment (e.g., airport terminal, railway and subway stations, bus depot, shopping mall, marketplace, school campus). It is often necessary to acquire and maintain 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. To address this surveillance problem, some recent works have proposed automated mechanisms to coordinate and control a network of static/fixed and active *pan-tilt-zoom* (PTZ) cameras to either (a) maximize the coverage of multiple mobile targets with a guaranteed pre-defined image resolution of observing them [2, 7, 8, 9] or (b) focus on one or few tar-

gets to be observed at high resolution [4, 10, 11, 12, 13]. However, all these *multi-camera coordination and control* (MC^2) algorithms share a common pitfall: In order to achieve either surveillance objective described above, they may "starve" some targets of observation by the active cameras for a prolonged period of time (e.g., see Fig. 1), especially those isolated ones with low likelihood of observing them; in the worst case, they may not be observed at all.

Figure 1.   Existing MC^2 algorithms may starve some targets of observation by any active camera.

Surprisingly, this issue of starvation has not been tackled by the multi-camera surveillance community. It motivates the need to design and develop a MC^2 algorithm that can coordinate the PTZ actions of the active cameras to observe all targets fairly. Intuitively, this implies prioritizing the coverage of targets with the least observation time such that a fair observation of all the targets is achieved if and only if increasing the observation of any target necessarily results in a decrease in observation of some other target with equal or lower observation time. In practice, this is particularly desirable for surveillance environments where it is not possible to identify potential high-importance/priority targets (e.g., suspicious targets or adversaries) in real time prior to the occurrence of a security threat (e.g., bomb blast); a fair observation of all targets consequently provides the forensic investigation in the aftermath at least some footage of each of them. Such a notion of fairness is often known as the max-min fairness in resource allocation problems (e.g., bandwidth allocation in networking).

Our active multi-camera surveillance problem can be distinguished from the conventional fairness-driven resource allocation problems by the following practical, non-trivial issues surrounding it: (a) It is subject to real-world physical constraints such as the spatial localities of the active cameras (i.e., resources) and mobile targets (i.e., users), which restrict their interactions. For example, some cameras may not be able to observe any target at times because all the targets are beyond their possible *fields of view* (FoV's). On the other hand, some targets may occasionally move into regions that are occluded from observations by the active cameras; (b) the stochastic motion of the targets entail uncertain (hence, less predictable) trajectories, which complicate how the active cameras are to be coordinated to keep possibly multiple targets of the least observation time within their FoV's at a guaranteed predefined image resolution; and

---

[1]   Department of Computer Science, National University of Singapore, Republic of Singapore, email: prabhu@nus.edu.sg, {lowkh, mohan}@comp.nus.edu.sg

(c) a MC^2 algorithm, if poorly designed, incurs exponential time in the number of targets to be observed during surveillance, thus degrading its real-time performance.

This paper is the first to present a principled decision-theoretic MC^2 algorithm called fair-MC^2 that can coordinate and control the active cameras to achieve max-min fairness in the observation of crowds of targets moving stochastically (Sections 2 and 3). The contributions of our work here are novel in demonstrating how (a) the uncertainty in the locations, directions, speeds, and observation times of the targets arising from the stochasticity of their motion can be modeled probabilistically (Section 2.3), (b) the notion of fairness in observing targets can be formally realized in the domain of multi-camera surveillance for the first time by exploiting the max-min fairness metric (Section 2.4) to formalize our surveillance objective, that is, to maximize the expected minimum observation time over all targets while guaranteeing a predefined image resolution of observing them (Section 3), and (c) a structural assumption in the state transition dynamics of a surveillance environment can be exploited to improve the scalability of our fair-MC^2 algorithm to linear time in the number of targets to be observed during surveillance (Section 3). Our fair-MC^2 algorithm is empirically evaluated in various realistic surveillance environmental setups through extensive simulations and its deployment on real AXIS 214 PTZ cameras (Section 4).

2   MODELING A SURVEILLANCE ENVIRONMENT

A surveillance environment comprises crowds of targets with stochastic motion and a network of static/fixed and active PTZ cameras calibrated to common ground plane coordinates. The wide-view static cameras can observe all the mobile targets at *low* resolution and estimate the targets' locations, directions, and speeds. These location, direction, and speed information of the targets are used by our fair-MC^2 algorithm to coordinate and control the actions of the active PTZ cameras to observe them fairly at a guaranteed predefined *high* image resolution. We will next describe how the surveillance environment is modeled by defining the states and state transition model of the m targets, the actions of the n active PTZ cameras, and the surveillance objective function.

