

Design and Implementation of a Wide Area, Large-Scale Camera Network

Thomas Kuo, Zefeng Ni, Carter De Leo, B.S. Manjunath

University of California, Santa Barbara

Department of Electrical and Computer Engineering, Santa Barbara, CA 93106

{thekuo, zefengni, cdeleo, manj}@ece.ucsb.edu

Abstract

We describe a wide area camera network on a campus setting, the SCALLOPSNet (Scalable Large Optical Sensor Network). It covers with about 100 stationary cameras an expansive area that can be divided into three distinct regions: inside a building, along urban paths, and in a remote natural reserve. Some of these regions lack connections for power and communications, and, therefore, necessitate wireless, battery-powered camera nodes. In our exploration of available solutions, we found existing smart cameras to be insufficient for this task, and instead designed our own battery-powered camera nodes that communicate using 802.11b. The camera network uses the Internet Protocol on either wired or wireless networks to communicate with our central cluster, which runs cluster and cloud computing infrastructure. These frameworks like Apache Hadoop are well suited for large distributed and parallel tasks such as many computer vision algorithms. We discuss the design and implementation details of this network, together with the challenges faced in deploying such a large scale network on a research campus. We plan to make the datasets available for researchers in the computer vision community in the near future.

1. Introduction

The analysis of video from a camera network covering a large area has been relatively unexplored. The technical, security, and privacy challenges have made it difficult to collect the data necessary for this research. Instead, most available datasets and networks use only a few cameras and observe a small area, such as a conference room or a courtyard. Larger camera networks do exist, but are run by governments and companies and are not publicly available.

We present the design and implementation of the Scalable Large Optical Sensor Network (SCALLOPSNet), a large camera network covering a wide area of a university campus. Using almost 100 stationary cameras, it observes indoor, urban, and remote scenes for human and natural ac-

tivity. In order to cover such diverse scenes, the network employs both wired cameras and battery or solar powered wireless nodes. It records and processes all of this data through a cloud computing-style cluster.

Our contributions include the deployment of one of the more diverse and expansive camera networks in the research community, the development and use of battery-powered wireless IP camera nodes for use in remote places, and the use of cloud computing infrastructure in our processing cluster for scalability. We will use this network to generate and make available datasets for computer vision research in the near future.

Section 2 describes existing datasets and camera networks available to the research community. Sections 3 and 4 describe the locations and architecture of the camera network. Section 5 demonstrates the application of tracking on our system. Finally, in Section 6, we describe some of the privacy and security issues we've encountered in the development of our system.

2. Prior Work

2.1. Available Datasets

The Workshops on Performance Evaluation of Tracking and Surveillance (PETS) [6] have released several free datasets with multiple, synchronized, and overlapping views of scripted events. The PETS 2001 dataset contains four short sequences observing people and vehicles moving on a small road from two cameras. The VS-PETS 2003 dataset contains views from three different corners of a soccer field during a game. These sets and their ground truth test object detection and tracking algorithms with single and multiple cameras.

PETS 2006 marks a transition to the higher level goal of detecting security and criminal events, specifically, abandoned luggage. A portion of a train station is viewed from four distinct views with actors leaving baggages. The following year, PETS 2007 added two more events, loitering and luggage theft, in a different four camera setup at a more crowded location.

Finally, PETS 2009 increased the number of cameras to eight and changed the goal to crowd observation. This included people counting and density estimation, tracking individuals in a crowd, and tracking the overall flow.

An unscripted dataset is available for a fee as the Multiple Camera Tracking dataset from the Image Library for Intelligent Detection Systems (i-LIDS) [1]. This data is sponsored by the UK Home Office Scientific Development Branch, and has been used in TRECVID Event Detection [14] and AVSS Multi-Camera Tracking Challenges [7]. It contains both overlapping and non-overlapping views from five CCTV cameras in the lobby of an airport.

2.2. Available Camera Networks

Without many available real-world datasets, several groups have opted to design their own smart cameras [13] or cameras networks [2].

Smart cameras developed by the research community include CITRIC [4], CMUCam3 [15], and MeshEye [8]. These generally use low-power processors like ARM or XS-scale, and 802.15.4 wireless communications. These cameras process the video on board and transmit only analyzed results.

