Intelligent Video Surveillance in Crowds

Dr. Alan J. Lipton
Chief Technology Officer
ObjectVideo
11600 Sunrise Valley Dr, Suite 290
Reston, VA 20191
ajl@objectvideo.com

Abstract

Organizations today face new and more insidious threats than they ever have in the past. To protect personnel and infrastructure alike requires a level of vigilance not previously anticipated. A new technology called Intelligent Video Surveillance employs state of the art computer vision technology to automate the process of watching CCTV video signals – making video a proactive defense sensor. Unfortunately, the state of the art in computer vision is somewhat limited when scenes become very complex – it's very hard to teach a computer to understand what is going on in a crowded scene. One approach to solving this problem is to detect and recognize the actions of individuals within a crowded scene. This technology is, as yet, in its infancy. Another approach is to use an ***optical flow*** *technique to track the gross motion of a crowd. From this, many threatening activities may be inferred even if the individual perpetrators are not identified. A specific example of this type of technology is described that can detect counter-flow – that is, someone moving against the flow of a crowd. This functionality has applications in public safety, traffic monitoring, and airport security.*

1. Introduction – Intelligent Video Surveillance (IVS)

Recent world events have prompted government and industry organizations alike to rethink their approach to physical security. The threats we face today are no longer large scale military attacks from known adversaries outside our borders. Our fears today derive from the possibility of a small group of individuals, perhaps already within our borders, having the ability to cause a large amount of damage. Such attacks could carry an extremely high cost in terms of economic and environmental damage, reduced national morale, and loss of human life. Not only has the nature of the threat changed, but recent events have redefined the nature of targets. No longer are prime targets military in nature – now public infrastructure and innocent civilians are facing attack. Organizations that control critical infrastructure and national assets such as airports, power production facilities, water supplies, and public transportation routes are feeling the pressure to increase their ability to detect "asymmetric threats" and respond to them in a timely manner.

These changes have forced a higher level of vigilance upon many organizations previously unconcerned with major attack. Accordingly, we see an increase in the awareness of physical security issues and technologies along with increases in physical security budgets. The relatively new Department of Homeland Security, for example, is working with a budget of some $37B. A very large piece of the physical security pie is being devoted to video surveillance infrastructure and research.

Why video? People like video. It's one of the most ubiquitous sensing modalities available. It is real-time, cheap, and highly intuitive (it's easy to understand what is happening in a video stream). Yet, curiously, video surveillance is not used primarily for real-time interdiction. It is used in two basic modes: as a deterrent and as a forensic tool. People are less likely to commit criminal activities if they believe they will be caught on camera; and if something does occur video is frequently used forensically to figure out what happened. Hence there is an apparent paradox: video is a ubiquitous, real-time, intuitive sensor that is *not* being used to provide real-time actionable intelligence.

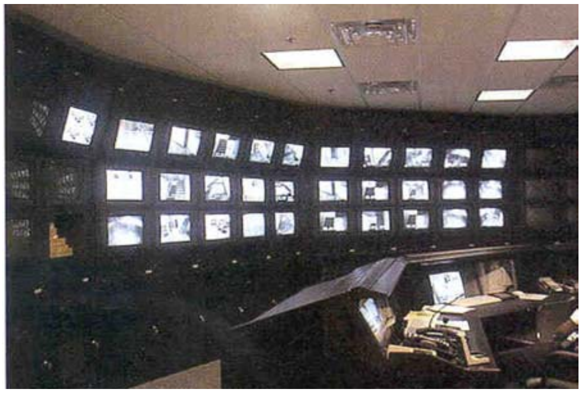


Figure 1 – Today's video surveillance system

Figure 1, for example, shows a "state of the art" video security system. Organizations often spend millions of dollars on video surveillance infrastructure consisting of hundreds or thousands of cameras. These camera feeds are usually backhauled to a central monitoring location where some of them are recorded for a period of time on local video storage media, and some of them are displayed in real-time to one or more security personnel on a bank of video monitors. No matter how highly trained or how dedicated a human observer, it is impossible to provide full attention to more than one or two things at a time; and even then, only for a few minutes at a time. A vast majority of surveillance video is permanently lost without any useful intelligence being gained from it. The situation is analogous to an animal with hundreds of eyes, but no brain to process the information.

