TWO EXAMPLES OF INDOOR AND OUTDOOR SURVEILLANCE SYSTEMS: MOTIVATION, DESIGN, AND TESTING

Ioannis Pavlidis
Honeywell Laboratories
3660 Technology Drive
Minneapolis, MN 55418
U.S.A. Partial funding provided by Honeywell Laboratories.
ioannis.pavlidis@honeywell.com

Vassilios Morellas
Honeywell Laboratories
3660 Technology Drive
Minneapolis, MN 55418
U.S.A. Partial funding provided by Honeywell Laboratories.
vassilios.morellas@honeywell.com

Abstract We examine the state of the security industry and market and underline the role that it plays in the R&D efforts. We also present a snapshot of the current state-of-the-art in indoor and outdoor surveillance systems for commercial applications. Then, we move on and describe in detail a prototype indoor surveillance system that we recently developed. The system is called Cooperative Camera Network (CCN) and reports the presence of a visually tagged individual throughout a building structure. Visual tagging is based on the color signature of a person. CCN is meant to be used for the monitoring of potential shoplifters in department stores. We also describe our prototype outdoor surveillance system, the DETER (Detection of Events for Threat Evaluation and Recognition). DETER can monitor large open spaces, like parking lots, and report unusual moving patterns by pedestrians or vehicles. To perform its function DETER fuses the field of views of multiple cameras into a super-view and performs tracking of moving objects across it. A threat assessment module with encoded suspicious behaviors performs the motion pattern identification. Both surveillance systems are good examples of technology transfer of state-of-the-art ideas from the research literature to the commercial domain. At the same time, they are

good study cases for the extra engineering methodology and effort that is needed to adapt initial research concepts into a successful practical technology.

Keywords: Surveillance, security systems, threat assessment, color recognition, multinormal representation.

1. Introduction

The current security infrastructure could be summarized as follows: (a) Security systems act locally and they do not cooperate in an effective manner. (b) Very high value assets are protected inadequately by antiquated technology systems. (c) Reliance on intensive human concentration to detect and assess threats.

Today there is a chasm between what is commercially available and the security technology demonstrated in R&D labs. On one hand, the most sophisticated security product one can find is a camera with some rudimentary motion detection capability. On the other hand, complete multi-camera prototype systems that cooperate at several levels and feature automated threat assessment have proved themselves in realistic test environments. Why the security industry has been so slow in productizing concepts and designs fully developed and tested by academic and commercial labs over the last 10 years? An understanding of the industry's peculiarities and the forces that shape up its current profile is essential for anyone who is interested to perform technology transfer in the security domain. Below we enumerate what we consider the most important characteristics of the current security market and industry.

Low Profit Margin The security market is very cost sensitive. In an era, where quarterly profits make or brake corporate giants in areas with much higher profit margin, the security industry always struggles to "make the numbers." Its strategic horizon usually does not extend beyond six months.

Resistance to Change Like most traditional industries the security industry is not an advocate of innovation by nature. It is characteristic that most of the commercial security R&D was initiated and financed directly by corporate mandate and not by the individual business units.

Low Tech Culture The security industry is permeated by low tech culture. The management and the engineers of the security business units are trained and grown within a low tech environment and are ignorant and suspicious to state-of-the-art developments. Their users and customers are often underpaid and under-educated

security guards and facility managers that also view high technology with skepticism.

Hardware Mentality The most advanced members of the security industry are probably the camera manufacturers. Even these, although they produce some advanced electronic products have difficulty outfitting them with the necessary software.

Despite the presence of many negative factors the future of the security industry can be viewed only in positive light. And, although the transformation of the industry and the market will take time to complete, it has already started happening in small steps. As a result of upcoming technology offerings the Freedonia group [fre, 2000] is projecting significant growth of the security service market during the next several years. This growth will fuel further research and development and will hopefully bootstrap the process of incorporating the security industry to the new economy.

