Collaborative Real-Time Control of Active Cameras in Large Scale Surveillance Systems

Nils Krahnstoever, Ting Yu, Ser-Nam Lim, Kedar Patwardhan, Peter Tu

Collaborative Real-Time Control of Active Cameras in Large Scale Surveillance Systems

Nils Krahnstoever, Ting Yu, Ser-Nam Lim, Kedar Patwardhan, Peter Tu

Visualization and Computer Vision Lab, GE Global Research, Niskayuna, NY 12309

{*krahnsto,yut,limser,patwardh,tu*}*@research.ge.com*

Abstract. A system that controls a set of Pan Tilt Zoom (PTZ) cameras for acquiring close-up imagery of subjects in a surveillance site is presented. The PTZ control is based on the output of a multi-camera, multi-target tracking system operating on a set of fixed cameras, and the main goal is to acquire imagery of subjects for biometrics purposes such as face recognition, or non-facial person identification. For this purpose, this paper introduces an algorithm to address the generic problem of collaboratively controlling a limited number of PTZ cameras to capture an observed number of subjects in an optimal fashion. Optimality is achieved by maximizing the probability of successfully completing the addressed biometrics task, which is determined by an objective function parameterized on expected capture conditions such as distance at which a subject is imaged, angle of capture and several others. Such an objective function serves to effectively balance the number of captures per subject and quality of captures. Qualitative and quantitative experimental results are provided to demonstrate the performance of the system which operates in real-time under real-world conditions on four PTZ and four static CCTV cameras, all of which are processed and controlled via a single workstation.

1 Introduction

Most commercially available automatic surveillance systems today operate with fixed CCTV cameras as this allows the use of efficient detection and tracking algorithms. Unfortunately, even the resolution of high-quality CCTV cameras is limited to 720x480 with the resolution of low-end cameras being much lower. This makes tasks that are sensitive to resolution and image quality, such as face detection, face recognition, and forensic examinations, very difficult, especially if subjects are imaged from a distance. Mega-pixel sensors can potentially overcome the resolution problem, but their real-world deployment is still very limited, mostly due to challenges associated with managing, processing and recording non-standard imagery from such sensors. This has led to the extensive use of Pan Tilt Zoom (PTZ) cameras for operator-based security applications, since PTZ cameras provide an inexpensive way of obtaining close-up imageries of activities in a site. Given the autonomy of a fixed camera tracking system and the versatility of PTZ cameras, it is only natural to combine them in a master-slave configuration, where detection and tracking information obtained from the fixed cameras is used to automatically control one or more PTZ cameras. The main challenge faced by such a system arises when the number of subjects and activities exceeds the number of

PTZ cameras, in which case the scheduling and controlling of the PTZ cameras become non-trivial.

Intuitively, a good plan for scheduling and controlling PTZ cameras should meet the following requirements: (i) no subject should be neglected, (ii) no subject should receive excessive preference over other subjects, (iii) the quality of captures should be optimized, and (iv) the assignment of PTZ cameras to subject should consider their capacities to meet the quality requirement. In this paper, we present a formalized method that satisfies these criteria, with a particular interest in obtaining high-quality close-up captures of targets for performing biometrics tasks. Towards this goal, a novel objective function, based on a set of quality measures that characterize the overall probability of successfully completing these tasks, is defined. Specifically, the assignments of all PTZ cameras to targets are considered jointly, so that the optimization of the objective function in effect guides a set of PTZ cameras, collaboratively, to obtain high-quality imageries of all site-wide activities.

The rest of the paper is organized as follow. We first discuss related work in Section 2, followed by a formal definition of the problem in Section 3. We then present the objective function in Section 4, and discuss how we could optimize it in Section 5. Experimental results are given in Section 6, after which discussions and conclusions are made in Section 7.

2 Related Work

A substantial number of papers in the literature has focused on low and middle-level vision problems in the context of multi-camera surveillance systems. The main problems highlighted in these papers are object detection and tracking [1,2,3,4], and multi-target, multi-camera tracking in a site-wide manner [5,6]. The importance of accurate detection and tracking is obvious, since tracking information is needed, as a first step, towards controlling a set of PTZ cameras to acquire high-quality imageries, and towards, for example, building biometrics signatures of the tracked targets automatically. In addition to accurate tracking information, the acquisition of high-quality imageries, particularly for biometrics purposes, also requires (i) accurate calibration between the fixed and PTZ cameras for establishing correspondences [7] and computing pan, tilt and zoom settings, and (ii) an effective and efficient camera scheduling algorithm for assigning PTZ cameras to targets [8].