One of the older and larger efforts in multiple visual sensors was VSAM (Video Surveillance and Monitoring) [5] at Carnegie Mellon University and Sarnoff Corporation. It incorporated multiple types of sensors including Pan-Tilt-Zoom (PTZ) cameras, infrared cameras, omnidirectional cameras, airborne cameras, and relocatable vehicle-mounted cameras in an effort to give a human operator the most useful information possible.

Other research efforts tend to be more modest. Many of them are temporary networks. For example, Wang et al. [16] collected two outdoor scenarios with four marginally overlapping cameras. Some more permanent camera networks exist, such as the KNIGHT^M system [9], which used a few cameras for detection and tracking in overlapping and non-overlapping views. At the Robotic Vision Lab at Purdue University, 18 Cyclops smart cameras with Mica-Z motes form a network, which provides an overhead view of one room [10]. UC Santa Barbara has a 10 camera overlapping-view network called VISNET [12], which is used for camera calibration and selection research.

One of the larger networks being run by a research institute is VideoWeb [11] at UC Riverside. It contains 80 Internet Protocol (IP) PTZ cameras mounted in and around a single building. The cameras connect wirelessly to low-level and mid-level processors that can be easily rearranged to test different processing paradigms. The system has been used to generate a dataset 2.5 hours of video data annotated with dozens of activities.



Figure 1: Camera (yellow circles) locations on one floor of the building; five per floor observing the door and elevators.

3. Camera Network Placement

One of our goals in the design of SCALLOPSNet is to enable wide-area camera network research. First, we selected three distinct regions that would benefit from large area studies: (1) inside a multi-story building, studying indoor human traffic flow and activity; (2) along a bicycle path, walkway, and traffic circle to study outdoor human traffic and activities; and (3) in a wilderness reserve enabling remote sensing and environmental studies.

3.1. Indoor Environment

The indoor camera network was placed inside a 60,000 square foot, five floor building. All floors consist of a single hallway leading to offices and lab spaces. Our aim is to observe the entrance and exits to every single floor, an area too large for one camera to cover. These locations are shown in Figure 1.

The relevant section of the hallways are about 8 feet tall by 254.5 feet long by about 7.5 feet wide. Network cable trays, which run on the south edge of every level except the ground floor, hang 10 inches from the ceiling. These trays carry Ethernet cables to each floor’s communication room, and, according to building code, cannot run AC power. Since they were our best option to mount the cameras and manage communications, we required the Power over Ethernet standard to power the cameras. On the first floor, the cameras hang from the ceiling and are connected by cables in the plenum.

There are 27 cameras in the building pointed at the entrances and exits to each floor. The first floor has seven cameras viewing five entrances, two stairwell doors, and the elevators. The remainder of the floors have five cameras viewing the two elevators and three stairwell entrances. The stairwells and elevators themselves, however, are not observed, creating definite non-overlapping views between floors.

This network can create good vision datasets with real-world challenges like compression artifacts, transmissions errors, and dynamic illumination resulting from non-uniform lighting, reflections, and open doors and windows. Potential tasks include:

- Detecting and tracking people with single and multiple

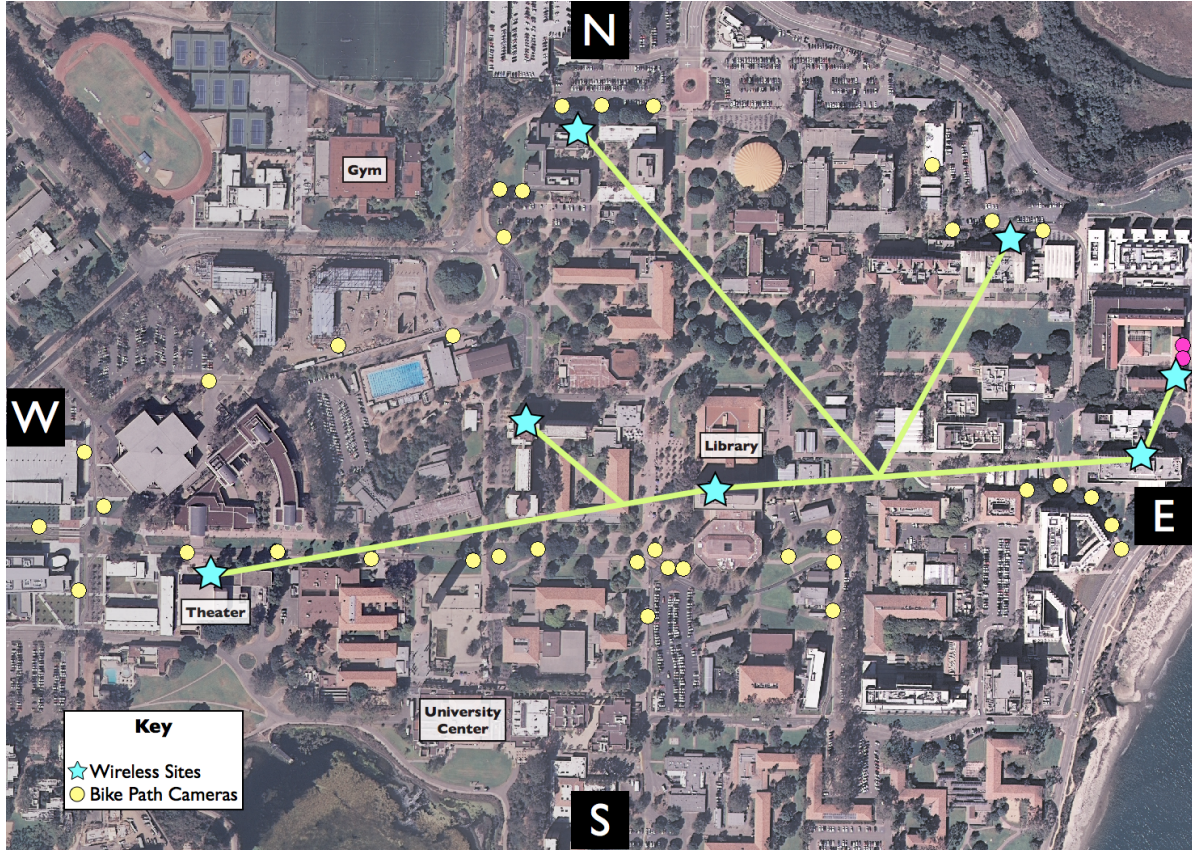


Figure 2: Camera (yellow circles) and wireless access point (blue stars) locations along the bicycle paths. The paths (green lines) connecting the square represent the 802.11a backhaul connection.

cameras, as shown in Sec. 5.

- Detecting indoor events like opening and closing a door, walking, running, and loitering.
- Counting the number of people on each floor.
- Modeling and summarizing typical indoor activities in order to extract abnormal behaviors.

3.2. Outdoor Urban Environment

The outdoor camera network covers 2 km of the main campus bicycle path, a popular walkway, and a vehicle traffic circle. These are shown in Figures 2 and 3. To discourage vandalism, cameras in these areas are mounted at 12 feet on light poles. As the entire bicycle path does not contain sufficiently tall poles, portions of the path are unobserved. From these vantage points, the cameras generally observe intersections, traffic circles, and other areas with high activity. However, care is also made in observing new areas not covered by other cameras.

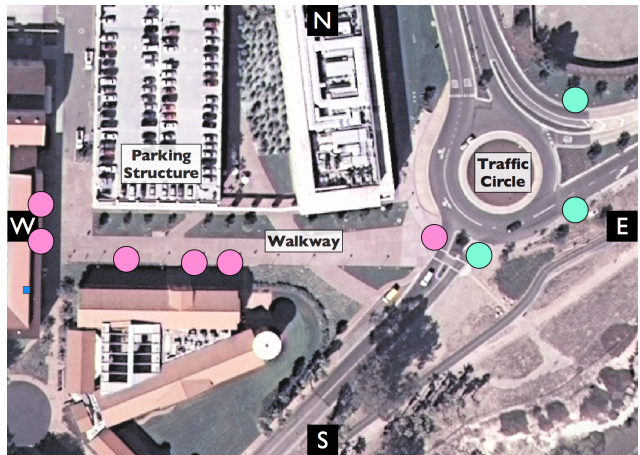


Figure 3: Camera (circles) locations at a walkway and traffic circle.