2.1   States of Targets

Let $\mathcal{M} \triangleq \{1, 2, \ldots, m\}$ denote a set of indices of the m targets and $x_i \triangleq (x_{l_i}, x_{d_i}, x_{s_i}, x_{o_i}) \in \mathcal{X}_l \times \mathcal{X}_d \times \mathcal{X}_s \times \mathcal{X}_o$ denote a state of each target i for $i = 1, \ldots, m$ where $x_{l_i} \in \mathcal{X}_l$, $x_{d_i} \in \mathcal{X}_d$, $x_{s_i} \in \mathcal{X}_s$, and $x_{o_i} \in \mathcal{X}_o$ represent target i's location, direction, speed, and observation time, respectively. Let $\mathcal{X} \triangleq \mathcal{X}_l \times \mathcal{X}_d \times \mathcal{X}_s \times \mathcal{X}_o$ denote a set of all possible states of a target and $X_{\mathcal{M}} \triangleq (x_1, x_2, \ldots, x_m) \in \mathcal{X}^m$ denote a joint state of the m targets.

2.2   Actions of PTZ Cameras

Let a_j denote an action (specifically, a PTZ configuration) of each PTZ camera j for $j = 1, \ldots, n$ and $A \triangleq (a_1, a_2, \ldots, a_n) \in \mathcal{A}$ denote a joint action of the n PTZ cameras where $\mathcal{A}$ represents a set of all possible joint actions (i.e., PTZ configurations) of the n PTZ cameras. Let $FoV(a_j) \subset \mathcal{X}_l$ denote a subset of all possible locations of a target lying within the depth-limited FoV of camera j that is in the PTZ configuration associated with its action a_j (e.g., see Fig. 2); by adjusting its zoom to focus its FoV and limiting the depth of its FoV, images of the targets observed within its depth-limited FoV can

satisfy a guaranteed predefined resolution. Supposing the n cameras are in the PTZ configurations associated with their joint action A, their joint FoV's is denoted by $FoV(A) \triangleq \bigcup_{j=1}^{n} FoV(a_j)$.

2.3   State Transition Model of Targets

To model the uncertainty in the states (i.e., locations, directions, speeds, and observation times) of the m targets arising from the stochasticity of their motion, their joint state transition is represented by a probability $P(X'_{\mathcal{M}}|X_{\mathcal{M}}, A)$ of changing from their current joint state $X_{\mathcal{M}}$ to their next joint state $X'_{\mathcal{M}}$ given the joint action A of the n cameras. By exploiting the structural assumption that the next state x'_i of each target i is conditionally independent of the other $m-1$ targets' states given its current state x_i and cameras' joint action A for $i = 1, \ldots, m$, the joint state transition probability $P(X'_{\mathcal{M}}|X_{\mathcal{M}}, A)$ of the m targets can be factored into state transition probabilities $P(x'_i|x_i, A)$ of individual targets $i = 1, \ldots, m$:

$$P(X'_{\mathcal{M}}|X_{\mathcal{M}}, A) = \prod_{i=1}^{m} P(x'_i|x_i, A) . \tag{1}$$

Such a conditional independence assumption (1) is central to improving the scalability of our fair-MC^2 algorithm from exponential to linear time in the number m of targets, as shown later in Section 3.

The state transition of target i from x_i to x'_i includes stochastic transitions of its location from x_{l_i} to x'_{l_i}, its direction from x_{d_i} to x'_{d_i}, its speed from x_{s_i} to x'_{s_i}, and its observation time from x_{o_i} to x'_{o_i}. So, the state transition probability of target i can be factored into transition probabilities of its location, direction, speed, and observation time:

$$P(x'_i|x_i, A) = P(x'_{l_i}|x_{l_i}, x'_{d_i}, x'_{s_i})P(x'_{d_i}|x_{d_i})P(x'_{s_i}|x_{s_i}) \\ P(x'_{o_i}|x'_{l_i}, x_{o_i}, A) .$$

The transition probability $P(x'_{d_i}|x_{d_i})$ ($P(x'_{s_i}|x_{s_i})$) of target i's direction (speed) is modeled as a normal distribution $\mathcal{N}(\mu_d, \sigma_d^2)$ ($\mathcal{N}(\mu_s, \sigma_s^2)$) where the mean parameter μ_d (μ_s) is the current direction (speed) of target i estimated from the tracking module in the static cameras and the variance parameter σ_d^2 (σ_s^2) is learned from a dataset of targets' trajectories in the surveillance environment. The transition probability $P(x'_{l_i}|x_{l_i}, x'_{d_i}, x'_{s_i})$ of target i's location is constructed using a general direction-speed motion model similar to that in [14]. To define the transition probability $P(x'_{o_i}|x'_{l_i}, x_{o_i}, A)$, the observation time x_{o_i} of target i is increased to $x'_{o_i} = x_{o_i} + 1$ if it is observed in the FoV of any active camera, and its observation time remains the same otherwise, that is, $x'_{o_i} = x_{o_i}$:

$$P(x'_{o_i} = x_{o_i} + 1|x'_{l_i}, x_{o_i}, A) \triangleq \begin{cases} 1 & \text{if } x'_{l_i} \in FoV(A), \\ 0 & \text{otherwise.} \end{cases}$$

$$P(x'_{o_i} = x_{o_i}|x'_{l_i}, x_{o_i}, A) \triangleq \begin{cases} 1 & \text{if } x'_{l_i} \notin FoV(A), \\ 0 & \text{otherwise.} \end{cases}$$