A camera scheduling algorithm would typically utilize tracking information, provided by one or more fixed cameras performing detection and tracking, for computing a schedule that controls the assignments of targets to PTZ cameras over time. Each PTZ camera would then servo, based on calibration data, to aim itself at different targets in a timely fashion as specified by the schedule. Such a setup, commonly known as master/slave [9], is adopted by our system. Several different flavors of master/slave control algorithm have been proposed previously, including those that are based on heuristics [9,10]. However, while these heuristics based algorithms provide a simple and tractable way of computing schedules, they quickly become non-applicable as the number of targets increases and exceeds the number of PTZ cameras, in which case the scheduling problem becomes increasingly non-trivial. To overcome the problem,

other papers, such as [11], have proposed a number of different camera scheduling algorithms designed for different application goals. They include, for example, a round-robin method that assigns cameras to targets sequentially to achieve uniform coverage.

Rather than assigning equal importance to each target, [12] instead proposed ordering the targets in a priority queue based on their arrival time and frequency with which they were captured. Currently available PTZ cameras are also ranked based on the ease of adjusting their PTZ states for an assignment, and the most suitable camera is then selected. [13] ranked the targets by the estimated deadlines by when they will leave the surveillance area. An optimal subset of the targets, which satisfies the deadline constraint, is obtained through an exhaustive search.

The scheduling problem is further complicated by crowded scenes due to occlusions caused by target interaction. [14] proposed estimating such occlusion moments based on camera geometry and predicted target motion. This is followed by constructing a visibility interval for each capture, which is defined as the complement of the occlusion moment. Based on these visibility intervals, they schedule the cameras using a greedy graph search method [15].

3 Problem Setup

Given a surveillance site and $\mathbf{N}^{\mathrm{f}}$ fixed cameras, let $\mathbf{C}_i^{\mathrm{f}}$ represent the fixed cameras, which perform tracking [3,4] and are calibrated with respect to a common site-wide metric coordinate system [7,16]. We assume that a ground-plane based tracker observes *targets* $\mathbf{O} = \{\mathbf{O}_i | i = 1, \dots\}$ during its operation. At time t, the state of a target is given by $\mathbf{O}_i^t = (x_i^t, v_i^t)$, with ground plane location $\mathbf{x} = (x, y)$ and ground plane velocity $\mathbf{v} = (v_x, v_y)$, and the system observes targets with indices $\mathbf{S}^t = \{s_i^t | i = 1, \dots, \mathbf{N}^t\}$.

We are also given $\mathbf{N}^{\mathrm{p}}$ PTZ cameras, each of which is denoted as $\mathbf{C}_i^{\mathrm{p}}$, with projective geometries $\{\mathbf{P}_i^{\mathrm{p}} | i = 1, \dots, \mathbf{N}^{\mathrm{p}}\}$. The Pan Tilt Zoom state of each PTZ camera is defined as $\mathbf{s} = (\phi, \theta, r)$, where ϕ and θ represent pan and tilt angles, and r the zoom factor. The state $(0, 0, 1)$ corresponds to the camera's *home* position, in which its projective matrix is defined. Our system has the ability to instruct a PTZ camera to focus on an object of a certain size at a particular world location, the details of which are omitted here due to size limitations.