2. The State of the Art in Video-Based Security Technology

The computer vision community has performed extensive research in the area of video-based surveillance for the past 20 years. This research, although initially military in nature, it turned very quickly into civilian (security) with the end of the cold war. One can identify two rather distinct application categories: indoor surveillance and outdoor surveillance systems. From the technology point of view outdoor surveillance is much more challenging because of the greater variability in lighting conditions. It is ironic, however, that most of the R&D work has been directed towards outdoor rather than indoor surveillance. This trend may be partly due to the military heritage of the computer vision community and partly to the wider social acceptability of outdoor as opposed to indoor (e.g. home, workplace) surveillance and monitoring.

In indoor surveillance the state of the art [Cai et al., 1995, Cai and Aggarwal, 1999a, Cai and Aggarwal, 1999b, Ziliani and Cavallaro, 1999] features robust motion detection and tracking algorithms. Indoor environments are composed of many relatively small spaces that are separated with walls and communicate with each other through doors and corridors. In this situation it is important to associate the presence of a particular individual in different parts of the building structure. It is less important to track continuously the motion of an individual as this motion is bound to break quite often due to the building's topography.

In outdoor surveillance the state of the art [Kanade et al., 1998, Stauffer and Grimson, 1999, Staufer and Grimson, 2000] also features robust

motion detection and tracking algorithms. The difference is that these algorithms are usually much more sophisticated than the corresponding indoor algorithms due to the complexity introduced by highly variable lighting. The topology of outdoor environments is also very different than that of indoor environments. The large open spaces invite for continuous object tracking. Moving objects are not only humans but vehicles as well, traveling at significantly higher speeds. Faster moving objects necessitate faster processing speeds, yet the algorithms are much more computationally intensive than those applied to indoor surveillance scenarios. These contradictory requirements simply add up to the technical challenges of an advanced outdoor security system.

3. CCN - A Prototype Indoor Surveillance System

We have developed a prototype indoor surveillance system that monitors human presence in and around our lab. We dubbed the system *Cooperative Camera Network or CCN* as it demonstrates cooperation between the different camera nodes through sharing of visual information. By camera node here we mean the combination of a camera with a PC. The hardware architecture of the system is depicted in Fig. 1. Each camera sends its live video feed to a networked PC. Depending on how powerful the processor is, the computational requirements of more than one live video feeds (cameras) can be accommodated by the same PC. The CCN architecture is highly modular allowing easy expansion at a low cost since it uses common off-the-shelf components. We consider that CCN type of systems should be preferably outfitted with USB or 1394 digital cameras. Both the USB and 1394 are ubiquitous serial interfaces and provide easy camera connectivity to the PC processors. Since the information pipeline is all digital no time is lost in Digital to Analog conversion and the quality of the video signal remains high. Also, full computer control of all the camera functions is possible, including the setting of brightness, contrast, aperture, and zoom values. In the incarnation of CCN in our lab we use the Sony 1394 DFW-VL digital camera model. The cameras sit atop Pan/Tilt devices that are controlled by the PCs through RS-232 connections. The Pan/Tilt devices facilitate easy camera repositioning and enable in-room tracking if desirable.

Currently, the CCN system consists of 3 camera nodes. One camera is physically located in our lab. The other two are located in two rooms consecutive to our lab (see Fig. 2). All the cameras are directed towards the rooms' doors. This is an exemplary setup for controlling human traffic in a subset of a building. We are interested in detecting

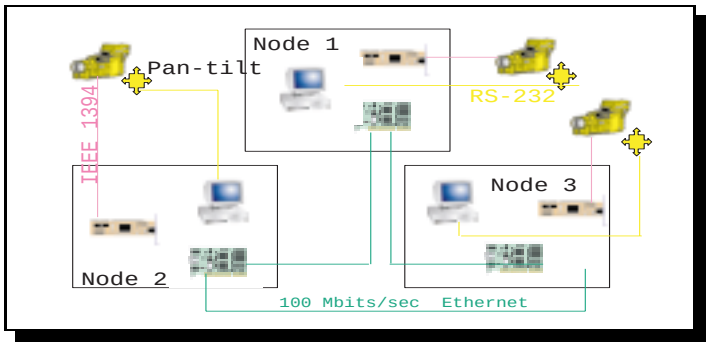

Figure 1: The CCN hardware architecture. More nodes can be added to the system to control larger portions of the building.

