

# Planning Ahead for PTZ Camera Assignment and Handoff

Faisal Z. Qureshi  
Faculty of Science  
University of Ontario Institute of Technology  
Oshawa, ON L1H 7K4, Canada  
<http://faculty.uoit.ca/qureshi>

Demetri Terzopoulos  
Computer Science Department  
University of California, Los Angeles  
Los Angeles, CA 90095, USA  
<http://www.cs.ucla.edu/~dt>

**Abstract**—We present a visual sensor network, comprising wide field-of-view (FOV) passive cameras and pan/tilt/zoom (PTZ) active cameras, which automatically captures high quality surveillance video of selected pedestrians during their prolonged presence in an area of interest. A wide-FOV static camera can track multiple pedestrians, while any PTZ active camera can follow a single pedestrian at a time. The proactive control of multiple PTZ cameras is required to record seamless, high quality video of a roaming individual despite the observational constraints of the different cameras. We formulate PTZ camera assignment and handoff as a planning problem whose solution achieves optimal camera assignment with respect to predefined observational goals.

## I. INTRODUCTION

Automated human surveillance systems comprising fixed CCTV cameras can detect and track multiple objects. These systems, however, perform poorly on tasks that require higher resolution images, such as closeup facial images for biometric identification. Pan/tilt/zoom (PTZ) cameras provide an effective way of capturing high-quality video of relevant activities in the scene. This has led to surveillance systems that combine traditional passive CCTV cameras with active PTZ cameras in master/slave configurations. Tracking information collected by the fixed CCTV cameras is used to actively control the PTZ cameras in order to capture closeup video of selected subjects.

When the number of pedestrians and activities in the scene exceeds the number of available PTZ cameras, the automatic management of these cameras becomes a challenge. Ideally, the cameras in an autonomous multicamera network should decide among themselves how best to allocate their time observing different pedestrians in order to satisfy the observational goals. The dynamic nature of the scenario further complicates the problem of assigning cameras to observe different pedestrians.

In this paper, we formulate PTZ camera assignment and handoff as a *planning* problem whose solution achieves optimal camera utilization with respect to predefined observational goals. We develop a planning framework for controlling active PTZ cameras in order to acquire seamless closeup video of pedestrians of interest as they move through a designated region, entering and exiting the observational ranges of different cameras. In general, no single camera is able to achieve this goal, and multiple cameras must work together to ensure that

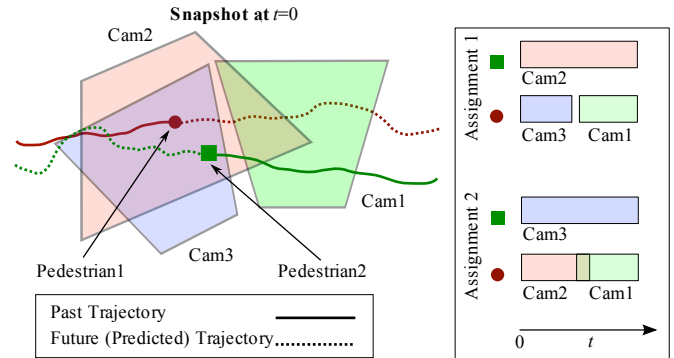


Fig. 1. The need for planning in camera assignment and handoff

at least one of them is zoomed in and actively tracking each of the selected pedestrians. The camera network must be able to resolve conflicts that might arise when tasking multiple cameras to observe several pedestrians simultaneously. Additionally, we argue that a successful camera assignment planner should consider both the short-term and long-term effects of camera assignments while searching for the best camera control strategy to carry out an observational task.

Consider, for example, the scenario shown in Fig. 1. A control strategy that does not *reason* about the long-term effects of camera assignment might task Cam3 to observe the red pedestrian (shown as a circle) and task Cam2 to observe the green pedestrian (shown as a square). While this assignment may satisfy the immediate goals, it creates a problem as the green pedestrian continues moving to the right while the red pedestrian continues moving to the left. It is impossible to capture seamless closeup video of the red pedestrian as he moves from Cam3 to Cam1 since there is no overlap between the two cameras. On the other hand, a camera control strategy that reasons about the long-term effects of camera assignment should assign Cam2 to the red pedestrian and Cam3 to the green pedestrian, assuming that both pedestrians will continue moving in their current directions. This allows for a seamless handoff between Cam2 and Cam3.

A planning strategy for controlling active PTZ cameras to capture closeup video of selected pedestrians during their