We refer to the move and follow strategy for a PTZ camera as a *plan*. A plan for the PTZ camera $\mathbf{C}_i^{\mathrm{p}}$ is denoted as a set of tasks $\mathbf{E}_i = \left[e_i^1, e_i^2, \dots, e_i^t\right]$ where e_i^t is the index of the target that will be visited by $\mathbf{C}_i^{\mathrm{p}}$ at the tth time step. Every target in the plan will be followed for a fixed amount of time Δt^{fo}, and, if two consecutive entries in the plan belong to different targets, the PTZ camera would take some time Δt^{mv} to move from one target to another. This time is, in practice, dependent on the PTZ state of the camera when the camera starts moving and the PTZ state of the camera when it reaches the next target. A plan for *all* PTZ cameras is denoted as $\mathcal{E} = [\mathbf{E}_1, \dots, \mathbf{E}_{\mathbf{N}^{\mathrm{p}}}]$. When a PTZ camera has reached a target, it will start to capture close-up imagery. We assume that every time a PTZ camera is following a target, it obtains one capture. We denote the jth capture of target $\mathbf{O}_i$ by $\mathbf{c}_{ij}$; the frequencies and qualities of these captures significantly affect the success of a PTZ plan, which we determine using a performance objective function discussed in the following section.

4 Objective Function

The objective function used in our system is designed to effectively guide the acquisition of imageries suitable for facial biometrics tasks, but can be easily carried over to similar (possibly non-facial) tasks. Additionally, we are also interested in a *probabilistic* objective function that optimizes the success probability of the tasks.

We assume here that we have a quantitative measure, q_{ij}, for the quality of the jth capture of target i, namely $\mathbf{c}_{ij}$. We associate with $\mathbf{c}_{ij}$ a probability, p_{ij}, that our task will succeed for target i, denoted as the event $S = \{\text{the biometrics task will succeed}\}$. While the proper relationship between q_{ij} and the success probability $p_{ij} = p(S|c_{ij}) = f(q_{ij})$ is dependent on the application, our choice, specified in Section 4.2, is designed to be a joint probability of several aspects of the capture quality. Once the relationship between a single capture and its success probability has been established, we can formulate the global objective function. For the set of captures for target $\mathbf{O}_i$, namely $\mathbf{C}_i = \{\mathbf{c}_{ij}|j = 1, \ldots, \mathbf{N}_i^{\text{c}}\}$, the probability of success (i.e., not failing every single time) is

$$p(S|\mathbf{C}_i) = \begin{cases} 0 & \mathbf{N}_i^{\text{c}} = 0, \\ 1 - \prod_j (1 - p(S|c_{ij})) & \text{otherwise} \end{cases} \tag{1}$$

and the total probability of success for capturing every target is

$$p(S) = \sum_i p(S|\mathbf{C}_i)p(\mathbf{C}_i) = \frac{1}{N} \sum_i \left[1 - \prod_j (1 - p(S|c_{ij})) \right]. \tag{2}$$

This overall success probability is appealing: its value increases as the number and quality of captures increase for every target if there is no PTZ resource constraint. However, with limited number of PTZ cameras and potentially large number of targets, the number of captures that each individual target can receive is often small. The overall quality of the captures measured over all targets then serves to guide the competition for PTZ resources among the targets, where the goal is to find a camera scheduling plan that optimizes the objective function (2) given the current site and PTZ camera (which may still be executing a previously obtained plan) activities, and past captures.

It is obvious that the efficacy of our objective function, as given above, is greatly dependent on q_{ij}. Since we are interested in acquiring close-up imageries of targets' faces, so that face detection and recognition can be applied, q_{ij} should desirably encapsulate requirements for frontal and high-resolution imageries among others. We achieve this based on a set of quality measures defined as follow.

4.1 Quality Measures

View Angle: Since our interest is in facial biometrics, a camera looking at a person from behind or a top-down viewpoint brings little value. This viewpoint concern is characterized by the angle, α, between the normal of the face of a person and the line direction defined by the PTZ camera location and the face position of the person. To

measure α, instead of using the normal of the face directly, which is difficult to estimate, we use the travel direction of the person. We model this quantity as

$$q^{\mathrm{vw}}(\alpha) = \exp^{-\frac{\alpha^2}{2\sigma_{vw}^2}} . \tag{3}$$

Target-Camera Distance: The quality of captured imageries generally degrades as the distance, d_{tc}, between the target and the camera increases. Hence, our second quality measure is based on d_{tc} using a two-component model

$$q^{\mathrm{dst}}(d_{tc}) = \gamma + (1-\gamma)\exp^{-\lambda d_{tc}} . \tag{4}$$

The second term models the trend that the quality of a capture decreases as the target moves away from the camera, while the first term represents a slight baseline quality of capturing a target even if it is far away.