The solution to this problem is intelligent video surveillance (IVS) [3,4,5,6,7,8]. That is, computer software that watches video streams to determine activities, events, or behaviors that might be considered suspicious and provide an appropriate response when such actions occur. The key technology is called ***Computer Vision***. This is a somewhat specialized branch of mainstream artificial intelligence research involving teaching machines to understand what they "see" through a camera. Traditionally, computer vision has had limited success in real-world commercial applications. The best available automated video surveillance capability involved a somewhat simplistic technology called Video Motion Detection (VMD) that is notorious for false alarms in realistic operational environments (see Figure 2). But recent advances in technology and computational power along with a move of key talent from academia into industry have allowed computer vision to come out of the lab and into commercial video surveillance products.

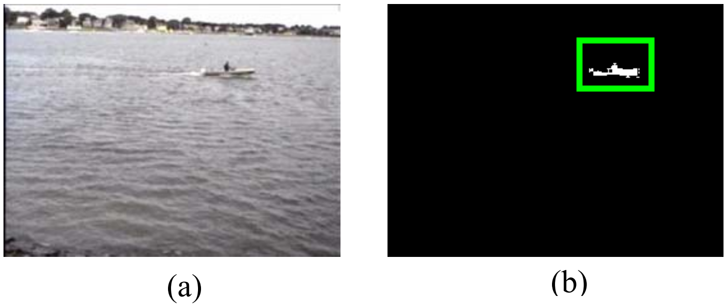


Figure 2 - VMD vs. IVS. (a) The source image - everything is moving.
(b) IVS accurately detects the object

ObjectVideo is one company that has successfully expanded and commercialized some government funded computer vision research technology. This technology is used as the basis of an IVS product called VEW (Video Early Warning) that monitors video streams in real-time and detects activities that have been prescribed as interesting or suspicious.

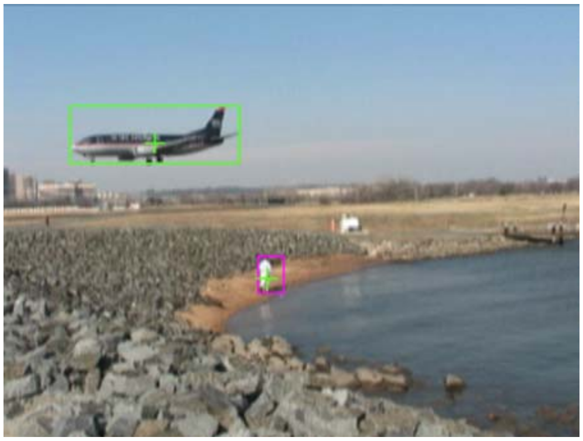


Figure 3 - Objects detected on the perimeter of Reagan National Airport

ObjectVideo's solution watches video streams and extracts descriptions of all relevant objects. It employs sophisticated algorithms for detection [3] and tracking of all relevant objects in the camera's view. It also contains algorithms for classification [10] of objects into specific types[1]. Figure 3 illustrates an image from ObjectVideo's application. Here an unauthorized human has been detected on the perimeter of an airport.

[1] The basic product classifies objects into "human" and "vehicle" classes, however the product also provides a simple mechanism for custom classification algorithms to be developed for customers with more specific needs such as to distinguish humans from animals, or trucks from aircraft.