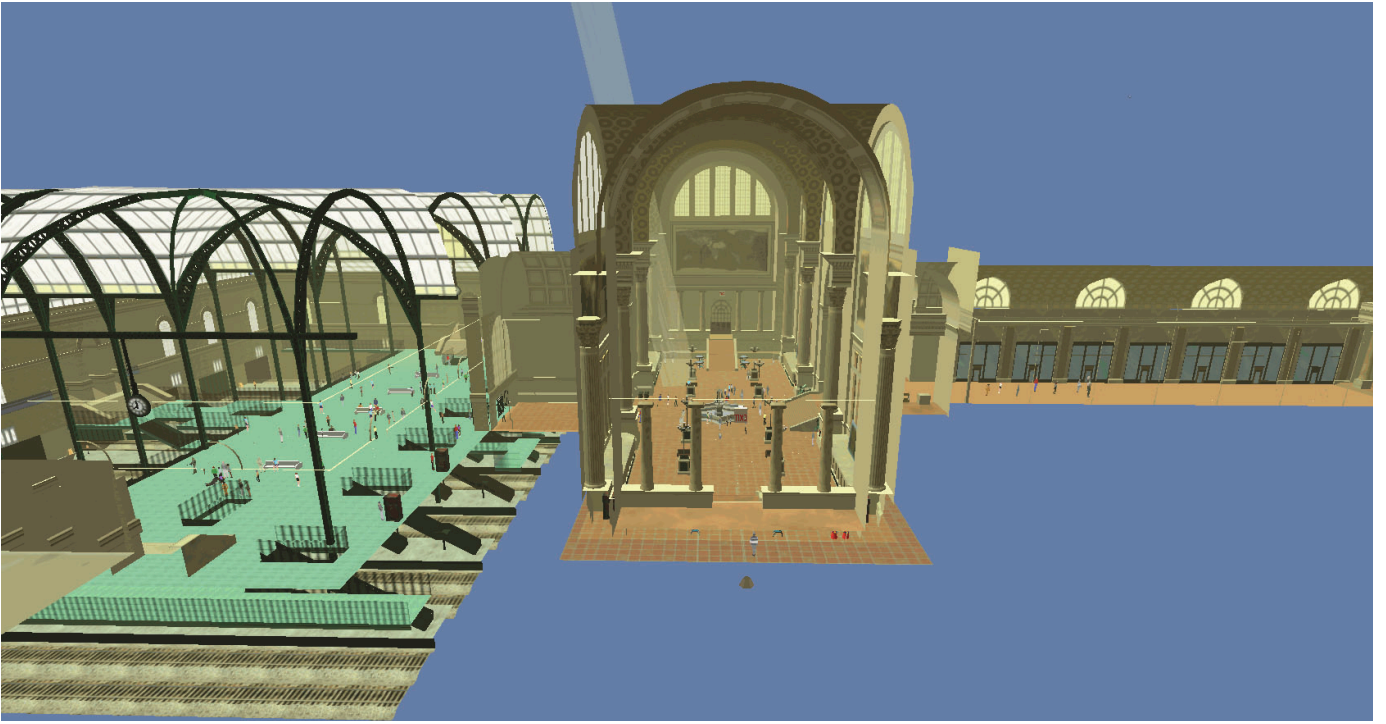


Fig. 2. A cutaway side view of the virtual train station. The main waiting room is at the center, the arcade at the right, and the concourses at the left with stairs leading down to the train platforms and tracks. The train station is populated by autonomous, self-animating virtual pedestrians.

presence in a designated area should meet the following criteria:

- 1) the quality of the captured video should be optimized,
- 2) the cameras should anticipate and successfully perform handoffs as a subject exits the range of one camera and enters the range of another camera, and
- 3) the constructed plan should be able to resolve any conflicts that may arise during the multicamera/multiperson assignment.

#### A. Virtual Vision

In view of the obstacles to deploying and experimenting with large-scale physical camera systems in the real world, we work within the *Virtual Vision* paradigm [1]–[3], implementing our surveillance system in a virtual reality simulator. The visual sensor network comprises passive and active simulated video cameras that provide perceptive coverage of an extensive virtual indoor public space—a train station populated by virtual pedestrians that autonomously engage in various natural activities (Fig. 2). The (OpenGL) graphics pipeline on a commodity PC workstation renders the busy urban scene with considerable geometric and photometric detail. Virtual surveillance cameras are situated throughout the expansive chambers of the train station. They generate multiple synthetic video feeds that emulate those generated by real surveillance cameras monitoring public spaces. The Virtual Vision simulation paradigm offers important advantages with regard to developing advanced, high-level camera control algorithms, as we do in this paper. For a detailed discussion of the

advantages, see [1], [3].

## II. RELATED WORK

Work on multicamera systems has advanced along two fronts. Several authors (e.g., [4]–[10]) have studied multicamera issues related to low-level sensing, distributed inference, and tracking. More recently, however, the research community is paying increasing attention to the problem of controlling or scheduling active cameras in order to capture high-resolution imagery of interesting events in order to support biometric analysis and to increase the situational awareness of the surveillance system operators. In a typical setup, information gathered by stationary wide-FOV cameras is used to control one or more active cameras [3], [11]–[15]. Generally speaking, the cameras are assumed to be calibrated and the total coverage of the cameras is restricted to the FOV of the stationary camera. Nearly all PTZ scheduling schemes rely on site-wide multitarget, multicamera tracking.

The critical problems of camera assignment and handoff have been studied in the context of smart camera networks. Park *et al.* [16] construct a distributed lookup table to perform camera handoffs. The lookup table encodes the suitability of a camera to observe a specific location. Jo and Han [17] propose the use of a handoff function for continuous tracking across multiple cameras. The handoff function is defined as the ratio of co-occurrence to occurrence for point pairs in two views. Their approach does not require calibration or 3D scene information. Li and Bhanu [18] developed a game theoretic approach to achieve camera handoffs. When a target

is visible in multiple cameras, the best camera is selected based upon its expected utility. They also propose a number of criteria to construct the utility function, such as the number of pixels occupied by the selected target in an image. Their approach eschews spatial and geometric information. Kim and Kim [19] develop a probabilistic framework for selecting the “dominant” camera for observing a pedestrian, defined as the camera with the highest proximity probability, which is computed as the ratio of the foreground blocks occupied by the selected pedestrian and the angular distance between the camera and that pedestrian. Song *et al.* [20] present a game-theoretic strategy for cooperative control of a set of decentralized cameras. The cameras work together to track every target in the area at acceptable image resolutions. The camera network can also be tasked to record higher resolution imagery of a selected target.

Our work presented in this paper differs from prior work in an important way. With the notable exception of [15], existing schemes for camera assignment do not *reason* about the long-term effects of camera assignments and handoffs. In essence, existing schemes are purely reactive. By contrast, the strategy introduced in this paper is *proactive* and deliberative. When searching for the best current camera assignment, it considers the future effects of possible camera assignments. The ability to reason enables our system to avoid camera assignments that might appear optimal at present, but will eventually lead to observation and tracking failures. For a good overview of planning and search techniques, we refer the reader to [21].