2.4   Surveillance Objective Function

Supposing the state transition models of all targets are deterministic (i.e., the states of all targets in the next time step are known), the surveillance objective can be defined directly in terms of the max-min fairness metric, that is, to maximize the minimum observation time over all targets while guaranteeing a predefined image resolution of observing them. Such a surveillance objective can be achieved

by defining an objective function U that measures the minimum observation time over all targets:

$$U(X_{\mathcal{M}}, A) \triangleq \min_{i \in \mathcal{M}} x_{o_i} . \tag{2}$$

It is noteworthy to point out the usefulness of other popular fairness metrics to the domain of multi-camera surveillance such as the Jain's fairness index [3]. Jain's fairness index $(\sum_{i \in \mathcal{M}} x_{o_i})^2 / (|\mathcal{M}| \sum_{i \in \mathcal{M}} x_{o_i}^2)$, which is revised to reflect our notations, measures whether the users receive their fair share of the resources. The value of Jain's fairness index lies between 0 and 1. The observation of all targets is perfectly fair if the index is 1. It is unfair if the index is 0. Jain's fairness index is not suitable for measuring fairness in the observation of crowds of targets in active multi-camera surveillance: For example, when 999 targets are always observed by any active camera and only 1 target is not being observed at all, Jain's fairness index yields 0.999, which is close to perfect. However, the single target is not observed at all and may potentially be a suspicious target that is critical to be observed by surveillance.

3 FAIR-MC^2 ALGORITHM

In practice, the states (i.e., locations, directions, speeds, and observation times) of the targets in the next time step are uncertain due to the stochasticity of their motion. Therefore, our fair-MC^2 algorithm has to instead maximize the expected minimum observation time over all targets in the next time step:

$$A^* \triangleq \arg\max_{A \in \mathcal{A}} Q(X_{\mathcal{M}}, A) \tag{3}$$

$$Q(X_{\mathcal{M}}, A) \triangleq \sum_{X'_{\mathcal{M}} \in \mathcal{X}^m} U(X'_{\mathcal{M}}, A) P(X'_{\mathcal{M}} | X_{\mathcal{M}}, A) \tag{4}$$

where $X'_{\mathcal{M}}$ is a joint state of the m targets in the next time step and A is a joint action of the n cameras. Computing an optimal joint action A^* (3) incurs $\mathcal{O}(|\mathcal{A}||\mathcal{X}|^m)$ time that is exponential in the number m of targets. This exponential time complexity can be significantly reduced by exploiting the conditional independence assumption in the targets' state transition model (1) (Section 2.3). As a result, the value function Q (4) can be simplified to

$$Q(X_{\mathcal{M}}, A) = t_{\min} + \prod_{i \in \mathcal{Y}} \sum_{x'_i \in \mathcal{X}_A} P(x'_i | x_i, A) \tag{5}$$

where $\mathcal{Y} \subseteq \mathcal{M}$ denotes a set of indices of all targets with minimum observation time in the current time step (i.e., $\mathcal{Y} \triangleq \{i \in \mathcal{M} \mid x_{o_i} = \min_{k \in \mathcal{M}} x_{o_k}\}$), $\mathcal{X}_A$ denotes a subset of all possible states of a target whose corresponding locations lie within the joint FoV's of the n cameras that are in the PTZ configurations associated with their joint action A (i.e., $\mathcal{X}_A \triangleq \{x'_i \in \mathcal{X} \mid x'_{l_i} \in FoV(A)\}$), and $t_{\min} \triangleq \min_{k \in \mathcal{M}} x_{o_k}$ denotes a constant representing the minimum observation time over all targets in the current time step. The derivation of (5) is given in Appendix A.

By plugging (5) into (4), (3) reduces to

$$A^* = \arg\max_{A \in \mathcal{A}} \prod_{i \in \mathcal{Y}} \sum_{x'_i \in \mathcal{X}_A} P(x'_i | x_i, A) . \tag{6}$$

To interpret (6), our current fair-MC^2 algorithm simply chooses a joint action $A \in \mathcal{A}$ that maximizes the product of likelihoods of observing, in the next time step, all targets with minimum observation time in the current time step by the active cameras.

The following result indicates that an optimal joint action A^* (6) can be derived using linear time in the number m of targets to be observed during surveillance:

Theorem 1 *If (1) holds, then computing the optimal joint action A^* (6) incurs $\mathcal{O}(|\mathcal{A}||\mathcal{X}|m)$ time.*

In (6), computing the likelihood of observing a target with minimum observation time (i.e., sum of probabilities) incurs $\mathcal{O}(|\mathcal{X}|)$ time. Computing the product of $|\mathcal{Y}|$ likelihoods then incurs $\mathcal{O}(|\mathcal{X}|m)$ time since the size of $\mathcal{Y}$ can be m in the worst case and Theorem 1 follows.

As mentioned above, computing an optimal joint action A^* (6) only needs to consider all targets with minimum observation time. These targets may be beyond the FoV's of some active cameras due to their spatial localities, which is an issue stated in Section 1. Consequently, multiple possible optimal joint actions are possible because any action of such a camera is optimal. In the worst case, all targets with minimum observation time are beyond the FoV's of all active cameras (i.e., $Q(X_{\mathcal{M}}, A^*) = t_{\min}$), thus causing all the cameras to be in "limbo".