2. Monitoring Crowds

What may be obvious from the example presented above is that IVS technology is extremely robust in applications involving physical isolation such as perimeter security, intrusion detection, and after-hours monitoring. This is because, until recently, computer vision technologists have shied away from crowded scenes. It's not that there are no security applications for crowded environments – quite the reverse – but the science of computer vision, in its current form, starts to break down when there is a lot of activity present. When a human looks at complex video scene, we intrinsically understand the interplay between objects in the scene. We know what people and vehicles look like and we can disambiguate one object from another. Computers, at this point, are not generally that smart – without very rigorous training, they can't distinguish one object from another when there is a complex jumble of motion in a scene. As far as the computer is concerned there is significant activity that cannot be resolved into individual objects or actions – see Figure 4.

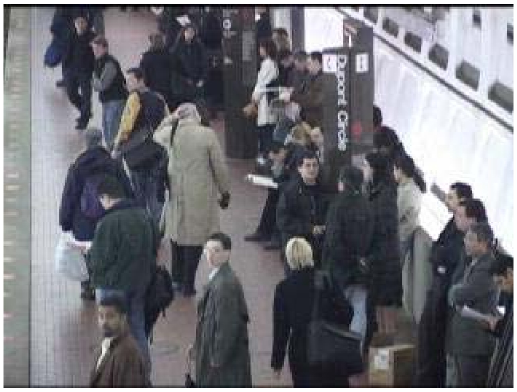


Figure 4 - Complex crowded scene. Here it is dificult for a computer to recognize individual actions.

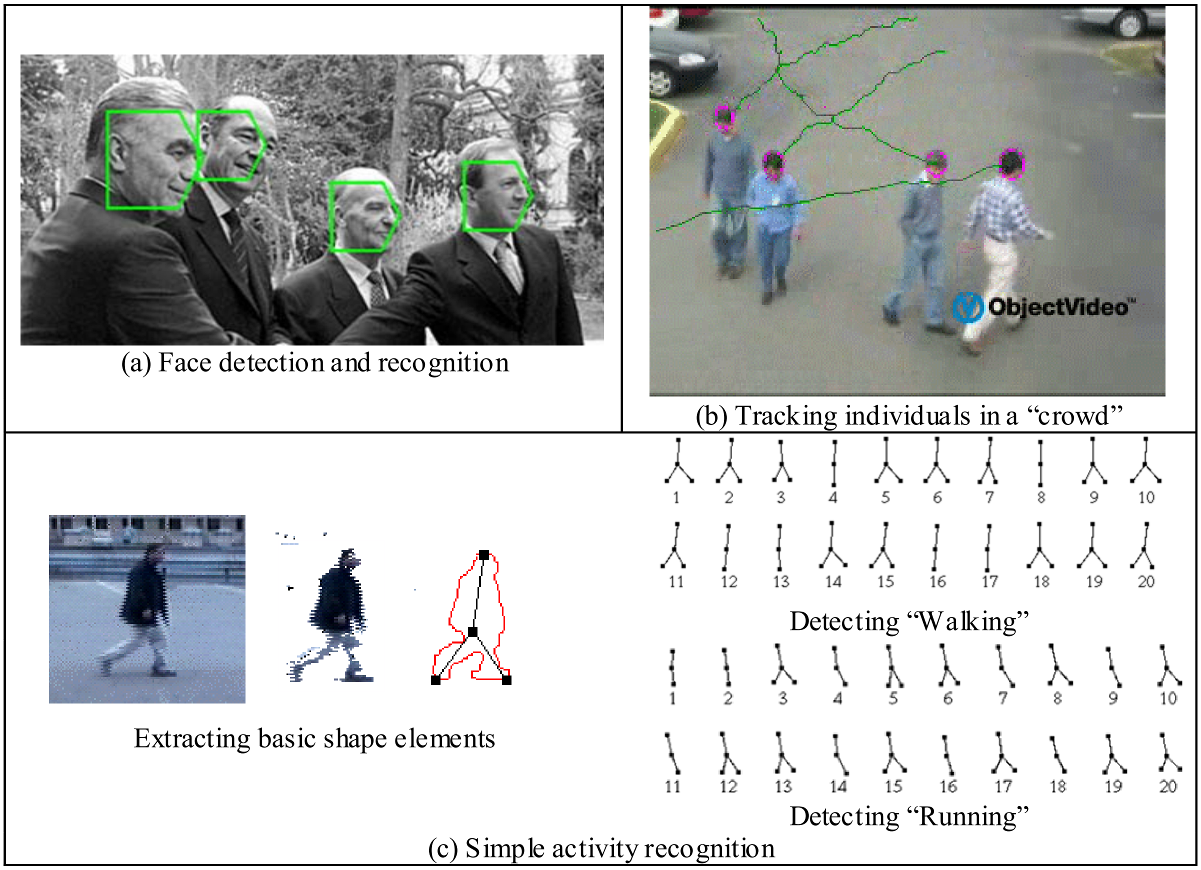
There are, however, just as many pressing applications for video surveillance automation in crowded scenes as there are in relatively isolated ones. Consider the relatively difficult task of maintaining public safety in an airline terminal or on a crowded railway platform. In such applications, the threat may be as subtle as someone reaching into an overcoat pocket for a concealed weapon. At this point, the state of the art in intelligent video surveillance is in its infancy with respect to understanding the actions of individuals in crowds. However, once the individual has reached for the weapon, there is a chain of events leading to some very simple gross crowd motion – in general, there will be disturbance to the pattern of crowd motion, and people will start moving away from the source of the threat. Using this crowd-motion analysis concept, companies like ObjectVideo are producing products like VEW that are starting to address crowd surveillance problems and provide security professionals with tools never before conceived of to help address security applications.