### III. PROBLEM STATEMENT

Consider a camera network comprising  $N_p$  calibrated wide FOV passive cameras and  $N_a$  PTZ active cameras. The static cameras track and estimate the 3D positions and velocities of the observed pedestrians. Let  $\mathcal{H} = \{h_j | j = 1, 2, \dots\}$  denote the set of pedestrians observed during the operation of the camera network. At time instant  $t$ , the state of the pedestrians observed by the camera network is given by  $(\mathbf{x}_i^t, \mathbf{v}_i^t)$ , where  $\mathbf{x}_i$  and  $\mathbf{v}_i$  represent the ground plane position and velocity, respectively, of observed pedestrian  $i$ . Let  $C = \{c_i | i \in [1, N_a]\}$  denote the set of active PTZ cameras. Each PTZ camera is described by a tuple  $\langle \mathbf{o}, \alpha_{\min}, \alpha_{\max}, \beta_{\min}, \beta_{\max} \rangle$ , where we assume that the 3D position  $\mathbf{o}$  of each PTZ camera is known *a priori*, and where  $[\alpha_{\min}, \alpha_{\max}]$  and  $[\beta_{\min}, \beta_{\max}]$  represent pan and tilt limits, respectively, for each PTZ camera. Furthermore, we assume that each PTZ camera stores a map between the gaze direction parameters  $(\alpha, \beta)$  and 3D world locations. In [3] we describe how such a map can be automatically learned by observing pedestrians present in the scene. Thus, given the 3D location of the pedestrian, a PTZ camera is able to direct its gaze towards the pedestrian. We model each PTZ camera as an autonomous agent—complete with low-level pedestrian tracking routines and fixate, zoom, and search behaviors—that is capable of recording closeup video of a designated pedestrian without relying on continuous feedback from static cameras.

With the above assumptions, we formulate collaborative camera control as a planning problem. Intuitively, planning for multicamera control is similar to *multiagent* planning problems. However, we cast the problem of controlling multiple cameras as a *multibody* planning problem. Multiagent planning typically assumes distributed planning. Multibody planning on the other hand assumes a central planner that controls the actions of multiple physical agents. We intentionally avoid distributed planning due to the enormous communication requirements. Additionally, in the case of distributed planning, each planning agent needs a decomposition of the overall task. It is not obvious how to decompose the task of capturing closeup video of the selected pedestrians into mutually independent subtasks. We therefore resort to a centralized planning strategy. Centralized planning is not suitable for large networks of PTZ cameras due to *timeliness* concerns. A good compromise is to restrict the planning to the group of “relevant” cameras. We use a similar strategy for conflict resolution in camera assignment tasks [1].

A planning problem is characterized by states, actions, and goals, and the solution to the problem requires finding a sequence of actions that will take an agent from its current state to a goal state. We begin by formally defining the states, actions, and goals for our system:

**Definition 1 (State)** Tuple  $s^t$  represents the state of the system during time interval  $[t, t + 1)$ .  $s^t = \langle s_i^t | i = 1, \dots, N_a \rangle$ , where  $s_i^t$  denotes the status of the PTZ camera  $i$  at time  $t$ . Possible values for  $s_i^t$  are *Free*( $c_i$ ), *Acquiring*( $c_i, h_j$ ), or *Recording*( $c_i, h_j$ ), for  $h_j \in \mathcal{H}$ , and for  $c_i \in C$ .

**Definition 2 (Actions)** Each PTZ camera has a repertoire of four actions: *Acquire*, *Record*, *Continue*, and *Idle*. Table I lists these actions, along with their preconditions and effects.  $a_i^t$  denotes the action for PTZ camera  $c_i$  at time  $t$ . Fig. 3 illustrates the relationship between actions, states, and low-level behaviors. The *Continue* action instructs a PTZ camera to continue its current behavior.

**Definition 3 (Joint Action)** At any given instant, each PTZ camera is executing exactly one action (possibly a *Continue*). The concurrent set of actions across the different cameras is called a *joint action*. The tuple  $a^t = \langle a_i^t | i = 1, \dots, N_a \rangle$  represents the joint action of  $N_a$  PTZ cameras at time  $t$ .

**Definition 4 (Action Sequence)** Let  $\mathcal{A} = \{a^t | t = 0, 1, \dots\}$  denote an action sequence. Note that the elements of an action sequence are joint actions.

**Definition 5 (State Sequence)** Let  $\mathcal{S} = \{s^t | t = 0, 1, \dots\}$  denote a state sequence. Sequence  $\mathcal{S}$  is obtained by starting in some initial state  $s^0$  and applying an action sequence  $\mathcal{A}$ . We express the quality of the sequence  $\mathcal{S}$  as  $Q(\mathcal{S})$ , which we define in the next section.

**Definition 6 (Goal)** The goal of the system is to capture closeup video of selected pedestrians during their presence in a designated area. Our choice of goal leads to the notion of *admissible* state sequences. An admissible state

TABLE I  
ACTION SCHEMA FOR PTZ CAMERAS ( $c_i, i \in [1, n]$ )

Actions	Preconditions	Effects	Description
$Continue(c_i)$	none	none	Do nothing
$Idle(c_i)$	none	$s_i = Free(c_i)$	Stop recording
$Acquire(c_i, h_j)$	$s_i \neq Acquiring(c_i, h_j) \wedge s_i \neq Recording(c_i, h_j)$	$s_i = Acquiring(c_i, h_j)$	Start recording pedestrian $h_j$
$Record(c_i, h_j)$	$s_i = Acquiring(c_i, h_j)$	$s_i = Recording(c_i, h_j)$	Keep recording pedestrian $h_j$

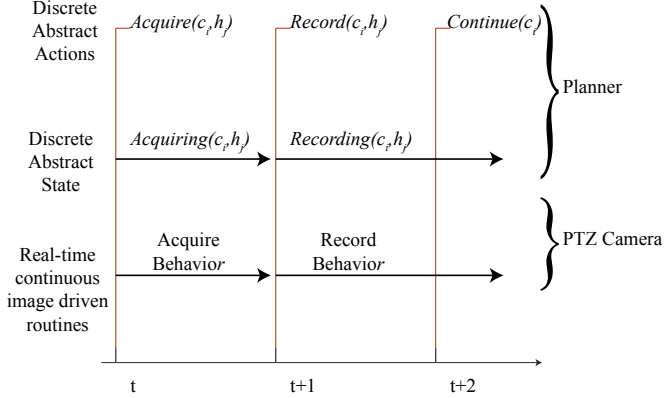


Fig. 3. Relationship between actions, states, and low-level camera behaviors.

sequence satisfies the observational constraints. Consider, for example, the goal of observing pedestrians  $h \subset \mathcal{H}$  during the time interval  $[t_s, t_e]$ . Then,  $\mathcal{S}_a = \{s^t | t = t_s, \dots, t_e\}$  represents an admissible state sequence if  $(\forall t \in [t_s, t_e])$   
 $(\exists i \in [1, N_a])$

$$s_i = Acquiring(c_i, h_j) \vee Recording(c_i, h_j),$$

where  $h_j \in h$  and  $c_i \in \mathcal{C}$ . Clearly, our notion of an admissible state sequence needs to be revised to cope with situations where the FOVs of the cameras do not overlap.