and recording human presence in the rooms under surveillance. We are also interested in communicating detection information from one camera node to the other. This information could be as simple as the location and time of detection or as complicated as the color signature of the detected individual. Nominal location and time of incident information enables the system to reason as to where to expect the next move. For example, the information that an individual was detected in Room 1304 puts on alert the camera node in the adjacent Room 1303 (see Fig. 2). The color signature information enables the reporting of a visually tagged individual as he moves from room to room. Visual tagging of a particular individual is realized on demand by pointing and clicking on his silhouette in the corresponding live video stream at the central console of the system. CCN isolates the blob of the present human from the rest of the background thanks to its differencing and segmentation [Otsu, 1979] algorithm. Then, CCN produces the color signature [Funt and Finlayson, 1995] for the individual and propagates it across the network of camera nodes. A copy of the signature is stored on each local model database under a single unique id. From that point on, every time a camera node detects human presence extracts the color signature of the detected individual and compares it against the color models stored in the database. If a match is found, the presence of the particular individual stamped by location and time is reported back to the central console of the system. A separate reporting window is maintained for each individual whose color signature has been stored in the database.

CCN is meant to enhance the effectiveness of a security guard that monitors human traffic in a commercial building. A very typical application is the monitoring of various parts of a department store for potential shoplifters. A security guard in such a case has to maintain

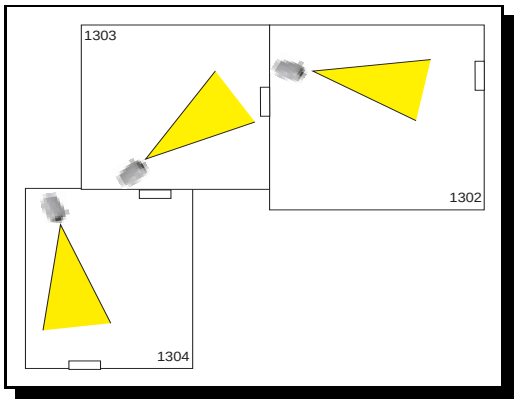

Figure 2: The topology of the CCN system in and around our lab (Room 1304).

awareness about video information displayed on dozens of screens. If at some point he notices a suspect it is very difficult to keep track of him as he moves out of the specific camera's field of view to another part of the store. It is typical in those cases for security the officer in charge to dispatch another guard, the so called 'shadow', to physically follow the suspect while he is present in the store. Since many suspect cases do not materialize this is an expensive use of security resources in a major department store. CCN allows the security officer to track not one but multiple suspects at one time semi-automatically, obviating the need of 'shadows'. Therefore, CCN does not only increase overall security in the department store scenario but also reduces its cost. Per our argumentation in Section 1 this would be a major selling point to conservative customers such as facility managers of department stores.

3.1. Testing of the CCN System

The CCN system has been tested extensively in a real building scenario. It performs flawlessly under the following two conditions:

1 The humans moving around the building are not dressed very similarly or the same.

2 The human traffic is sparse. Dense human traffic tends to occlude part of the lower body, thus reducing the discriminatory power of the color signature.

In a typical commercial building there is enough variability in people's apparel to render the first condition a non-issue. Of course, CCN would never be an ideal surveillance system for a boot-camp. Dense human

traffic, however, is typical in department stores during the holiday season. Therefore, this is a condition that should be addressed before CCN becomes a viable product.

Fig. 3 shows some results from the segmentation and color recognition algorithm of CCN. It is worth noting that the CCN color recognition algorithm is sufficiently robust to orientation and size changes. While the subject's orientations differ 90^0 the color recognition algorithm still performs successfully.