To remedy this, the key idea is to repeatedly refine the set of optimal joint actions by preserving fairness in the observation of the remaining targets using (6) after ignoring those with minimum observation time. To elaborate, the first step is to compute the set $\mathcal{A}^*$ of optimal joint actions of the active cameras satisfying (6). Then, ignore the targets with minimum observation time by removing $\mathcal{Y}$ from $\mathcal{M}$, that is, $\mathcal{M} \leftarrow \mathcal{M} \setminus \mathcal{Y}$. Finally, consider $\mathcal{A}^*$ to be the new joint action space in (6), that is, $\mathcal{A} \leftarrow \mathcal{A}^*$. These steps are repeated until there is a unique optimal joint action A^* or the number $|\mathcal{M}|$ of remaining targets after ignoring those with minimum observation time is 0 (see Algorithm 1).

---

while $|\mathcal{A}| > 1 \vee |\mathcal{M}| > 0$ **do**
  Compute $\mathcal{A}^* = \{A^*\}$ by (6)
  $\mathcal{M} \leftarrow \mathcal{M} \setminus \mathcal{Y}$
  $\mathcal{A} \leftarrow \mathcal{A}^*$
end

Algorithm 1: Fair-MC^2($X_{\mathcal{M}}, \mathcal{A}$).

4 EXPERIMENTS AND DISCUSSION

This section empirically evaluates our proposed fair-MC^2 algorithm in different realistic surveillance environmental setups through extensive simulations (Section 4.1) and its deployment on real AXIS 214 PTZ cameras (Section 4.2). Our video demo [2]. visually explains the environmental setup and interesting observations from our real camera surveillance experiments. The surveillance performance of our fair-MC^2 algorithm is compared to that of the following state-of-the-art MC^2 algorithms as well as some baseline methods:

- *Fair-MC^2 algorithm without prediction* (WoP): This algorithm is based on optimizing the objective function (2) without accounting for uncertainty in the targets' locations, directions, speeds, and observation times arising from the stochasticity of their motion;
- *Equality in observation of targets* (equal-MC^2): This MC^2 algorithm aims to achieve equal observation times of all targets by coordinating and controlling the cameras such that the expected difference between the maximum and minimum observation times of the targets is minimized;

---

[2] http://www.comp.nus.edu.sg/~lowkh/camera.html

- *Maximizing coverage of targets* (COV): The MC^2 algorithm of [8] coordinates and controls the active cameras such that the expected number of targets observed in their joint FoV's is maximized;
- *Round robin method* (RRB): For this baseline method, 50% of the targets in the surveillance environment are given high priorities at every time step in a round robin fashion. The active cameras are controlled to observe the targets based on their priorities;
- *Auto-panning of PTZ cameras* (AP): This is another baseline method in which the active cameras are panned to each of their PTZ configurations in a round robin fashion. Surprisingly, this method seems to be the current practice of controlling the PTZ cameras in the multi-camera surveillance industry; for example, see the list of cameras in [1] that are controlled to do auto-panning.

The following performance metrics are used to evaluate the tested MC^2 algorithms and baseline methods:

$$\text{Fairness} \triangleq \min_{i \in \mathcal{M}} x_{o_i}^T$$

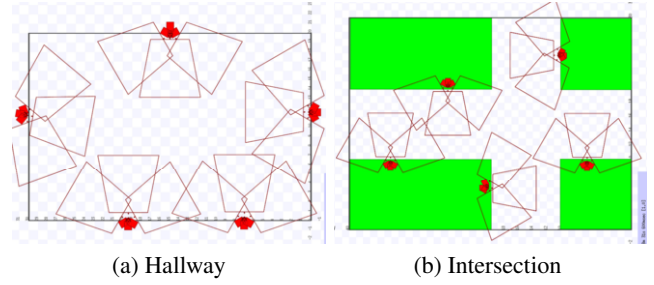$$\text{Coverage} \triangleq \frac{1}{T} \sum_{k=1}^{T} m_{obs}^k$$

where T (i.e., set to 100) is the total number of time steps taken in our experiments, $x_{o_i}^T$ is the observation time of target i at time step T, and m_{obs}^k is the total number of targets observed in the joint FoV's of the active cameras at time step k. That is, the Fairness metric measures the minimum observation time over all $m = |\mathcal{M}|$ targets in the surveillance environment at time step T and the Coverage metric measures the number of targets observed in the joint FoV's of the active cameras averaged over the T time steps.

4.1   Simulated Surveillance Environments

Our Fair-MC^2 algorithm is evaluated in two realistic simulated surveillance environments using the Player/Stage simulator: Hallway and intersection setups (Fig. 2). In both setups, a network of $n = 5$ PTZ cameras is deployed and each camera can perform 3 possible actions (i.e., PTZ configurations), as shown in Fig. 2. Each PTZ camera is configured such that images of targets observed within its depth-limited FoV satisfy a guaranteed predefined resolution. Both setups contain up to $m = 50$ mobile targets whose trajectories are generated using a general direction-speed motion model similar to that in [14] and speeds are $s = 1$ cell per time step. The simulator itself acts as a static camera that is used to observe/estimate the locations, directions, and speeds of the mobile targets at every time step. The duration of a time step is set to 5 seconds. The state transition probabilities of every target are pre-computed offline in both setups.