Given the above definitions, we can represent the solution to the planning problem as

$$\mathcal{S}^* = \operatorname{argmax}_{\mathcal{S} \in \mathcal{S}_a} \mathcal{Q}(\mathcal{S}),$$

and the corresponding action sequence is  $\mathcal{A}^*$ .

#### IV. COMPUTING STATE SEQUENCE QUALITY

The overall performance of the camera network is intimately tied to how capable the individual PTZ cameras are at carrying out the observation tasks assigned to them. In order for the planner to find the plan with the highest probability of success, we must quantify the quality of a state sequence (or a plan) in terms of the expected performance of individual PTZ cameras. We construct a probabilistic objective function that describes the quality of a state sequence in terms of the success probabilities of the individual PTZ cameras. Such an objective function then enables the planner to compute the plan that has the highest probability of achieving the goals.

##### A. Computing PTZ Camera Relevance

We begin by formulating the relevance  $r(c_i, O)$  of a PTZ camera  $c_i$  to an observation task  $O$ . The relevance encodes our expectation of how successful a PTZ camera will be at satisfying a particular observation task, i.e.,

$$p(c_i | O) = r(c_i, O),$$

where  $p(c_i | O)$  denotes the success probability of a camera  $c_i$  given task  $O$ .<sup>1</sup>

The computation of the relevance of a camera to the task of recording closeup videos of selected pedestrians encodes the intuitive observations that 1) the turn and zoom limits of cameras should be taken into account when assigning a camera to observe a pedestrian; i.e., a camera that has more leeway to turn and zoom may be able to follow a pedestrian for a longer period of time, 2) tracking becomes harder as camera-to-pedestrian distance grows; e.g., due to occlusions, 3) it is more desirable to select a camera that can capture facial imagery of a pedestrian, and 4) it is better to avoid unnecessary reassignment of cameras to observe different pedestrians, as doing so may degrade the performance of the underlying computer vision routines.

We formalize the above intuitions by describing the relevance of a camera to the task of observing a pedestrian in terms of the following five factors:

- **Camera-pedestrian distance**  $r_d$ : gives preference to cameras that are closer to the pedestrian.
- **Frontal viewing direction**  $r_\gamma$ : gives preference to cameras having a frontal view of the pedestrian.
- **PTZ limits**  $r_{\alpha\beta\theta}$ : takes into account the turn and zoom limits of the PTZ camera.
- **Observational range**  $r_o$ : reflects the observational constraints of a camera. It is set to 0 when the pedestrian is outside the observational range of a camera; otherwise, it is set to 1.
- **Handoff success probability**  $r_h$ : gives preference to handoff candidates in the vicinity of the camera currently observing the pedestrian. The idea is that nearby cameras have a similar viewpoint, making appearance-based pedestrian signature more relevant for the candidate camera. Factor  $r_h$  is considered only during camera handoffs; otherwise, it is set to 1. A consequence of using this factor during camera relevance computation is that

<sup>1</sup>Ideally,  $p(c_i | O) = \mathcal{F}(r(c_i, O))$ , where function  $\mathcal{F}$  should be learned over multiple trials. Krahnstoeber *et al.* [15] arrive at a similar conclusion.



the planner prefers plans with fewer handoffs, which is desirable.

These terms are defined below: Let  $r(c_i, h_j)$  represent the relevance of a camera  $c_i$  to the task of recording closeup video of a pedestrian  $h_j$ , then

$$r(c_i, h_j) = \begin{cases} 1 & \text{if } c_i \text{ is idle;} \\ r_d r_\gamma r_{\alpha\beta\theta} r_o r_h & \text{otherwise} \end{cases}$$

and

$$r_d = \exp\left(-\frac{(d - \hat{d})^2}{2\sigma_d^2}\right),$$

$$r_\gamma = \exp\left(-\frac{\gamma^2}{2\sigma_\gamma^2}\right),$$

$$r_{\alpha\beta\theta} = \exp\left(-\frac{(\theta - \hat{\theta})^2}{2\sigma_\theta^2} - \frac{(\alpha - \hat{\alpha})^2}{2\sigma_\alpha^2} - \frac{(\beta - \hat{\beta})^2}{2\sigma_\beta^2}\right),$$

$$r_o = \begin{cases} 1 & \text{if } \alpha \in [\alpha_{\min}, \alpha_{\max}] \\ & \text{and } \beta \in [\beta_{\min}, \beta_{\max}] \\ & \text{and } d < d_{\max}; \\ 0 & \text{otherwise,} \end{cases}$$