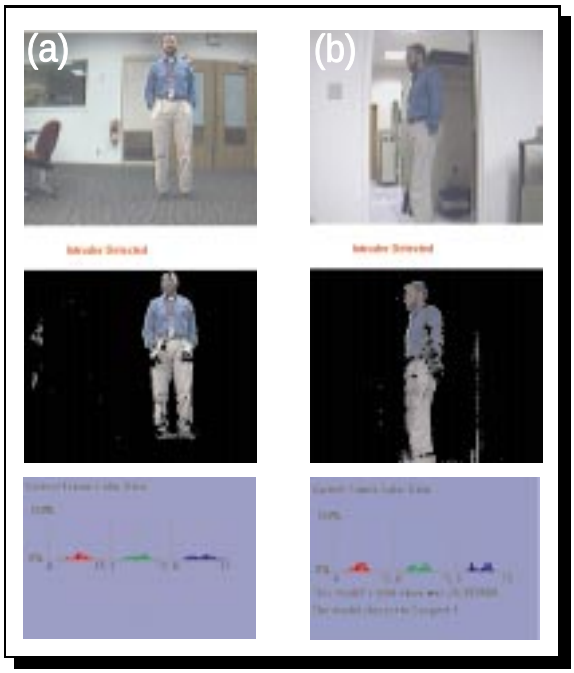

Figure 3: The color matching algorithm of CCN in action. (Column a) The top frame represents a snapshot from a live video feed. The mid frame shows the result of segmentation. The bottom frame shows the color signature of the segmented object as it was recorded in the CCN database. (Column b) Snapshot of the individual in a different room under surveillance. The mid frame shows the result of segmentation this time. The bottom frame shows the color signature of the segmented object this time and the successful matching result. At the time of this experiment a total of 3 color signatures were active in the model database.

4. DETER - A Prototype Outdoor Surveillance System

A comprehensive urban video surveillance system, such as DETER, depends primarily on two different technologies: computer vision and threat assessment. The computer vision part consists of the optical and system design, the moving object segmentation and tracking, and the multi-camera fusion stages. The threat assessment part consists of the feature assembly, the off-line training, and the threat classification stages (see Fig. 4). We will give a brief overview of each stage and compare our solutions to others proposed in the literature.

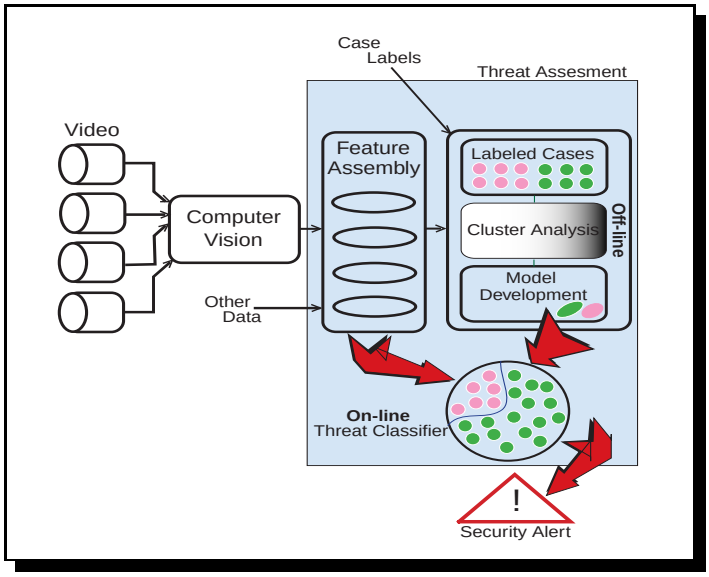

Figure 4: Architecture of the DETER system.

Our system is probably the only one that features a formal optical and system design stage. Most of the efforts reported in the literature had as their main objective to demonstrate the feasibility of a novel idea and they didn't pay any attention to the practical aspects of fielding a surveillance system. There is a number of requirements that a surveillance system needs to fulfill to function properly and be commercially viable. First, it should ensure full coverage of the open space or blind spots may pause the threat of a security breach. It is often argued in the technical literature that video sensors and computational power are getting cheaper and therefore can be employed in mass to provide coverage for any open space [Grimson et al., 1998]. In reality things are not so rosy. Most of the cheap video sensors do not still have the required

resolution to accommodate high quality object tracking. Both cheap and expensive cameras also need to become weather proof for employment outdoors, which increases their cost substantially. Then, it is the issue of installation cost that includes the provision of power and the transmission of video signals, sometimes at significant distances from the building. The installation cost for each camera is usually a figure many times its original value. Even if there were no cost considerations, cameras cannot be employed arbitrarily in public places. There are restrictions due to the topography of the area (e.g. streets, tree lines) and due to city and building ordinances (e.g. aesthetics). All these considerations severely curtail the allowable number and positions of cameras for an urban surveillance system.