3. Two Technology Directions – Detecting Individuals and Understanding Crowds

Ultimately, solutions to the problems of understanding complex actions in crowded scenes will have to rely on two, somewhat different, computer vision technologies. The first is the detection of an individual in a crowd and thence the understanding of individual action and motion. An excellent example of this is the aforementioned individual reaching into a pocket for a weapon. The second technology is the detection of "crowd behavior". This is a statistical understanding of crowd action and the ability to detect activities through observing motion flow, speed, and direction of elements of the crowd itself. A good example of this might be the detection of a fight in a crowd at a ball-game. Without understanding individual human motion, it might be enough to detect that there is a disturbance in the crowd; a break in the regular motion pattern. It might, then, be enough to track that disturbance over time to isolate the epicenter of the disturbance – that is, "who started it!"

Detecting individuals in crowds

Currently, research in this area of computer vision is focused on two fronts: detecting an individual (usually, in fact, an individual face) for the purpose of identifying that person by some biometric technique such as face recognition [11] and understanding individual actions for the purpose of detecting some threatening behavior (see Figure 5, next page)**.** At present, both technologies are in their infancy. Simple methods for finding individuals in crowds focus on skin-tone detection, face detection [12] or human model detection. These techniques are adequate for detecting people in crowds of up to tens of individuals – but not really large crowds. Human activity detection is becoming good at identifying simple actions such a person loitering or walking into a restricted area, but only in research labs do algorithms exist for detecting people walking/running, throwing, punching, etc. [13]. Clearly, for today's security needs, these techniques are not ready for prime time – and more generic crowd-based surveillance approaches are required.


(a) Face detection and recognition

(b) Tracking individuals in a "crowd"

Extracting basic shape elements

Detecting "Walking"