The light poles are not a reliable source for power and have no connections for communications. Therefore, we

needed a battery-powered camera node with a wireless connection. This is discussed in more detail in Section 4.

Datasets from this network can be used for vision tasks such as:

- Determining topology and links between camera views.
- Detecting and tracking pedestrians, bicycles, and vehicles across non-overlapping camera views.
- Modeling traffic flow and tracking crowds along the bicycle path.
- Detecting anomalous events such as accidents.
- Fusing information from multiple sources (e.g. visual data, class schedules, and classroom locations)

3.3. Outdoor Remote Environment

Our remote camera network is located in a natural reserve, which is shown in greater detail in our supplementary material. It is a protected habitat for the endangered Snowy Plover and is designated an Important Bird Area by the Audubon Society for the variety of birds that visit throughout the year.

Because it is remote and difficult to access, it has little power or communication infrastructure to rely upon. This necessitates again the use of wireless communications and battery power, especially with solar panels to maintain the charge of the battery.

As a natural reserve it lends itself to both vision and ecological problems such as:

- Counting the number of birds and recognizing their species.
- Correlating human activities with the population of animals.
- Tracking birds in 3D from multiple cameras.
- Detecting human trespassers into restricted areas.
- Tracking changes in plants throughout the season.

4. Camera Network Architecture

Our interests in wide-area camera network research lie in developing the appropriate software and algorithms as opposed to hardware. Therefore, we relied on commercially available equipment. Doing so required us to consider the camera, communications, and processing solutions available on the market. This resulted in the system shown in Fig. 4.

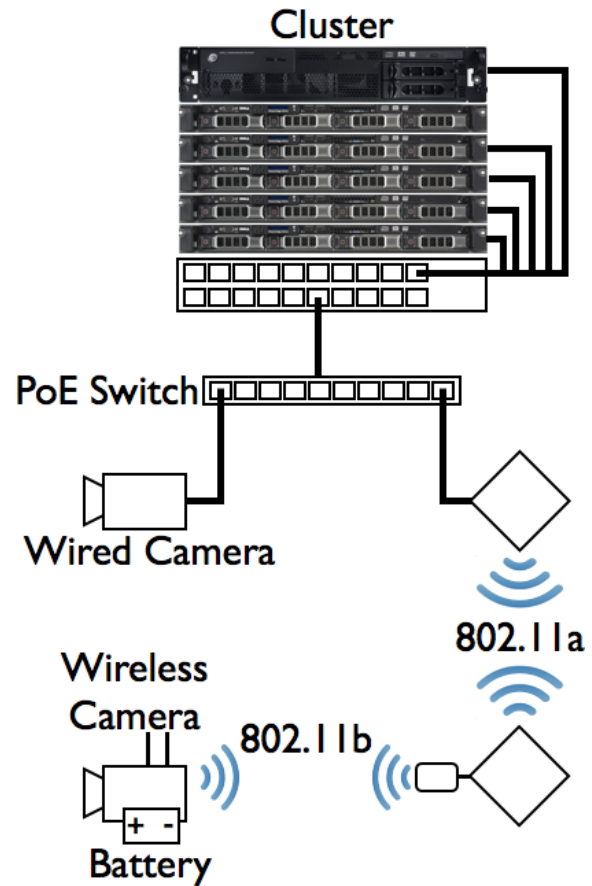


Figure 4: System architecture consists of cameras (Sec. 4.1), communications (Sec. 4.2), and compute cluster (Sec. 4.3).

4.1. Camera Nodes

The first camera requirement based on our site survey is that they use both wired and wireless communication and transmit at least VGA resolution (640×480) video over 320 meters. Indoor cameras also had to be able to receive power without running AC power in the cable trays, while outdoor cameras needed to be battery powered. Additionally, the outdoor cameras had to withstand environmental conditions like rain and fog.