$$r_h = \exp\left(-\frac{\varepsilon^2}{2\sigma_\varepsilon^2}\right),$$

where  $\alpha$  and  $\beta$  are, respectively, the pan and tilt gaze angles corresponding to the 3D location of the pedestrian as computed by the triangulation process and  $\theta$  corresponds to the field-of-view (zoom) setting required to capture closeup video of the pedestrian.  $\hat{\theta} = (\theta_{\min} + \theta_{\max})/2$ ,  $\hat{\alpha} = (\alpha_{\min} + \alpha_{\max})/2$ , and  $\hat{\beta} = (\beta_{\min} + \beta_{\max})/2$ , where  $\theta_{\min}$  and  $\theta_{\max}$  are extremal field-of-view settings,  $\alpha_{\min}$  and  $\alpha_{\max}$  are extremal vertical rotation pan angles, and  $\beta_{\min}$  and  $\beta_{\max}$  are extremal horizontal rotation tilt angles.  $d$  denotes the camera-to-pedestrian distance,  $\hat{d}$  is the optimal camera-to-pedestrian distance, and  $d_{\max}$  is the maximum distance at which a camera can reliably track a pedestrian.  $\gamma$  is the angle between the fixation vector of the camera and the velocity vector of the pedestrian and  $\varepsilon$  represents the angle between the fixation vector of camera  $c_i$  and the fixation vector of the camera currently observing the pedestrian (Fig. 4). The fixation vector (for a camera with respect to a pedestrian) is defined along the line joining the center of projection of the camera and the 3D position of the pedestrian. The values of the variances  $\sigma_d$ ,  $\sigma_\gamma$ ,  $\sigma_\theta$ ,  $\sigma_\alpha$ ,  $\sigma_\beta$ , and  $\sigma_\varepsilon$  associated with each attribute are chosen empirically; in our experiments, we set  $\sigma_d = 10$ ,  $\sigma_\gamma = \sigma_\theta = \sigma_\alpha = \sigma_\beta = 15.0$ , and  $\sigma_\varepsilon = 45.0$ .

### B. State Sequence Quality

The quality of a state sequence is

$$\mathcal{Q}(S) = \prod_{t \in [0, 1, \dots]} \mathcal{Q}(s^t), \quad (1)$$

where the quality of a state  $s^t$  is determined by the success probabilities of individual PTZ cameras. Omitting superscript  $t$  for clarity,

$$\mathcal{Q}(s) = \prod_{i \in [1, N_a]} p(s_i) = \prod_{i \in [1, N_a]} r(c_i, h_j). \quad (2)$$

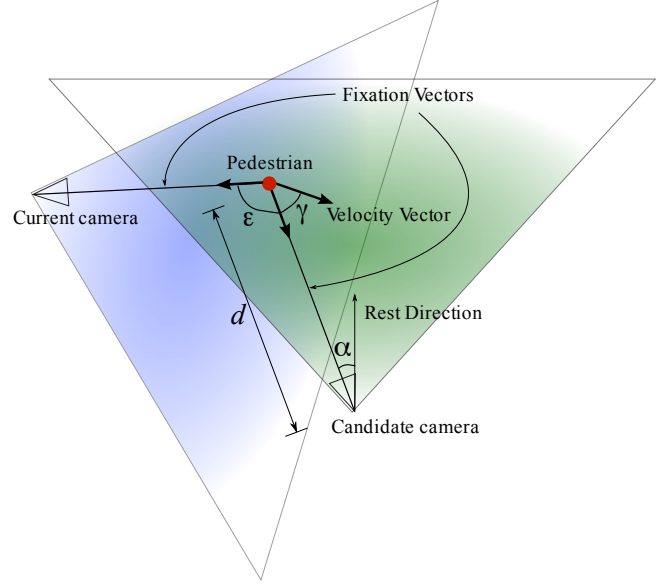


Fig. 4. Relevance of camera to the task of observing a pedestrian.

Rewriting (1), we obtain

$$\mathcal{Q}(S) = \prod_{t \in [0, 1, \dots]} \left( \prod_{i \in [1, N_a]} r(c_i, h_j) \right). \quad (3)$$

Thus,  $\mathcal{Q}(S)$  represents the probability of success of a state sequence  $S$  and it serves as a probabilistic objective function that enables the planner to compute state sequences (or plans) with the highest probability of success.

## V. PLANNING

Finding an optimal state sequence is a combinatorial search problem, which typically cannot be carried out in real time. This is especially true for longer plans that arise in scenarios involving multiple pedestrians and larger networks of PTZ cameras. Camera control, however, must be carried out in real time. Therefore, planning activity must proceed in parallel with real-time camera control. In our case, the planning activity requires reliable predictions of the states (position and velocity) of pedestrians. Pedestrian state predictions are provided by the static cameras. Obviously, the predictions become increasingly unreliable as the duration of the plan increases. It is therefore counterproductive to construct long-duration plans.

We regard plans of length 10 or more as being long-duration plans. The real-world time duration of a plan depends upon its length and the duration of each of its steps (in real-world time). We construct short-duration plans consisting of action/state sequences of lengths between 5 and 10. When a new plan is available, actions are sent to the relevant PTZ cameras. Fig. 5 outlines our planning strategy.

**Require:**  $\mathcal{A}_c$  {Current action sequence.}

**Require:**  $\mathcal{S}_c$  {Current state sequence. Initially,  $\mathcal{S}_c$  consists of a single state showing all PTZ cameras as idle.}