Figs. 3 and 4 show results of the Fairness and Coverage performance of the tested MC^2 algorithms and baseline methods in, respectively, the hallway and intersection setups deploying $n = 5$ PTZ cameras and containing $m = 5, 10, 15, \ldots, 50$ targets. Our observations from the simulated experiments are as follows:

Fairness and Coverage performance. Figs. 3a and 4a show that our fair-MC^2 algorithm outperforms equal-MC^2, COV, RRB, and AP in the Fairness metric in both setups. This is expected because fair-MC^2 directly optimizes the Fairness metric by prioritizing the coverage of targets with minimum observation time, hence resolving the issue of starvation. From Figs. 3b and 4b, it is expected that fair-MC^2 does not achieve a better Coverage performance than COV since COV directly optimizes the Coverage metric. However, fair-MC^2 can produce a Coverage performance that is generally better than that of equal-MC^2 and AP in both setups. It can also be observed from

(a) Hallway   (b) Intersection

Figure 2.   Setups of simulated surveillance environments: (a) Hallway ($|\mathcal{X}_l| = 30 \times 20$ cells) and (b) intersection ($|\mathcal{X}_l| = 40 \times 30$ cells).

Figs. 3a-b and 4a-b that the Fairness and Coverage performance of our fair-MC^2 algorithm degrade gracefully with an increasing number m of targets.

The equal-MC^2 algorithm performs less well than fair-MC^2 in both Fairness and Coverage metrics because it sometimes causes the cameras not to observe any target in order to minimize the expected difference between the maximum and minimum observation times of the targets. As a result, the targets (i.e., including those with minimum observation time) are starved of observation by the cameras. Interestingly, this observation reveals that obtaining equal observation times of all targets is not equivalent to achieving fairness in observing them.

The COV algorithm performs very poorly in the Fairness metric because, in order to maximize the expected number of observed targets, some targets are starved of observation by the cameras for a long time, as explained previously in Section 1.
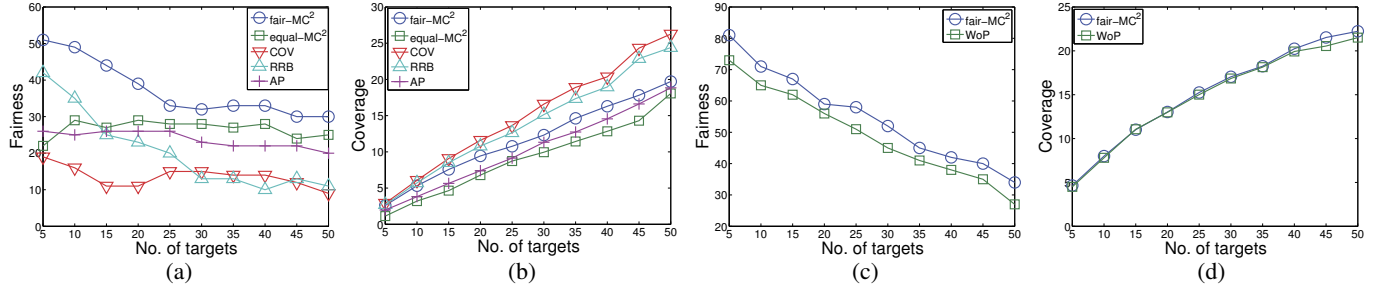
The RRB method also performs very poorly in the Fairness metric because when 50% of the targets are prioritized at every time step in a round robin fashion, the remaining unprioritized targets are starved of observation by the cameras until their turns are reached. The prioritized targets may also sometimes move beyond the FoVs of the cameras.