Available wireless smart cameras use short range 802.15.4 wireless standard, which lack the throughput to stream video, especially across so many hops. Thus, the only most feasible choice was Internet Protocol (IP) cameras. Many IP cameras use the Power over Ethernet (PoE) standard to receive power over the Ethernet cable. Some have built-in 802.11b/g Wi-Fi. These solutions also have enclosures rated for outdoor use by either an International Protection Code (IP) or National Electrical Manufacturers Association (NEMA) rating.



Figure 5: The camera, a Linksys PVC2300 for indoor scenarios and WVC2300 in enclosure for outdoor scenarios.

We evaluated the Axis 211W and Linksys WVC2300 wireless network cameras. Both performed similarly. The Axis camera had slightly better image quality, but, we chose the Linksys camera, which was significantly less expensive. It can transmit up to VGA resolution video in either Motion JPEG or MPEG-4 streams. Transmission rates for the two streams are about 300 KB/s and 40 KB/s, respectively.

Power to indoor cameras is supplied by Power over Ethernet, a standard which transmits power over a wire on a Category 5 Ethernet cable. Outdoor cameras are powered by a 12V 12Ah battery. In our tests, these cameras could stay active for over 24 hours on a single charge. Given the size of our network, it's infeasible to change the battery so often. Therefore, we added a programmable switch that activates the camera only during set times. For cameras in the difficult to access natural reserve, we also attached a 20W solar panel and charge controller to trickle charge the battery.

To house our camera and battery, we needed enclosures that are at least IP43 or NEMA 3R. Enclosures with these ratings protect against falling dirt, rain, sleet, snow, and the external formation of ice. The cameras have a NEMA 3R rated enclosure with external antenna connection manufactured by Videolarm, while the batteries are stored in a separate NEMA 3R junction box. The camera nodes have so far withstood several rainstorms without incident.

4.2. Communications

Indoor cameras are easily connected using an Ethernet cable connected to a PoE switch, i.e. a switch that injects power into the Ethernet cable. The switch is connected over

gigabit Ethernet to our cluster.

To connect to the Wi-Fi cameras, we evaluated a number of options: (1) 802.11b terminals with an 802.11a backhaul, (2) WiMAX, (3) commercially available mesh networking with dual 802.11a/b/g radios, and (4) draft 802.11s mesh networking. The first option was the most feasible. WiMAX would have required a wireless bridge at each camera, which would have made our camera nodes too large. The dual radio mesh network was constrained to 3-4 hops in order to transmit video, which was infeasible given the size of our network. Finally, the 802.11s is still a draft and would have required us to modify the firmware on our cameras, which could potentially break them.

The final network consists of 7 wireless sites, which are connected to each other by 802.11a connections. These connections are shown in Figure 2. Each site has at least one 802.11b radio in order to connect to the cameras. The network is wired into our cluster via one of the PoE switches used for the indoor cameras. Average throughput for our 802.11a connections is 28 Mb/s and 8 Mb/s for the 802.11b connections. The average ping time between our recording computer and all the cameras is 85 ms, though this varies from camera to camera, with some only a few microseconds and other in the hundreds of microseconds.

4.3. Processing

There are two main paradigms for processing in a camera network: centralized and distributed. In centralized processing all data is handled by a single server. The advantages include flexibility and speed in developing and deploying new algorithms. The disadvantages are that adding cameras generally requires more bandwidth. In distributed processing, the raw video is processed at the node, and the processed data is distributed to neighbors. This allows for greater scalability and lower power consumption. However, doing so requires more advanced algorithms and makes it more difficult to record the video both for evaluating results and collecting data for distribution.

The lack of suitable smart cameras precludes us from implementing the distributed paradigm. Therefore, our network is centralized with all data collected by a central cluster. The cluster consists of 21 computers: 20 rack servers with Intel Quad Core Xeon Processor E5530, 6GB 1333MHz UDIMM memory, 500 GB SATA hard drive; and 1 rack workstation with Quad Core Intel Xeon Processor E5420, 8GB DDR2 667 MHz memory, and 500 GB SATA hard drive. Also on the network is a 5 TB network storage drive.