Tracking Performance: The combined field of views of a fixed camera network defines a region that can be monitored by the system. Theoretically, any target inside this zone can be detected and tracked although it is conceivable that a target outside the area may still have correct state estimates due to motion prediction. In reality, however, a target that is currently close to the zone boundary may have a higher probability of leaving the area soon, and therefore may have a smaller chance to be captured well. By calculating the distance between the predicted location of the target and zone boundary, d_{tz}, where d_{tz} is positive if the target is inside the zone, and negative if outside, we obtain a model, $q^{\mathrm{trck}}(d_{tz})$, consisting of a logistic function and a uniform term

$$q^{\mathrm{trck}}(d_{tz}) = \beta + (1-\beta)\frac{\exp^{\sigma_{trck} d_{tz}}}{1+\exp^{\sigma_{trck} d_{tz}}}. \tag{5}$$

Again the small uniform term represents the baseline quality of captures.

PTZ Limits: Each PTZ camera has some mechanical limitation on its Pan Tilt Zoom parameter range, defined by $(\phi_{min}, \phi_{max}, \theta_{min}, \theta_{max}, r_{min}, r_{max})$. A capture requiring a PTZ camera to set its parameter state outside this physical range is impractical. Thus, a term, $q^{\mathrm{rch}}(\phi, \theta, r)$, which defines the mechanical limitation of a PTZ camera is introduced:

$$q^{\mathrm{rch}}(\phi,\theta,r) \propto \begin{cases} 1, & (\phi,\theta,r) \in ([\phi_{min},\phi_{max}],[\theta_{min},\theta_{max}],[r_{min},r_{max}]), \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

4.2 Probability Model

Ideally, one would learn the relationship, $p_{ij} = f(q_{ij})$, between the above quality measures q_{ij} and the success probability of a specific biometrics task p_{ij} given the application domain. However, in this paper, for simplicity, we chose to determine p_{ij} directly from q_{ij}, i.e., $p_{ij} = q_{ij} = q^{\mathrm{vw}} q^{\mathrm{dst}} q^{\mathrm{trck}} q^{\mathrm{rch}}$.

In addition, targets that have a long lifetime in the site will get captured multiple times, and the system spends less and less time on old targets in favor of new arrivals. This is a desirable system behavior. In practice, however, this leads to problems because (i) human operators may feel that the system starts to neglect targets, and (ii) if

the tracker accidentally switches target index, the system might actually neglect new or very recent arrivals. To avoid this, we add a temporal decay factor to the objective function (2) that serves to discount old captures of a target, which in turn prompts the objective function to consider re-capturing the target. We augment the success probability evaluated at time t as follows

$$p_{ij}(t) = p(S|c_{ij}, t) = q_{ij}\tau(t - t_{ij}, \sigma_\tau, \Delta t_s, \Delta t_{cut}), \tag{7}$$

where t_{ij} is the time at which capture c_{ij} was taken and $\tau()$ is defined as

$$\tau(t - t_{ij}, \sigma_\tau, \Delta t_{cut}, \Delta t_s) = \begin{cases} 1 & t - t_{ij} < \Delta t_s, \\ e^{-\frac{t - t_{ij} - \Delta t_s}{\sigma_\tau}} & \Delta t_s \leq t - t_{ij} < \Delta t_{cut}, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Within time Δt_s, captures are rated fully, while after Δt_{cut}, they do not contribute to the objective function. In between, the quality factor decays with a rate determined by σ_τ.

5 Optimization

5.1 Asynchronous Optimization

Due to the cost of determining PTZ plans, the plan optimization has to be performed asynchronously to the real-time tracking and PTZ control system. Our policy is that when we begin to optimize a new plan given the most recent states of the targets, say at time t_0, we predict the time, $t_0 + \Delta t_c$, when the computation will finish, and let each PTZ camera finish up the task it is executing at that time (see Figure 1 for illustration). This creates a predictable start state for the optimization and provides a simple approach for merging old and new PTZ plans.