In addition to optical considerations there are also system design considerations including the type of computational resources, the computer network bandwidth, and the display capabilities. Due to the cost sensitivity of the security market all these become critical issues and should be addressed in an optimal manner.

We achieve motion segmentation (see Fig. 5) through a multi-Normal representation at the pixel level. Our method resembles the method described in [Staufer and Grimson, 2000] with some interesting modifications. The method identifies foreground pixels in each new frame while updating the description of each pixel's mixture model. The labeled foreground pixels can then be assembled into objects using a connected components algorithm. Establishing correspondence of objects between frames (tracking) is accomplished using a linearly predictive multiple hypotheses tracking algorithm which incorporates both position and size.

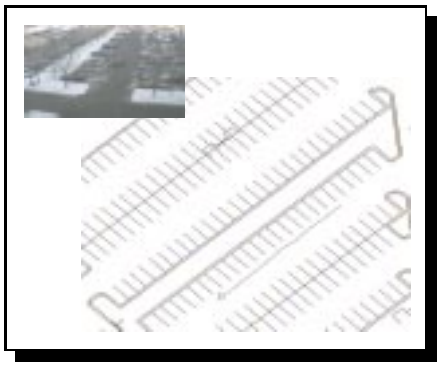

Figure 5: Live video snapshot of a car moving out of the parking lot and its itinerary (green line) as it is recorded by DETER at the CAD design level.

No single camera is able to cover large open spaces, like parking lots, in their entirety. Therefore, we need to fuse the Fields of View (FOV) of the various cameras into a coherent super picture to maintain global awareness. We fuse (calibrate) multiple cameras by computing the respective homography matrices. The computation is based on the identification of several landmark points in the common FOV between camera pairs.

The threat assessment portion of DETER consists of a feature assembly module followed by a threat classifier. Feature assembly extracts various security relevant statistics from object tracks and groups of tracks. The threat classifier decides in real time whether a particular point in feature space constitutes a threat. The classifier is assisted by an off-line threat modeling component (see Fig. 4).

DETER, like CCN, is a departure from the traditional proprietary security mentality. Both systems are implemented with off the shelf components. DETER can run on very moderate PC hardware (233 MHz) and its architecture provides for standard analog cameras that abound in commercial buildings. Therefore, PC surpluses and existing cameras can be used to build DETER, reducing dramatically its installation cost.

4.1. Testing of the DETER System

DETER is a fully operational prototype that performs perimeter surveillance in the Honeywell Laboratories building in Minneapolis. In addition to the qualitative testing performed by the actual users (security guards) we also performed quantitative testing for benchmarking purposes. Since August $11th$, 2000 we measured the tracking performance of DETER in the Honeywell Laboratories parking lot for 8 hrs. The testing was done in 1 hr increments spread over different days, times of day, and seasons. Table 1 shows the results of the DETER performance in the field tests. Parking lot activity included walking and running of a single individual, simultaneous walking of a number of individuals(following crossing or parallel paths), driving of a single and multiple cars, and finally a combination of cars and humans in motion. Some staged events included geometrically interesting walking patterns such as the ones we call *M-Patterns* (favorite by car thieves) and dangerous driving. These events were identified as suspicious by the Threat Assessment classifier.

Perfect Tracks	Split Tracks	Joint Tracks	False Alarms	Missed Tracks
554	77	16	5	3

Table 1: Experimental results for the 8-hour-long data set.