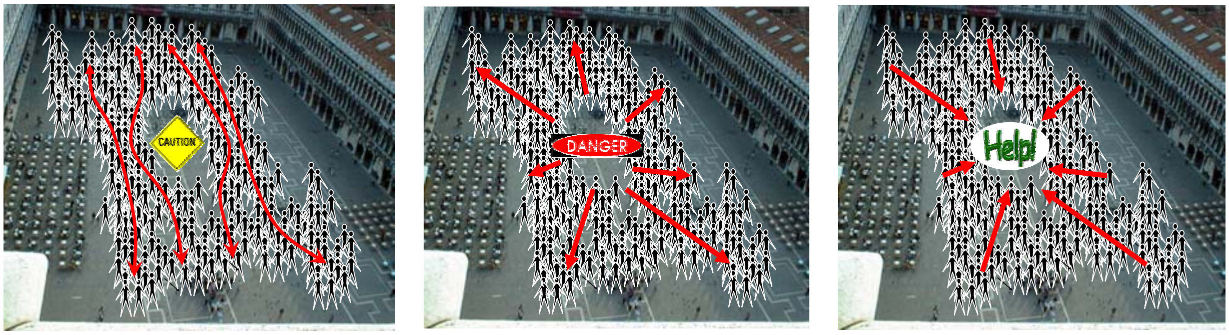
Detecting "Running"

(c) Simple activity recognition

Figure 5 - State of the art in human activity recognition.

Understanding Crowd Motion – Flow Monitoring

Even without getting down to the individual level of knowing which person precipitated a specific action within a crowd, a great deal of important security and public safety information can be gleaned by observing crowd flow in general. For example, as mentioned above, it may be possible to detect that a disturbance in the typical pattern of crowd motion indicates a fight has broken out amongst spectators at a sporting event. Unusual crowd flow may also indicate a threatening event. A crowd moving too fast may indicate that something dangerous is happening. A crowd avoiding a particular area may indicate something unsafe there. A crowd moving away from an epicenter may indicate a threat at that point. Figure 6, next page, illustrates several examples of this.



(a) Something is unusual *(b) Something is dangerous* *(c) Something is attracting a crowd*

Figure 6 - Extracting security information from observing crowd motion

To understand these events, it is only necessary to be able to monitor the flow-field of a crowd in a video scene. The locations and actions of individuals do not need to be identified – just the elements of the crowd as a whole. Fortunately, the technology of computer vision includes in its arsenal a technique called ***optical flow*** [14] which measures the motion of pixels in a video scene. Using this technique it is possible to tell at any given time, which elements of the crowd in a particular scene are moving in which directions. Unfortunately, optical flow algorithms are generally quite computationally expensive. So, companies like ObjectVideo have created high-speed versions of these techniques specifically for real-time intelligent video surveillance algorithms