During plan computation, the future locations of observed targets are predicted based on the expected elapsed time for moving the PTZ cameras and following the targets. The duration of a plan is given by the maximum completion time of all tasks in the plan. As the duration of the plan increases, state predictions become less reliable as targets start to deviate from their constant velocity path. Due to this reason and in order to limit the computational complexity for computing a plan, only plans up to a certain completion time are explored. We call this maximum completion time the *planning horizon*, and define it to be the time when the first previously scheduled task completes, t_1, plus an offset Δt_h, which is chosen such that at least a certain number of tasks can be assigned to every PTZ camera (see Figure 1).

5.2 Combinatorial Search

Finding a plan that fits into the plan horizon while maximizing the objective function (2) is a combinatorial problem. During optimization, we iteratively add target-to-camera assignments to the plan, which implicitly defines a (directed acyclic) weighted graph, where the weights represent changes in quality caused by the additional assignments,

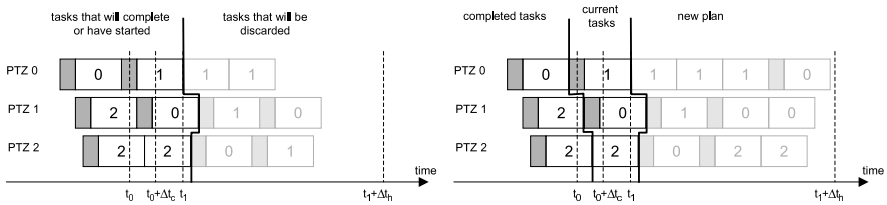

Fig. 1. Replacement of PTZ Plans. *Left: At time t_0 the system begins to compute a new PTZ plan, and the optimization is estimated to complete at time $t_0 + \Delta t_c$. All tasks that are being executed at that time will be allowed to finish, which is what the optimizer will take as a starting state. All later tasks will be discarded. Right: The discarded plans are replaced with an updated plan. The line $t_0 + \Delta t_h$ denotes the plan horizon. Gray boxes denote intervals during which a camera moves from one target to another. White boxes with numbers denote the index of the target that is scheduled to be captured.*

and nodes describe the feasible plans obtained so far. It is easy to see in Eq. 2 that the plan quality is non-decreasing in the number of target-to-plan assignments. Any plan that can not be expanded further without violating the plan horizon is a terminal node and a candidate solution. We can utilize standard graph search algorithms, such as best-first search, to find these candidate solutions. Finding the optimal solution requires exhaustively examining this expanding graph, which is possible if the number of observed targets is small. Solving larger problems through approaches such as A* or branch-and-bound searches requires an admissible optimistic estimator that can hypothesize the quality of a non-terminal node during graph expansion. Such a heuristic, however, has so far not been found. We currently resolve to a best-first strategy, which rapidly yields good candidate solutions, followed by coordinate ascend optimization through assignment changes. It rapidly improves candidate solutions by making small changes in the assignments in the currently found plan. The search is continued until a preset computation time is exceeded, and the algorithm returns the best solution found so far. See algorithm 1 for an overview.

At every optimization step and during PTZ control, if no target is currently observed, the system could go into idle mode and remain in its home position. We could however choose to be more "pro-active" by aiming the PTZ cameras at locations in the site where targets frequently appear. These locations are obtained by continuously clustering target arrival locations during operation [17]. Then, when no active target is currently present, the system is provided with virtual target locations corresponding to the cluster centers. This essentially leads to the system adaptively assuming a *ready* position whereby the PTZ cameras are aimed at typical target arrival locations. Since virtual targets are treated just like other target, the system automatically optimizes the virtual-target-to-camera assignments.

Data : Current time t_k. Predicted algorithm completion time Δt_c. Current PTZ activities up to time $t_k + \Delta t_c$. Plan horizon $t_h := t_k + \Delta t_h$. All captures c^o_{ij} that have been (or will have been made) after all current PTZ activities complete.