**Require:**  $t$  {Current time.}

**Require:**  $t_h$  {End time of the current plan.}

**Require:**  $t_b$  {Time budget available for planning.}

**Require:**  $t_p$  {Duration of the new plan.}

**Require:**  $\Delta t$  {Time step specifying the temporal granularity of the new plan.}

**Require:**  $e$  {Error flag indicating a re-planning request from a PTZ camera.}

```

1: while Keep planning do
2:   while  $t + t_b < t_h$  and  $e = \text{false}$  do
3:     Update  $e$  {Check for any re-planning requests from PTZ cameras.}
4:   end while
5:   if  $e = \text{false}$  then
6:      $s_0 \leftarrow \text{EndState}(\mathcal{S}_c)$  {Last element of the current state sequence.}
7:   else
8:      $s_0 \leftarrow \text{CurrentState}(\mathcal{S}_c)$  {Re-planning starts from the current state.}
9:   end if
10:   $(\mathcal{A}^*, \mathcal{S}^*) = \text{Plan}(s_0, t_b, \Delta t)$  {Find optimal action/state sequence starting from state  $s_0$ . Planning stops when the maximum plan depth  $t_p/\Delta t$  is reached or when the time budget for planning is exhausted.}
11:  if  $e = \text{true}$  then
12:     $\text{Replace}(\mathcal{S}_c, \mathcal{S}^*)$  {Replace current state sequence.}
13:     $\text{Replace}(\mathcal{A}_c, \mathcal{A}^*)$  {Replace current action sequence.}
14:  else
15:     $\text{Append}(\mathcal{S}_c, \mathcal{S}^*)$  {Append the new state sequence to the current state sequence.}
16:     $\text{Append}(\mathcal{A}_c, \mathcal{A}^*)$  {Append the new action sequence to the current action sequence.}
17:  end if
18:   $e \leftarrow \text{false}$  {Reset error flag; essentially ignoring any errors that might have been raised by the PTZ cameras during the current planning cycle.}
19:   $t_h \leftarrow t + t_p$  {Update end time (i.e., time horizon).}
20:  Send the new actions to the relevant PTZ cameras.
21: end while

```

Fig. 5. The planner.

#### A. Finding an Optimal Sequence

We employ *greedy best-first search* [21] to find the optimal sequence of actions/states. The starting state along with the successor function, which enumerates all possible camera-pedestrian assignments, induce a state graph with branching factor  $3^{h_s} N_a! / (N_a - h_s)!$ , where  $h_s$  is the number of pedestrians selected to be observed by at least one PTZ camera. Fortunately, the branching factor is much smaller in practice due to observational constraints of PTZ cameras and due to the preconditions imposed on camera actions (Table I). Eq. (2)

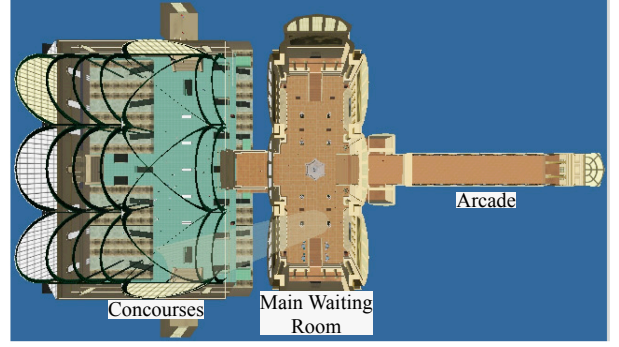


Fig. 6. Overhead view of the train station.

provides the state value, whereas the path value is given by (3). To keep the search problem tractable, we compute shallow plans (plans comprising 5 to 10 steps). The time granularity  $\Delta t$  may be used to control the actual duration of a plan without affecting the associated search problem. For additional information on greedy best-first search, see [21].

## VI. RESULTS

Our visual sensor network is deployed and tested within our Virtual Vision train station simulator. The simulator incorporates a large-scale environmental model (of the original Pennsylvania Station in New York City) with a sophisticated pedestrian animation system that combines behavioral, perceptual, and cognitive human simulation algorithms [22]. The simulator can efficiently synthesize well over 1000 self-animating pedestrians performing a rich variety of activities in the large-scale indoor urban environment. Like real humans, the synthetic pedestrians are fully autonomous. They perceive the virtual environment around them, analyze environmental situations, make decisions and behave naturally within the train station. They can enter the station, avoiding collisions when proceeding through portals and congested areas, queue in lines as necessary, purchase train tickets at the ticket booths in the main waiting room, sit on benches when they are tired, purchase food/drinks from vending machines when they are hungry/thirsty, etc., and eventually proceed downstairs in the concourse area to the train tracks. Standard computer graphics techniques enable a photorealistic rendering of the busy urban scene with considerable geometric and photometric detail (Fig. 2).

Fig. 7 shows a scenario consisting of 3 PTZ cameras that are tasked to record closeup video of a single pedestrian as he makes his way through the shopping arcade towards the concourses in the train station. The camera network successfully accomplishes this goal. Initially, only Cam 1 (shown as a blue triangle) is observing the pedestrian. Our planner anticipates that the pedestrian will soon enter the range of Cam 2, and Cam 2 is pre-tasked with observing the pedestrian, which results in a successful handoff between Cam 1 and Cam2. As stated earlier, the planner constructs short-duration plans, so Cam 3 is not considered at this time. During the next planning cycle, however, Cam 3 is also taken into account

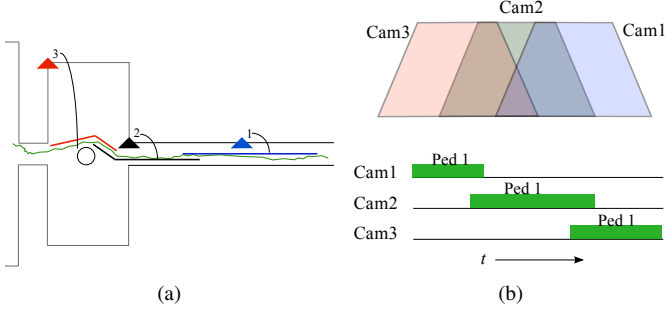


Fig. 7. Cameras 1, 2, and 3 perform handoffs to capture closeup video of the selected pedestrian. Outline (a) depicts the boundary walls of the train station shown in Fig. 6.

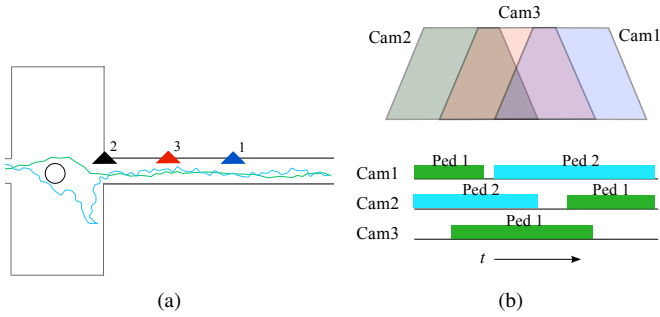


Fig. 8. Cameras 1, 2, and 3 successfully record closeup video of two selected pedestrians. Outline (a) depicts the boundary walls of the train station shown in Fig. 6.