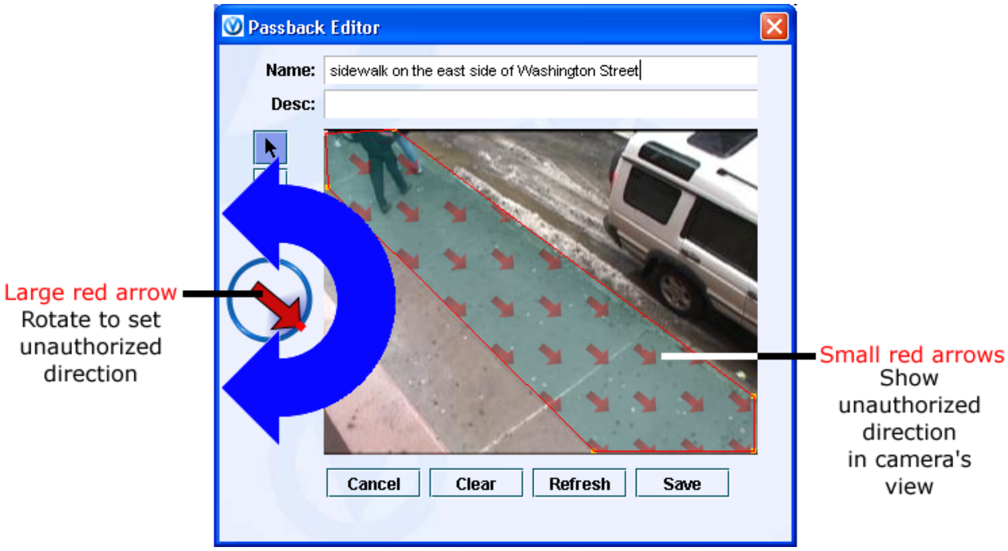
4. Example – Counter Flow Detection

One particular application where understanding crowd flow is important is counter flow detection. The idea is that in a particular area, monitored by a video camera, regulations require that people or objects move in a prescribed direction. The IVS system watches the flow of the crowd through that area and detects when objects move against the flow of traffic in the wrong direction – not by understanding the motion of individuals in the scene, but by monitoring the aggregate flow of the crowd itself. Applications for this functionality are legion – from public safety (people moving the wrong way along people-movers or escalators) to traffic monitoring (cars going the wrong way down one-way streets or lanes) to airport security. Let's consider airport security.

Recent changes made by the Transportation Security Agency (TSA) at US airports have made the gate areas more secure by enforcing additional security checks upon all passengers moving from main terminal areas to transit gate areas (called sterile areas). All airports enforce this screening on all passengers moving in to the sterile areas, but there is no special security screening applied to passengers leaving sterile areas. The TSA has only mandated that all exits from sterile areas be physically monitored by TSA personnel. If someone or something manages to enter through the exit portal and that object is not intercepted in time, dire consequences result. The entire terminal must be shut down and evacuated – this results in airline delays and airport down-time that can costs millions of dollars as well as a great deal of inconvenience to passengers and staff members alike.

This happens more commonly than is appreciated. It is simply too easy for a human agent to become distracted and miss someone going the wrong way or for the crowd to become simply too large for effective monitoring by a human agent.

Interestingly, most of these exit portals are already observed by a video surveillance system. So an application exists for an intelligent video surveillance system to monitor the exit portals of sterile areas at airport terminals for people or objects moving against the flow of traffic. Here, the goal is to observe the crowded exit portal as people stream through after disembarking from ‘planes. If anyone or anything moves against the accepted flow direction, an alert can be generated in real-time to allow TSA or other security personnel to intervene and interdict the perpetrator before the sterile area is breached.



Name: sidewalk on the east side of Washington Street

Desc:

Passback Editor
Cancel
Clear
Refresh
Save
Large red arrow
Rotate to set unauthorized direction
Small red arrows
Show unauthorized direction in camera's view