Result : Optimal plan $\mathcal{E}^* = [\mathbf{E}_1, \ldots, \mathbf{E}_{\mathbf{N}^{\mathrm{P}}}]$ such that objective (2) is maximized.

begin

- Set $\mathcal{E}^* = [[], \ldots, []]$.
- Compute the partial failure probability with all past captures $F_i = \prod_j (1 - p(S|c^o_{ij}, t_h))$.
- Set $p(S|\mathcal{E}^*) = \frac{1}{N} \sum_i [1 - F_i]$.
- Create priority queue Q^{open} and insert $(\mathcal{E}^*, p(S|\mathcal{E}^*))$.
- Create empty set L^{closed}.
- **while** Q^{open} *not empty and run-time is less than* Δt_c **do**
  - Retrieve best node $\mathcal{E} = [\mathbf{E}_1, \ldots, \mathbf{E}_{\mathbf{N}^{\mathrm{P}}}]$ from Q^{open} and add it to L^{closed}.
  - **if** $\mathcal{E}$ *is a terminal node and* $p(S|\mathcal{E}) > p(S|\mathcal{E}^*)$ **then**
    - Locally refine plan $\mathcal{E}$ by switching tasks labels until no further increase in quality possible.
    - Set $\mathcal{E}^* := \mathcal{E}$
    - Continue.
  - **for** *all PTZ cameras* $\mathbf{C}^{\mathrm{p}}_i$ *and all targets* s^k_l, $l = 1, \ldots, \mathbf{N}^t$ *that are active at time* t_k **do**
    - **if** *adding* s^k_l *to plan* $\mathbf{E}_i$ *exceeds time limit* t_h **then**
      - Continue
    - Create new plan $\mathcal{E}'$ from $\mathcal{E}$ by adding target s^k_l to $\mathbf{E}_i$.
    - **if** *not* $\mathcal{E}' \in L^{\mathrm{closed}}$ *and not* $(\mathcal{E}', p) \in Q^{\mathrm{open}}$ *for any* p **then**
      - Predict the parameters for this new capture c_l and compute quality q_l.
      - Assume that this plan leads overall to captures $\{c_{xy}\}$ for targets indexed by x and the number of captures (of target x) indexed by y.
      - Compute $p(S|\mathcal{E}') = \frac{1}{N} \sum_x \left[1 - F_x \prod_y (1 - p(S|c_{xy}, t_h)\right]$.
      - Insert $(\mathcal{E}', p(S|\mathcal{E}'))$ into Q^{open}.

end

The plan $\mathcal{E}^*$ is our best solution.

Algorithm 1: Estimating best PTZ plan. *A best first graph search is used to obtain the best solution within computational time window* Δt_c.

6 Experiments

The system was integrated on a dual CPU (Intel Quad Core) workstation running in real-time at around 15 FPS under load and tested in a corporate campus environment. It captures a total of eight views, four from fixed cameras and four from PTZ cameras. The fixed views are processed at a resolution of 320×240 while no image processing was performed on the PTZ views. The layout of the testbed and the cameras can be seen in Figure 5 with example views shown in Figure 4.

We first present the performance of the quality measures described in Section 4.1. Figure 2 shows the quality measures (left) and PTZ parameters (right) for a target that was walking along a straight line in positive X direction, passing underneath PTZ camera 2 (see Figure 5). One can see that, as the target approaches the camera, the quality due to the decreasing camera-target distance improves, while the quality due to the view angle (increasingly down-looking) decreases. Furthermore, the physical angle limitation of the camera and a degradation in target trackability, lead to a steep drop in quality, as the target passes under the camera. The optimal capture region is at about $10\mathrm{m}$ in front of the camera.

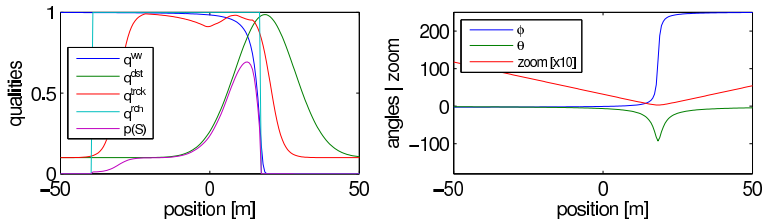

Fig. 2. Quality Objectives and PTZ Angles.

Figure 3 shows several PTZ snapshots of a 7 minute long video sequence. One can see from the figure that the system (i) effectively captures all targets in the field of view, and (ii) prefers to capture targets that are frontal to the views. The system attempts to capture frontal views of targets whenever it can, but will resolve to capturing targets from the side or behind, when no other option exists. As an example of the behavior of the system, in one instance shown in Figure 3 (bottom, left), the system was mostly focusing on the target in white shirt, being the only one facing any of the cameras. As the target turns around and walks away from the camera, the system changes its attention on two other targets that are now facing the cameras (bottom, right).