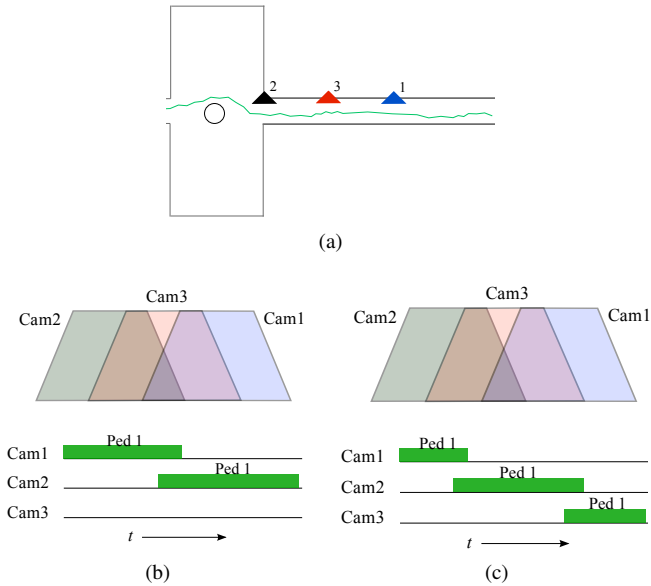


Fig. 9. Cameras 1, 2, and 3 successfully record closeup video of two selected pedestrians. Outline (a) depicts the boundary walls of the train station shown in Fig. 6. The planner selects the strategy in (b), as it requires fewer handoffs.

TABLE II  
SUCCESS RATES FOR THE CAMERA NETWORK SHOWN IN FIG. 10.

# of Selected Pedestrians	Short Duration Plans	Long Duration Plans
	5 Steps	10 Steps
1	99.8%	96%
2	95.1%	88%
4	67.2%	65.1%

as the pedestrian continues to walk towards the main waiting room. Cam 3 and Cam 2 perform a successful handoff.

Fig. 8 depicts a far more challenging scenario, where 3 cameras are tasked to record closeup videos of two selected pedestrians. The first pedestrian (green trajectory) has entered the arcade and is moving towards the concourses, and the second pedestrian (cyan trajectory) has entered the main waiting room and, after purchasing a ticket at one of the ticket booths, is walking towards the arcade. Here, Cam 3 temporarily takes over Pedestrian 1, thereby allowing Cam 1 to handoff Pedestrian 2 to Cam 2. Afterwards, Pedestrian 1 is handed off to Cam 1.

In Fig. 9, three PTZ cameras are tasked with recording closeup video of a single pedestrian (green trajectory). Notice how the fields of view of all three cameras overlap; consequently, there is more than one correct handoff strategy, as shown in Fig. 9(b)–(c). The planner selects the handoff strategy in Fig. 9(b), as it requires fewer handoffs.

Table II documents the success rates of capturing closeup videos of up to 4 pedestrians using a camera network comprising 7 PTZ cameras (shown in Fig. 10) plus static wide-FOV cameras. The results are aggregated over 5 runs each. As expected, when the network is tasked with closely observing a single pedestrian, the success rate is close to 100%; however, prediction errors prevent the success rate from being exactly 100%. When the network is tasked with simultaneously observing two pedestrians, the success rate falls to 95.1% for short-duration plans and it is below 90% for long-duration plans. Again, we can attribute this behavior to errors in predicting the state of the selected pedestrians. Next, the camera network is tasked to observe 4 pedestrians simultaneously. The success rate now falls to 67% for short-duration plans and 65% for long-duration plans. This is partly due to the fact that the planner cannot find an admissible state sequence when the four pedestrians aggregate in the arcade.

## VII. CONCLUSIONS AND FUTURE WORK

We have presented a planning strategy for intelligently managing a network of active PTZ cameras so as to satisfy the challenging task of capturing, without human assistance, closeup biometric videos of selected pedestrians during their prolonged presence in an extensive environment under surveillance. The ability to plan ahead enables our surveillance system to avoid camera assignments that might appear optimal at present, but will later lead to observation failures. The planning process assumes the reliable prediction of pedestrian states, which is currently provided by the supporting stationary

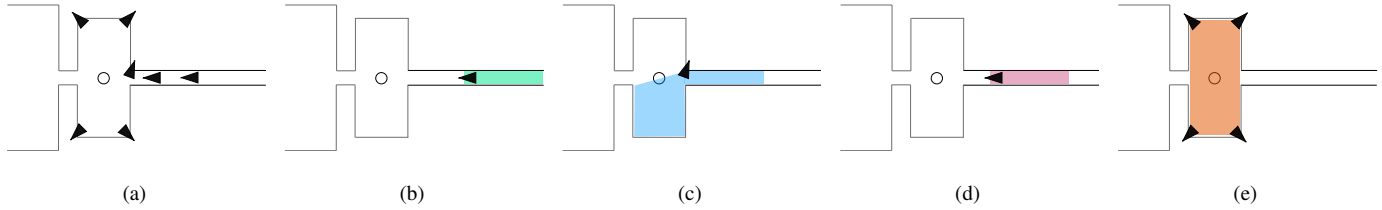


Fig. 10. A virtual camera network deployed in our Virtual Vision simulator. (a) The positions of the virtual cameras. (b)–(d) The observational ranges of individual cameras. (e) The observational range of the four cameras situated at the corners of the main waiting room; cameras cannot observe/track any pedestrian outside their observational ranges.

wide-FOV passive cameras. We have noticed that short duration plans are preferable to longer duration plans as 1) state predictions are less reliable for longer plans, and 2) longer plans take substantially longer to compute, which adversely affects the relevance of a plan when it is executed.

