Multimedia Systems (2006) 12:269–283
DOI 10.1007/s00530-006-0059-4

REGULAR PAPER

Surveillance camera scheduling: a virtual vision approach

Faisal Z. Qureshi · **Demetri Terzopoulos**

Published online: 8 November 2006

Abstract We present a surveillance system, comprising wide field-of-view (FOV) passive cameras and pan/tilt/zoom (PTZ) active cameras, which automatically captures high-resolution videos of pedestrians as they move through a designated area. A wide-FOV static camera can track multiple pedestrians, while any PTZ active camera can capture high-quality videos of one pedestrian at a time. We formulate the multi-camera control strategy as an online scheduling problem and propose a solution that combines the information gathered by the wide-FOV cameras with weighted round-robin scheduling to guide the available PTZ cameras, such that each pedestrian is observed by at least one PTZ camera while in the designated area. A centerpiece of our work is the development and testing of experimental surveillance systems within a visually and behaviorally realistic virtual environment simulator. The simulator is valuable as our research would be more or less infeasible in the real world given the impediments to deploying and experimenting with appropriately complex camera sensor networks in large public spaces. In particular, we demonstrate our surveillance system in a virtual train station environment populated by autonomous, lifelike virtual pedestrians, wherein easily reconfigurable virtual cameras generate synthetic video feeds. The video streams emulate those generated by real surveillance cameras monitoring richly populated public spaces.

A preliminary version of this paper appeared as [1].

F. Z. Qureshi (✉) · D. Terzopoulos
Department of Computer Science, University of Toronto, Toronto, ON M5S 3G4, Canada
e-mail: faisal@cs.toronto.edu

D. Terzopoulos
Computer Science Department, University of California, Los Angeles, CA 90095, USA
e-mail: dt@cs.toronto.edu

1 Introduction

We regard the design of an autonomous visual sensor network as a problem in resource allocation and scheduling, where the sensors are treated as resources necessary to complete the required sensing tasks. Imagine a situation where the camera network must acquire high-resolution videos of every pedestrian that passes through a region of interest. The captured video should be amenable to further biometric analysis; e.g., by gait, gesture, or facial recognition routines. Passive cameras alone cannot satisfy this task. Additional, active cameras, also known as pan/tilt/zoom (PTZ) cameras, are needed to capture high-quality videos of pedestrians. As there will often be more pedestrians in the scene than the number of available cameras, the PTZ cameras must intelligently allocate their time among the different pedestrians. A resource management strategy can enable the cameras to decide autonomously how best to allocate their time to observing the various pedestrians in the scene. The dynamic nature of the sensing task further complicates the decision making process; e.g., the amount of time a pedestrian spends in the designated area can vary dramatically among different pedestrians, or an attempted video recording by a PTZ camera might fail due to occlusion.

1.1 The virtual vision paradigm

Beyond the legal obstacles to monitoring people in public spaces for experimental purposes, the cost of deploying and repeatedly reconfiguring a large-scale camera network in the real world can easily be prohibitive for computer vision researchers. As was argued in [2], however, rapidly evolving computer graphics and virtual reality technologies present viable alternatives to the real world for developing computer vision systems. Legal impediments and cost considerations aside, the use of a virtual environment can also offer greater flexibility during the system design and evaluation process. Terzopoulos [3] proposed a virtual vision approach to designing surveillance systems using a virtual train station environment populated by fully autonomous, lifelike pedestrians that perform various activities (Fig. 1) [4]. Within this environment, virtual cameras generate synthetic video feeds. The video streams emulate those generated by real surveillance cameras, and low-level image processing mimics the performance characteristics of a state-of-the-art surveillance video system. Recently, we have further developed the virtual vision approach to surveillance in sensor networks [5].

1.2 The virtual sensor network

Within the virtual vision paradigm, we propose a sensor network comprising wide field-of-view (FOV) passive cameras and PTZ active cameras to automatically capture and label high-quality video for every pedestrian that passes through a designated region. The network described herein, a special instance of the sensor network architecture proposed in [5], is capable of performing common visual surveillance tasks using local decision making at each camera node, as well as inter-node communication, without relying on camera calibration, a detailed world model, or a central controller.

Unlike our earlier work [5], we assume here that the wide-FOV static cameras are calibrated,[1] which enables the network to estimate the 3D locations of pedestrians through triangulation. However, we do not require the PTZ active cameras to be calibrated. Rather, during a learning phase, the PTZ cameras learn a coarse map between the 3D locations and the gaze-direction by observing a single pedestrian in the scene. A precise map is unnecessary since each PTZ camera is an autonomous agent that can invoke a search behavior to find a pedestrian using only coarse hints about the pedestrian's 3D position. The network uses a weighted round-robin strategy to assign PTZ cameras to surveil the various pedestrians. A new observation request is inserted into the task queue for every pedestrian that is sensed. Initially, each observation request is assigned the same priority; however, the decision making process uses domain-specific heuristics, such as the distance of the pedestrian from a camera or the heading of the pedestrian, to continuously evaluate the priorities of the observation requests. The PTZ cameras handle each task in priority sequence. The surveillance system issues a warning when an observation request cannot be met.

[1] This assumption is justifiable given the existence of several static camera calibration schemes [6,7].

1.3 The virtual world simulator

Our visual sensor network is deployed and tested within the virtual train station simulator that was developed in [4]. 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. The simulator can efficiently synthesize well over 1,000 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 virtual train station. They can enter the station, avoiding collisions when proceeding though 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. 1).

1.4 Contributions and overview

In this paper, we introduce a sensor management scheme that appears well suited to the challenges of designing camera networks for surveillance applications capable of fully automatic operation. We also develop new gaze-direction controllers for active PTZ cameras. Our work demonstrates the conveniences of the virtual vision paradigm for implementing, experimenting with, and evaluating our surveillance system.

The remainder of the paper is organized as follows: Sect. 2 covers relevant prior work. Section 3 overviews our system. We explain the low-level vision emulation in Sect. 4. Section 5 describes the PTZ active camera controllers and proposes a scheme for learning the map between 3D locations and gaze directions. Section 6 introduces our scheduling strategy. We present our results in Sect. 7 and our conclusions and future research directions in Sect. 8.