For another illustration on how the system schedules PTZ cameras intelligently, we examine a situation where two targets are being tracked by the cameras and a third target enters the site from the left (left of Figure 5). The system has very limited time to capture a frontal view and quickly re-targets PTZ 4 to follow the new arrival. Figure 4 (left) shows the success probabilities for all three targets. The third target (id=16) that was captured from a side angle has a success probability of only $P(S|id = 16) = 0.18$.

Fig. 3. Control of PTZ Cameras. *Snapshots of the PTZ views, four views per snapshot. Image order is left to right, top to bottom (four PTZ views per snapshot).*

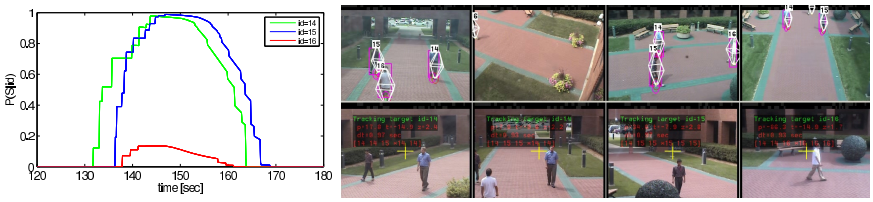

Fig. 4. Limited Time to Capture Target. *The left graph shows the target success probabilities. The right image shows the three targets moving in the site. See text for details.*

We now evaluate the numerical performance of the system. Figure 5 shows one of the four fixed cameras tracking several targets. The targets had just been acquired and each was captured several times. Upon execution of the plan optimization, the exhaustive search examined a total of 42592 plan candidates using 8.40 seconds. The best plan was determined to be $\mathcal{E}^* = [[24, 29], [28, 26, 26], [28, 28], [28, 28]]$, and was chosen because target $id = 28$ had just been recently acquired by the tracker. The total plan probability was $p(S|\mathcal{E}^*) = 0.247747$. The best-first search terminal node was found after visiting 100 nodes, which yielded the plan $[[29, 29], [26, 26, 26], [28, 28], [28, 28]]$ with plan probability $p = 0.235926$. The above optimal plan $\mathcal{E}^*$ was found after three steps of local plan refinements.

Table 1 shows, for several plan optimization runs, how the best-first refined estimate compares to the global estimate and when they were reached (measured by the number of nodes that have been explored). It shows that (i) the quality of the best-first estimate is very close to the globally optimal solution found through exhaustive search, (ii) the best-first node is found early during the optimization process, and (iii) very often, the global optimum is reached after only a fraction of all possible plans. However, as shown by the extreme case in the first row of Table 1, the exhaustive search sometimes has to explore 50% or more of all nodes. Thus, for practical purposes, if the plan horizon is not too large, the best-first refined node is indeed a very good solution to the control problem outlined in this work.

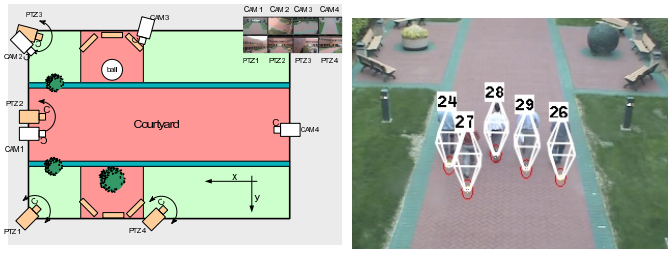

Fig. 5. Multi-Camera Surveillance System. *Left: The experimentation setup. Right: The employed tracker performs multi-camera multi-target tracking in a calibrated camera environment. The above images show five individuals being tracked by the system. The system performs centralized tracking, i.e., the identities of targets are maintained across camera views.*

total nodes	best first found at	best first probability	best node found at	best node probability
234256	81	0.148558	127866	0.161289
38016	72	0.191554	6407	0.201699
96668	112	0.267768	112	0.267768
36864	81	0.231526	8159	0.234387

Table 1. Best-First vs. Best Node. *The table shows the probabilities of the best-first search result vs. the global optimum. It also shows when the best-first node is reached and when the best node was found during the search. The left most column indicates the total number of nodes examined.*