DETER detected and tracked perfectly 554 objects out of 666 - a remarkable performance. In 77 instances DETER has lost momentarily track of the object but regained it very quickly. The result was a split track. That was typically the case with pedestrians as they ventured momentarily under the tree lines (summer and early fall trials). In a few occasions (16) where pedestrians were moving next to each other (party of two) DETER correctly detected and tracked the motion but as a single object. This is a camera resolution problem. If we covered less area with each camera the resolution would have been better and the segmentation of closely spaced moving objects more accurate. DETER produced a small number of false alarms. Four of the five false alarms were produced in a snowy day as accumulated iced snow was hovering from the top cover of one of the cameras.

5. Summary

We have described two prototype surveillance systems, one for indoor use (CCN) and one for outdoor use (DETER). As a way of introduction we also discussed the state of the security industry and market and the role it played in our project decisions. CCN features a relatively simple visual motion detector based on image differencing and adaptive thresholding, which suffices for the locally stable illumination in indoor spaces. CCN can also record on demand the color signature of the segmented individual and propagate it across the camera network of the building. From that point on all the cameras are actively matching the color signatures of passing humans and report back to a central console if a match was found. We have adopted an illumination independent color matching method to accommodate for the different illumination intensities and spectral compositions across a typical commercial building. CCN is quite resilient in variations of object size and orientation. It has only been tested though in sparse human traffic scenarios. The challenge would be to extend the CCN functionality to dense crowd situations, typical in commercial buildings during peak times. This is an area our current work is focusing on.

DETER consists of a computer vision module and a threat assessment module. The two primary components of the computer vision module is the moving object segmenter and the associated tracker. The threat assessment module reports suspicious patterns detected in the annotated trajectory data at the CAD level. The threat assessor also uses the information produced by the computer vision module to perform some non-security functions, like monitoring the capacity of the parking lot. We are working towards the improvement of the threat assessment mod-

ule with the inclusion of a clustering algorithm. The clustering algorithm will help in the partial automation of the off-line training, currently performed manually. DETER is scheduled for productization in 2002.

References

[fre, 2000] (2000). World security services to 2004. Technical Report 1348, The Freedonia Group.

[Cai and Aggarwal, 1999a] Cai, Q. and Aggarwal, J. (1999a). Automatic tracking of human motion in indoor scenes across multiple synchonized video streams. In *Proceedings Sixth International Conference on Computer Vision*, pages 356–362, Bombay, India.

[Cai and Aggarwal, 1999b] Cai, Q. and Aggarwal, J. (1999b). Tracking human motion in structured environments using a distributed camera system. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(11):1241–1247.

[Cai et al., 1995] Cai, Q., Mitiche, A., and Aggarwal, J. (1995). Tracking human motion in an indoor environment. In *Proceedings 1995 IEEE International Conference on Image Processing*, volume 1, pages 215–218, Washington D.C.

[Funt and Finlayson, 1995] Funt, B. and Finlayson, G. (1995). Color constant color indexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5):522–529.

[Grimson et al., 1998] Grimson, W., Stauffer, C., Romano, R., and Lee, L. (1998). Using adaptive tracking to classify and monitor activities in a site. In *Proceedings 1998 IEEE Conference on Computer Vision and Pattern Recognition*, pages 22–29, Santa Barbara, CA.

[Kanade et al., 1998] Kanade, T., Collins, R., Lipton, A., Burt, P., and Wixson, L. (1998). Advances in cooperative multi-sensor video surveillance. In *Proceedings DARPA Image Understanding Workshop*, pages 3–24, Monterey, CA.

[Otsu, 1979] Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66.

[Staufer and Grimson, 2000] Staufer, C. and Grimson, W. (2000). Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):747–767.

[Stauffer and Grimson, 1999] Stauffer, C. and Grimson, W. (1999). Adaptive background mixture models for real-time tracking. In *Proceedings 1999 IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 246–252, Fort Collins, CO.

[Ziliani and Cavallaro, 1999] Ziliani, F. and Cavallaro, A. (1999). Image analysis for video surveillance based on spatial regularization of a statistical model-based change detection. In *Proceedings 1999 International Conference on Image Analysis and Processing*, pages 1108–1111, Venice, Italy.