Our server is unique in the camera network community in that it takes advantage of technologies intended for cluster and cloud computing. At the infrastructure level, we've installed Apache Hadoop, a framework for running applications on large clusters built on commodity hard-

ware. At the heart of Hadoop is the computational paradigm Map/Reduce, which divides an application into small fragments of work, each of which may be executed on any node in the cluster. This fits very well with many video analysis algorithms, which can often be parallelized across the cameras, and across the spatial and temporal domains of one video.

5. Network Evaluation

Fig. 6 and 7 contain snapshots from our network – some of these videos will be available in the supplementary material. To better understand some of the challenges computer vision algorithms will face, we ran a multiple instance learning tracking algorithm by Babenko et al. [3].

In the indoor videos, see Fig. 6, two people enter the building, go to another floor, then return to the first floor and exit. In these videos, you can see the effect of non-uniform overhead lighting. Additionally, the reflections of people and lights are noticeable on the floor and walls. Cameras that can see windows and door ways must deal with larger dynamic ranges than other networks generally consider. Frame rates in these videos average about 26 fps, yet still show occasional artifacts as in the legs of the closest person in Fig. 6f.

The videos along the bicycle path, see Fig. 7, can contain pedestrian, bicycular, and vehicular traffic. Average frame rates of these videos are 23 fps, but can be 13 fps on a few cameras. The snapshots show fairly active intersections and the crowds that can be present. Also of note is the effect of the inclement weather, creating reflections and altering the dress and appearance of people.

The tracking results in both sequences perform reasonably well with expected errors during occlusions and severe scale changes, neither of which are accounted for by the algorithm.

6. Privacy and Security

Privacy and security are sensitive topics in the context of camera networks. Our discussion with the police established that it is legal for us to record video in any public area, but not audio. While United States federal law and most state laws allow for single party consent, i.e. only one party needs to be aware of the recording, some state laws require all parties to be aware. This is known as two party consent. Additionally, legal precedent in the US states that cameras not give the false impression of security. Therefore we've attached appropriate signs to inform passersby.

An ethicist emphasized that recorded data would have to be evaluated on a case-by-case basis before use in research. Her chief concern was that the data not record any instance when a person can reasonably expect privacy. In

cases where this might be violated, we should remove the identities of those concerned.

In the near future, we will record video from across our network. After the videos are synchronized and relevant identities anonymized, they must be approved for distribution. Then, they will be made available to the research community for studies like the ones mentioned in Sec. 3.

7. Conclusion

In this paper, we have described the design and implementation of SCALLOPSNet camera network, a large camera network covering very diverse regions including indoor and outdoor regions, and populated and remote areas. To do so with the available technology, we designed our own battery-powered wireless camera nodes for outdoor use and created a central cluster running cluster and cloud computing infrastructure, especially Apache Hadoop. We will soon make available datasets to address research problems such as object tracking and event detection in indoor and outdoor scenarios.

Future tasks focus on the development of computer vision algorithms designed for expanding cloud computing infrastructures and processes. Specifically, they need to be more parallel in order to fit into this framework.

With respect to the design of future wide-area camera networks, more effort is necessary in the design of wireless smart cameras that can also transmit video for evaluating the performance of algorithms. Additionally, our cameras are placed and pointed based upon our own subjective determination. Further work can help determine how the cameras can be better positioned to more meaningfully cover the region.

References

- [1] Home Office Scientific Development Branch Imagery Library for Intelligent Detection Systems (i-LIDS). <http://scienceandresearch.homeoffice.gov.uk/hosdb/cctv-imaging-technology/i-lids/>, March 2010. 2
- [2] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury. A survey on wireless multimedia sensor networks. *Computer Networks*, 51(4):921–960, 2007. 2
- [3] B. Babenko, M.-H. Yang, and S. Belongie. Visual tracking with online multiple instance learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 983–990, 2009. 6
- [4] P. Chen, P. Ahammad, C. Boyer, S. I. Huang, L. Lin, E. Lobaton, M. Meingast, S. Oh, S. Wang, P. Yan, A. Yang, C. Yeo, L. C. Chang, J. Tygar, and S. Sastry. CITRIC: A low-bandwidth wireless camera network platform. In *Int. Conf. on Distributed Smart Cameras*, pages 1–10, Sept. 2008. 2
- [5] R. T. Collins, A. J. Lipton, T. Kanade, H. Fujiyoshi, D. Dugins, Y. Tsing, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt,