7 Discussion and Conclusion

We have presented a systematic approach to achieving real-time collaborative control of multiple PTZ cameras. The main novelty of our approach lies in optimizing the captures of targets using a probabilistic objective function that encapsulates a set of quality measures, which includes capture distance, view angle, target reachability and trackability. While the reported experiments demonstrate the effectiveness and intelligence of the system in obtaining high-quality imageries of all site-wide activities, we are interested in conducting more detailed quantitative performance evaluation in the future. This is challenging, since any active camera system experiment has to be run in real-time. We plan to overcome such challenges by validating imageries captured online with biometrics tasks such as face detection or recognition that can be conducted offline.

Acknowledgement

This project was supported by grant #2007-RG-CX-K015 awarded by the National Institute of Justice, Office of Justice Programs, US Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the Department of Justice.

References

1. Stauffer, C., Grimson, W.: Learning patterns of activity using real-time tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2000) 747–757
2. Zhao, T., Nevatia, R.: Tracking multiple humans in complex situations. IEEE Transactions on Pattern Analysis and Machine Intelligence **26** (2004) 1208– 1221
3. Krahnstoever, N., Tu, P., Sebastian, T., Perera, A., Collins, R.: Multi-view detection and tracking of travelers and luggage in mass transit environments. In: In Proc. Ninth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS). (2006)
4. Tu, P., Wheeler, F., Krahnstoever, N., Sebastian, T., Rittscher, J., Liu, X., Perera, A., Doretto, G.: Surveillance video analytics for large camera networks. (2007)
5. Krumm, J., Harris, S., Meyers, B., Brumitt, B., Hale, M., Shafer, S.: Multi-camera multi-person tracking for easyliving. In: IEEE Workshop on Visual Surveillance. (2000)
6. Collins, R., Lipton, A., Kanade, T.: A system for video surveillance and monitoring. In: Proceedings of the American Nuclear Society (ANS) Eighth International Topical Meetingon Robotics and Remote Systems. (1999)
7. Krahnstoever, N., Mendonça, P.: Bayesian autocalibration for surveillance. In: Proc. of IEEE International Conference on Computer Vision (ICCV'05), Beijing, China. (2005)
8. Tarabanis, K.A., Allen, P.K., Tsai, R.Y.: A survey of sensor planning in computer vision. IEEE Transactions on Robotics and Automation **11** (1995) 86–104
9. Marchesotti, L., Marcenaro, L., Regazzoni, C.: Dual camera system for face detection in unconstrained environments. In: IEEE Conf. Image Processing. Volume 1. (2003) 681–684
10. Zhou, X., Collins, R., Kanade, T., Metes, P.: A master-slave system to acquire biometric imagery of humans at a distance. In: ACM SIGMM Workshop on Video Surveillance. (2003) 113–120
11. Hampapur, A., Pankanti, S., Senior, A., Tian, Y.L., Brown, L.B.: Face cataloger: Multi-scale imaging for relating identity to location. In: Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, Washington D.C. (2003) 13–21
12. Qureshi, F., Terzopoulos, D.: Surveillance camera scheduling: A virtual vision approach. ACM Multimedia Systems Journal, Special Issue on Multimedia Surveillance Systems **12** (2006) 269–283
13. Bimbo, A.D., Pernici, F.: Towards on-line saccade planning for high-resolution image sensing. Pattern Recognition Letters **27** (2006) 1826–1834
14. Lim, S.N., Davis, L.S., Mittal, A.: Constructing task visibility intervals for a surveillance system. ACM Multimedia Systems Journal, Special Issue on Multimedia Surveillance Systems **12** (2006)
15. Lim, S.N., Davis, L., Mittal, A.: Task scheduling in large camera network. In: Proceedings of the Asian Conf. on Computer Vision, Tokyo, Japan. (2007)
16. Krahnstoever, N., Mendonça, P.: Autocalibration from tracks of walking people. In: Proc. British Machine Vision Conference (BMVC), Edinburgh, UK, 4-7 September. (2006)
17. Stauffer, C.: Estimating tracking sources and sinks. In: Proceedings of the Second IEEE Workshop on Event Mining. (2003)