Since scalability is an issue when dealing with numerous active cameras spread over an extensive area, we hope to tackle the scalability issue in the future by investigating distributed multiagent planning strategies. In the shorter term, we will address the scalability issue by restricting planning activity to the relevant cameras by first grouping cameras with respect to the active tasks. Our strategy assumes a static camera setup; it currently does not support *ad hoc* camera deployment, a limitation that we intend to address in the future. We have prototyped our surveillance system in a virtual train station environment populated by autonomous, lifelike pedestrians. However, we intend to evaluate our planning strategy using a physical camera network, which will involve significant technical challenges.

#### ACKNOWLEDGEMENTS

The work reported herein was supported in part by a UOIT Startup Grant and an NSERC Discovery Grant. We thank Wei Shao and Mauricio Plaza-Villegas for their invaluable contributions to the implementation of the Penn Station simulator.

#### REFERENCES

- [1] F. Z. Qureshi and D. Terzopoulos, "Smart camera networks in virtual reality," *Proceedings of the IEEE (Special Issue on Smart Cameras)*, vol. 96, no. 10, pp. 1640–1656, October 2008.
- [2] F. Z. Qureshi and D. Terzopoulos, "Virtual vision and smart cameras," in *Proc. of the First ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC07)*, Vienna, Austria, Sep. 2007, pp. 87–94.
- [3] F. Z. Qureshi and D. Terzopoulos, "Surveillance camera scheduling: A virtual vision approach," *ACM Multimedia Systems Journal*, vol. 12, pp. 269–283, Dec 2006.
- [4] R. Collins, O. Amidi, and T. Kanade, "An active camera system for acquiring multi-view video," in *Proc. International Conference on Image Processing*, Rochester, NY, Sep. 2002, pp. 517–520.
- [5] D. Comaniciu, F. Berton, and V. Ramesh, "Adaptive resolution system for distributed surveillance," *Real Time Imaging*, vol. 8, no. 5, pp. 427–437, Oct 2002.
- [6] M. Trivedi, K. Huang, and I. Mikic, "Intelligent environments and active camera networks," in *Proc. IEEE International Conference on Systems, Man and Cybernetics*, vol. 2, Nashville, TN, Oct 2000, pp. 804–809.
- [7] O. Javed, Z. Rasheed, O. Alatas, and M. Shah, "KNIGHT: A real time surveillance system for multiple and non-overlapping cameras," in *Proc. International Conference on Multimedia and Expo (ICME03)*, vol. 1, Baltimore, MD, July 2003, pp. 649–652.
- [8] R. Farrell and L. S. Davis, "Decentralized discovery of camera network topology," in *Proc. of the Second International Conference on Distributed Smart Cameras (ICDSC08)*, Menlo Park, CA, September 2008.
- [9] P. Meijer, C. Leistner, and A. Martinieri, "Multiple view camera calibration for localization," in *Proc. of the First IEEE/ACM International Conference on Distributed Smart Cameras (ICDSC07)*, Vienna, Austria, September 2007, pp. 228–234.
- [10] K. Heath and L. Guibas, "Multi-person tracking from sparse 3d trajectories in a camera sensor network," in *Proc. of the Second International Conference on Distributed Smart Cameras (ICDSC08)*, Menlo Park, CA, September 2008.
- [11] R. Collins, A. Lipton, H. Fujiyoshi, and T. Kanade, "Algorithms for cooperative multisensor surveillance," *Proceedings of the IEEE*, vol. 89, no. 10, pp. 1456–1477, Oct. 2001.
- [12] X. Zhou, R. T. Collins, T. Kanade, and P. Metes, "A master-slave system to acquire biometric imagery of humans at distance," in *Proc. ACM SIGMM International Workshop on Video Surveillance*. New York, NY, 2003, pp. 113–120.
- [13] C. J. Costello, C. P. Diehl, A. Banerjee, and H. Fisher, "Scheduling an active camera to observe people," in *Proc. ACM International Workshop on Video Surveillance and Sensor Networks*. New York, NY, 2004, pp. 39–45.
- [14] A. Hampapur, S. Pankanti, A. Senior, Y.-L. Tian, L. Brown, and R. Bolle, "Face cataloger: Multi-scale imaging for relating identity to location," in *Proc. IEEE Conference on Advanced Video and Signal Based Surveillance*, Washington, DC, 2003, pp. 13–21.
- [15] N. Krahnstoever, T. Yu, S. Lim, K. Patwardhan, and P. Tu, "Collaborative real-time control of active cameras in large-scale surveillance systems," in *Proc. ECCV Workshop on Multi-camera and Multi-modal Sensor Fusion*, Marseille, France, October 2008, pp. 1–12.
- [16] J. Park, P. C. Bhat, and A. C. Kak, "A look-up table based approach for solving the camera selection problem in large camera networks," in *Working Notes of the International Workshop on Distributed Smart Cameras (DSC 2006)*, B. Rinner and W. Wolf, Eds., Boulder, CO, Oct 2006, pp. 72–76.
- [17] Y. Jo and J. Han, "A new approach to camera hand-off without camera calibration for the general scene with non-planar ground," in *Proc. of the 4th ACM international workshop on Video surveillance and sensor networks (VSSN06)*. Santa Barbara, CA, October 2006, pp. 195–202.
- [18] Y. Li and B. Bhanu, "Utility-based dynamic camera assignment and hand-off in a video network," in *Proc. of the Second International Conference on Distributed Smart Cameras (ICDSC08)*, Menlo Park, CA, September 2008, pp. 1–9.
- [19] J. Kim and D. Kim, "Probabilistic camera hand-off for visual surveillance," in *Proc. Second ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC08)*, Stanford, CA, September 2008, pp. 1–8.
- [20] B. Song, C. Soto, A. K. Roy-Chowdhury, and J. A. Farrell, "Decentralized camera network control using game theory," in *Proc. of the Second IEEE/ACM International Conference on Distributed Smart Cameras (ICDSC08)*, Menlo Park, CA, September 2008.
- [21] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 2nd ed. Prentice Hall, 2003.
- [22] W. Shao and D. Terzopoulos, "Autonomous pedestrians," *Graphical Models*, vol. 69, no. 5–6, pp. 246–274, Sept./Nov. 2007.