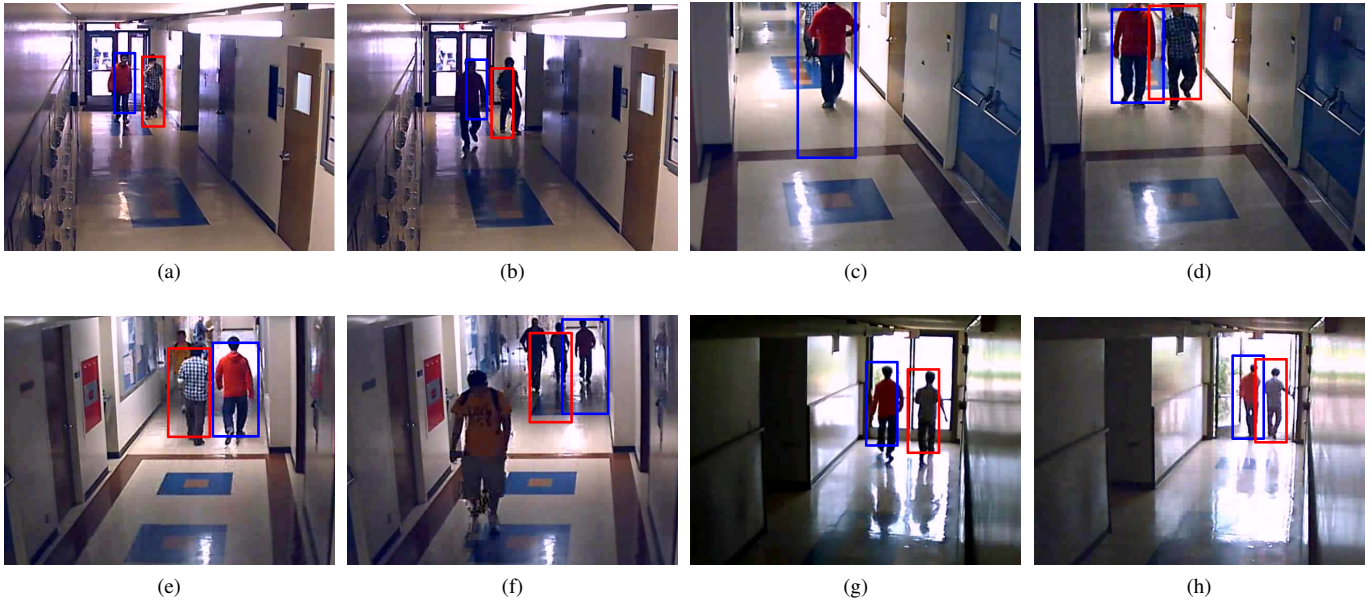


Figure 6: Tracking test with the indoor camera network. Note the dynamic lighting conditions and occasional artifacts, (f).