The AP method suffers from poor Fairness and Coverage performance because they do not exploit the targets' state information.
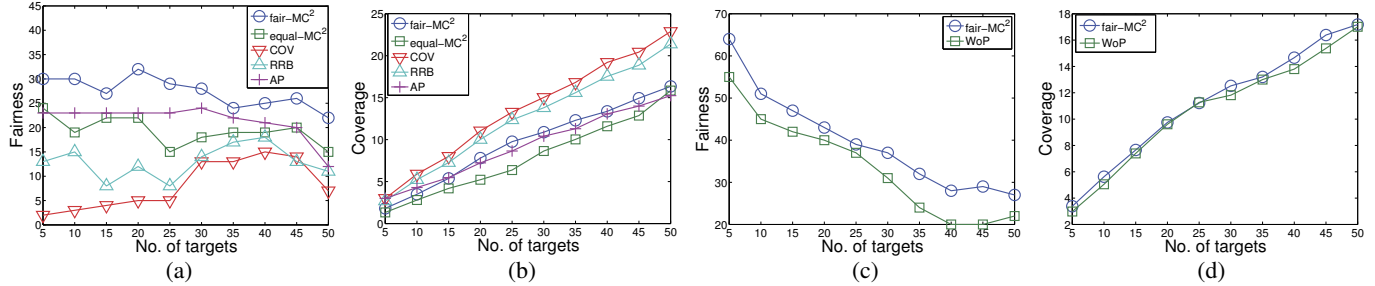
Modeling uncertainty in targets' states. To better demonstrate the importance of modeling the uncertainty in the targets' states in order to predict their stochastic motion well, we increase the targets' speeds to $s = 2$ cells per time step. Figs. 3c and 4c show that our fair-MC^2 algorithm outperforms WoP in the Fairness metric in both setups because fair-MC^2 coordinates and controls the cameras based on the predicted locations and observation times of the targets whereas WoP coordinates and controls the cameras based on the current locations and observation times of the targets since it does not account for the uncertainty in the targets' states. As a result, when the targets are at the edge of the FoV's of the active cameras, fair-MC^2 tries to keep the targets in the center of the FoV based on the predicted locations of the targets. In contrast, WoP loses the targets in the next time step, which causes the loss of observation of the targets with minimum observation time. This also explains the slightly better Coverage performance of fair-MC^2 over WoP (Figs. 3d and 4d).

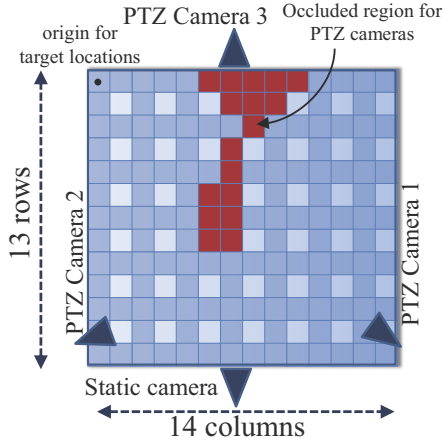## 4.2   Real Camera Surveillance Environment

The feasibility of deploying our fair-MC^2 algorithm on a network of $n = 3$ real AXIS 214 PTZ cameras is investigated using the experimental setup/testbed in our indoor lab surveillance environment, as shown in Fig. 5. Each PTZ camera can perform 3 possible actions.

Figure 3.   Graphs of (a,c) Fairness and (b,d) Coverage performance vs. number m of targets in a simulated hallway setup.

Figure 4.   Graphs of (a,c) Fairness and (b,d) Coverage performance vs. number m of targets in a simulated intersection setup.

Figure 5.   Setup of real camera surveillance environment ($|\mathcal{X}_l| = 14 \times 13$ cells): The occluded region (i.e., red cells) is deliberately introduced to cause the observation times of the targets to be more varied when they move into it.

The FoV's of the PTZ cameras are configured manually based on the resolution of the observed targets. The wide-view static camera, which is placed opposite to PTZ camera 3, can observe the entire surveillance environment at a low resolution. The static camera and 3 PTZ cameras are calibrated to common ground plane coordinates. The targets are the Lego robots that are programmed to move in a random manner. The robots are tracked in the static camera based on color properties using OpenCV libraries. Using the location, direction, and speed information of the targets observed/estimated by the static camera, the 3 PTZ cameras are coordinated and controlled by our fair-MC^2 algorithm to observe the targets fairly.

Our fair-MC^2 algorithm is evaluated using up to $m = 5$ targets in different interesting real camera surveillance experiments. For example, in the case of $m = 2$ targets, one of the targets is programmed to move into the occluded region and wait there for a few time steps. When that target re-enters the observable region, the PTZ cameras try to focus and observe it because the observation time of that target is less than the other target. In the case of $m = 5$ targets, three

targets are made to move from the occluded region to the observable region so that their observation times remain zero for a few time steps. The snapshots of the observation times of these 5 targets are shown in Fig. 6. As mentioned before, at time step 2, targets 1, 2, and 3 are inside the occluded region and their observation times are thus zero. When these targets move forward into the observable region, the cameras try to observe them and their observation times hence increase at time step 7. At time step 9, the observation times of all targets become equal and our fair-MC^2 algorithm alternates the active cameras' observation over all the targets in order to maintain a fair observation time of them. Table 1 shows the Fairness performance of our fair-MC^2 algorithm taken at time step $T = 50$ with varying number $m = 2, 3, 4, 5$ of targets in these real camera surveillance experiments.
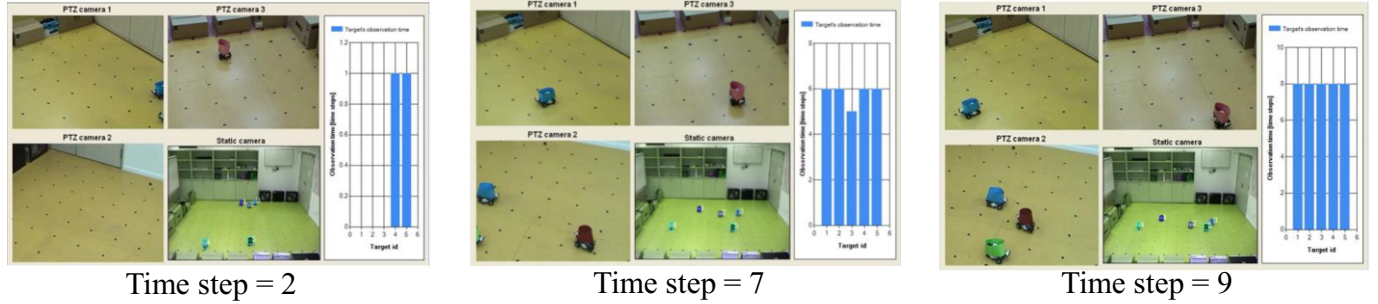
m	2	3	4	5
Fairness	36	32	28	27

Table 1.   Fairness performance of our fair-MC^2 algorithm with varying number m of targets in a real camera experimental setup.

5   CONCLUSION

This paper describes a novel decision-theoretic fair-MC^2 algorithm that can coordinate and control the active cameras to maximize the expected minimum observation time over all targets with a guaranteed predefined image resolution of observing them. As a result, the issue of starvation that plagues the existing MC^2 algorithms [2, 4, 7, 8, 9, 10, 11, 12, 13] can be resolved. Our fair-MC^2 algorithm accounts for the uncertainty in the locations, directions, speeds, and observation times of the targets arising from the stochasticity of their motion by modeling them probabilistically. Through our work in this paper, the notion of fairness in observing targets is formally realized in the domain of multi-camera surveillance for the first time. We have exploited the conditional independence assumption in the targets' state transition model to significantly reduce the exponential time complexity of fair-MC^2 to that of linear time in the number

Time step = 2   Time step = 7   Time step = 9

Figure 6. Snapshots of the observation times of $m = 5$ targets in a real camera surveillance experiment.

of targets. Empirical evaluation through simulations reveals that fair-MC^2 outperforms the state-of-the-art and baseline MC^2 algorithms. We have also implemented our fair-MC^2 algorithm on real AXIS 214 PTZ cameras to demonstrate its practical feasibility in a real surveillance system. In our future work, we like to extend our fair-MC^2 algorithm to operate in a partially observable surveillance environment where wide-view static cameras are no longer required and observations are made only through the PTZ cameras [9]. Also we would like to extend our fair-MC^2 algorithm into robotic surveillance [5, 6].

ACKNOWLEDGEMENTS

REFERENCES

[1] http://www.earthcam.com.
[2] A. Abdelkader, M. Mokhtar, and H. El-Alfy, 'Angular heuristics for coverage maximization in multi-camera surveillance', in *Proc. AVSS*, pp. 373–378, (2012).
[3] R. Jain, D.M. Chiu, and W. R. Hawe, 'A quantitative measure of fairness and discrimination for resource allocation in shared computer system', DEC Research Report TR-301, Eastern Research Laboratory, Digital Equipment Corporation, (1984).
[4] N. Krahnstoever, T. Yu, S. N. Lim, K. Patwardhan, and P. Tu, 'Collaborative Real-Time Control of Active Cameras in Large Scale Surveillance Systems', in *Proc. M2SFA2*, (2008).
[5] 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*, pp. 28–33, (2004).
[6] 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).
[7] V. P. Munishwar, S. Tilak, and N. B. Abu-Ghazaleh, 'Coverage management for mobile targets in visual sensor networks', in *Proc. MSWiM*, pp. 107–116, (2012).
[8] 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*, pp. 155–162, (2012).
[9] 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).
[10] F. Z. Qureshi and D. Terzopoulos, 'Planning ahead for PTZ camera assignment and handoff', in *Proc. ICDSC*, (2009).
[11] E. Sommerlade and I. Reid, 'Probabilistic surveillance with multiple active cameras', in *Proc. ICRA*, (2010).
[12] B. Song, C. Ding, A. T. Kamal, J. A. Farrell, and A. K. Roy-chowdhury, 'Distributed camera networks', *IEEE Signal Processing Magazine*, **28**(3), 20–31, (2011).
[13] M. T. J. Spaan and P. U. Lima, 'A decision-theoretic approach to dynamic sensor selection in camera networks', in *Proc. ICAPS*, pp. 279–304, (2009).
[14] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*, MIT Press, Cambridge, MA, 2005.