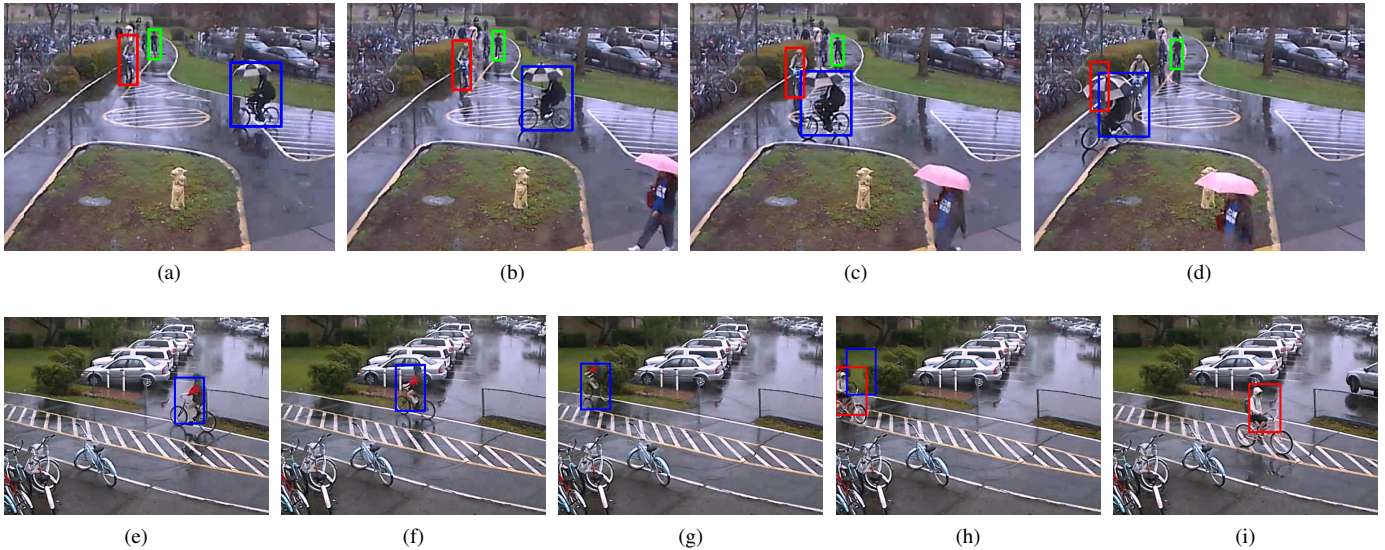


Figure 7: Tracking test with the outdoor camera network on a rainy day.

and L. Wixson. A system for video surveillance and monitoring. Technical report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, May 2000. 2

[6] J. Ferryman. PETS. In *IEEE Int. Workshops on Performance Evaluation of Tracking and Surveillance (PETS)*, <http://www.visualsurveillance.org>, 2000-2009. 1

[7] J. Ficus, J. Garofolo, T. Rose, and M. Michel. AVSS multiple camera person tracking challenge evaluation overview. In

Int. Conf. on Advanced Video and Signal Based Surveillance, pages 219 – 219, 2009. 2

[8] S. Hengstler, D. P. ans S. Fong, and H. Aghajan. MeshEye: A hybrid-resolution smart camera mote for applications in distributed intelligent surveillance. In *Int. Symposium on Information Processing in Sensor Networks*, 2007. 2

[9] O. Javed, Z. Rasheed, O. Alatas, and M. Shah. Knightm: A real time surveillance system for multiple overlapping and

non-overlapping cameras. *4th International Conference on Multimedia and Expo*, 2003. 2

- [10] H. Medeiros, J. Park, and A. C. Kak. A light-weight event-driven protocol for sensor clustering in wireless camera networks. In *Int. Conf. on Distributed Smart Cameras*, 2007. 2
- [11] H. Nguyen, B. Bhanu, A. Patel, and R. Diaz. VideoWeb: Design of a wireless camera network for real-time monitoring of activities. In *Int. Conf. on Distributed Smart Cameras*, 2009. 2
- [12] M. Quinn, R. Mudumbai, T. Kuo, Z. Ni, C. DeLeo, and B. S. Manjunath. VISNET: A distributed vision testbed. In *Int. Conf. on Distributed Smart Cameras*, 2008. 2
- [13] B. Rinner and W. Wolf. Towards pervasive smart camera networks. In H. Aghajan and A. Cavallaro, editors, *Multi-Camera Networks*, chapter 20, pages 483–496. Elsevier, 1st edition, 2009. 2
- [14] T. Rose, J. Ficus, P. Over, J. Garofolo, and M. Michel. The trecvid 2008 event detection evaluation. In *Workshop on the Applications of Computer Vision*, pages 1 – 8, 2009. 2
- [15] A. Rowe, A. G. Goode, D. Goel, and I. Nourbakhsh. CMU-Cam3: An open programmable embedded vision sensor. Technical report CMU-RI-TR-07-13, Carnegie Mellon University, May 2007. 2
- [16] X. Wang, K. Tieu, and W. E. L. Grimson. Correspondence-free activity analysis and scene modeling in multiple camera views. *Trans. on Pattern Analysis and Machine Intelligence*, 32(1):56–71, January 2010. 2