A Derivation of (5)

$$Q(X_\mathcal{M}, A)$$
$$= \sum_{X'_\mathcal{M} \in \mathcal{X}^m} U(X'_\mathcal{M}, A) P(X'_\mathcal{M}|X_\mathcal{M}, A)$$
$$= \sum_{X'_\mathcal{M} \in \mathcal{X}^m} \min_{i \in \mathcal{M}} x'_{o_i} P(X'_\mathcal{M}|X_\mathcal{M}, A)$$
$$= \sum_{X'_\mathcal{Y} \in \mathcal{X}^y} \min_{i \in \mathcal{Y}} x'_{o_i} P(X'_\mathcal{Y}|X_\mathcal{Y}, A) \sum_{X'_{\overline{\mathcal{Y}}} \in \mathcal{X}^{m-y}} P(X'_{\overline{\mathcal{Y}}}|X_{\overline{\mathcal{Y}}}, A)$$
$$= \sum_{X'_\mathcal{Y} \in \mathcal{X}^y} \min_{i \in \mathcal{Y}} x'_{o_i} P(X'_\mathcal{Y}|X_\mathcal{Y}, A)$$
$$= \sum_{X'_\mathcal{Y} \in \mathcal{X}^y \setminus \mathcal{X}^y_A} \min_{i \in \mathcal{Y}} x'_{o_i} P(X'_\mathcal{Y}|X_\mathcal{Y}, A) + \sum_{X'_\mathcal{Y} \in \mathcal{X}^y_A} \min_{i \in \mathcal{Y}} x'_{o_i} P(X'_\mathcal{Y}|X_\mathcal{Y}, A)$$
$$= \sum_{X'_\mathcal{Y} \in \mathcal{X}^y \setminus \mathcal{X}^y_A} t_{\min} P(X'_\mathcal{Y}|X_\mathcal{Y}, A) + \sum_{X'_\mathcal{Y} \in \mathcal{X}^y_A} (t_{\min} + 1) P(X'_\mathcal{Y}|X_\mathcal{Y}, A)$$
$$= \sum_{X'_\mathcal{Y} \in \mathcal{X}^y \setminus \mathcal{X}^y_A} t_{\min} P(X'_\mathcal{Y}|X_\mathcal{Y}, A) + \sum_{X'_\mathcal{Y} \in \mathcal{X}^y_A} t_{\min} P(X'_\mathcal{Y}|X_\mathcal{Y}, A)$$
$$\quad + \sum_{X'_\mathcal{Y} \in \mathcal{X}^y_A} P(X'_\mathcal{Y}|X_\mathcal{Y}, A)$$
$$= \sum_{X'_\mathcal{Y} \in \mathcal{X}^y} t_{\min} P(X'_\mathcal{Y}|X_\mathcal{Y}, A) + \sum_{X'_\mathcal{Y} \in \mathcal{X}^y_A} P(X'_\mathcal{Y}|X_\mathcal{Y}, A)$$
$$= t_{\min} \sum_{X'_\mathcal{Y} \in \mathcal{X}^y} P(X'_\mathcal{Y}|X_\mathcal{Y}, A) + \sum_{X'_\mathcal{Y} \in \mathcal{X}^y_A} P(X'_\mathcal{Y}|X_\mathcal{Y}, A)$$
$$= t_{\min} + \sum_{X'_\mathcal{Y} \in \mathcal{X}^y_A} P(X'_\mathcal{Y}|X_\mathcal{Y}, A)$$
$$= t_{\min} + \prod_{i \in \mathcal{Y}} \sum_{x'_i \in \mathcal{X}_A} P(x'_i|x_i, A) .$$

The first and second equalities are due to (4) and (2), respectively. The third equality follows by partitioning $X_\mathcal{M}$ into $X_\mathcal{Y} \triangleq (x_i)_{i \in \mathcal{Y}}$ and $X_{\overline{\mathcal{Y}}} \triangleq (x_i)_{i \in \overline{\mathcal{Y}}}$ where $\mathcal{M} = \mathcal{Y} \cup \overline{\mathcal{Y}}$ and $y = |\mathcal{Y}|$. The fourth equality follows from the law of total probability: $\sum_{X'_{\overline{\mathcal{Y}}} \in \mathcal{X}^{m-y}} P(X'_{\overline{\mathcal{Y}}}|X_{\overline{\mathcal{Y}}}, A) = 1$. The fifth equality is due to $\mathcal{X}^y = (\mathcal{X}^y \setminus \mathcal{X}^y_A) \cup \mathcal{X}^y_A$ where $\mathcal{X}^y_A$ denotes a subset of all possible joint states of the y targets in $\mathcal{Y}$ whose corresponding locations all lie within $FoV(A)$, i.e., $\mathcal{X}^y_A \triangleq \{X'_\mathcal{Y} \in \mathcal{X}^y \mid \forall x'_i \in X'_\mathcal{Y} \ x'_{l_i} \in FoV(A)\}$. To obtain the sixth equality, for a given $X'_\mathcal{Y} = (x'_i)_{i \in \mathcal{Y}}$,

$$\min_{i \in \mathcal{Y}} x'_{o_i} = \begin{cases} t_{\min} & \text{if } X'_\mathcal{Y} \in \mathcal{X}^y \setminus \mathcal{X}^y_A, \\ t_{\min} + 1 & \text{if } X'_\mathcal{Y} \in \mathcal{X}^y_A; \end{cases}$$

where $t_{\min}$ is a constant with respect to the expectations. The second last equality follows from the law of total probability: $\sum_{X'_\mathcal{Y} \in \mathcal{X}^y} P(X'_\mathcal{Y}|X_\mathcal{Y}, A) = 1$. The last equality is due to the conditional independence assumption in the y targets' state transition model (1).