Figure 7 - Defining a rule for counter-flow detection

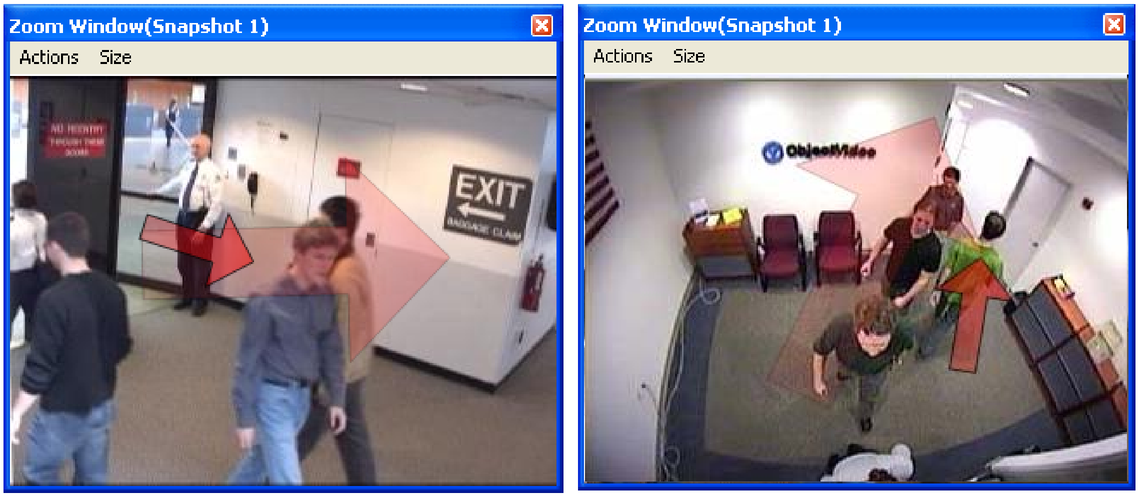


Figure 8 - Alerts generated by automatic counter-flow detection

This application exists as an add-on to ObjectVideo's commercial intelligent video surveillance product VEW. An operator can configure the system so that it knows the acceptable direction of crowd flow through the exit – as viewed by a surveillance video camera (see Figure 7). If any motion flow is detected in the direction counter to the acceptable direction, an alarm is generated highlighting the area where the counter-flow occurred (see Figure 8). Here, it is not necessary for the computer system to be able to identify an individual in a crowd, or what actions that individual is undertaking. The system highlights the area and the determination of what is happening is left to a human operator. This system works not by replacing a human, but by providing a tool to make the human more effective.

5. Conclusions

Intelligent Video Surveillance (IVS) systems are going to be an essential component providing the necessary levels of vigilance required to protect both populations and critical infrastructure from external (and internal) threats. IVS systems provide a scalable means for turning passive video surveillance systems into proactive defense sensors – by adding a robot pair of eyes to every surveillance video stream. Unfortunately, the state of the art in computer vision technology is in its infancy with respect to monitoring actions within crowded scenes. There are two explicit technology paths for dealing with these situations: extracting and recognizing the actions of individuals within crowded scenes; and understanding the gross actions of crowds themselves.

Whilst great strides are being made in the advancement of science to detect and understand the actions of individuals within crowds, there is good mileage to be made from understanding the patterns of motion of a crowd itself. From the simple expedient of being able to monitor the flow of motion within a video scene, it is conceivable that it will be possible to determine when threats and disturbances are occurring around gatherings of people such as sporting events and public transportation areas. A specific example of this technology is demonstrated by ObjectVideo's VEW product which includes the application of detecting counter-flow in a crowd; not by measuring the motion of each individual within the crowd, but by observing the flow of the crowd itself. This technology may justly be used in airports to protect sterile areas from wrong-way movement of people or objects. It can also be used to ensure public safety by detecting wrong-way traffic up or down escalators in public places or to monitor vehicle traffic on public roads. Many other applications undoubtedly exist.

6. References

[1] www.cs.cmu.edu/~VSAM

[2] H.H. Nagel, “Formation of an Object Concept by Analysis of Systematic Time Variations in the Optically Perceptible Environment”, Computer(14), No. 8, Aug. 1978, pp. 29-39

[3] A.J. Lipton, M. Allmen, N. Haering, W. Severson, T.M. Strat, “Video Segmentation Using Statistical Pixel Modeling”, US Patent #20020159634 Pending

[4] I.Haritaoglu, D. Harwood, and L.S. Davis, “W4s: A Real-Time System for Detecting and Tracking People in 2 ½ D”, in 5th European Conference on Computer Vision, 1998, Freiburg, Germany: Springer Verlag.

[5] C. Wren, A. Azarbayejani, T. Darrel, and A. Pentland, “Pfinder: Real-time tracking of the human body”, in IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): pp. 780-785.

[6] M.A. Isard and A. Blake, “ICondensation: Unifying low-level and high-level tracking in stochastic framework”, in Proceedings of the 5th European Conference on Computer Vision, 1998, pp. 893-908.

[7] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, “Wallflower: Principles and Practice of Background Maintenance”, in Proceedings of the International Conference on Computer Vision, 1999, pp. 255-261.

[8] A. Lipton, H. Fujiyoshi, and R. Patel, “Moving Target Detection and Classification from Real-Time Video”, in Proceedings of the IEEE Workshop on Applications of Computer Vision, 1998.

[9] M. Onoe, N. Hamano, K. Ohba, “Computer Analysis of Traffic Flow Observed by Subtractive Television”, in the Journal of Computer Graphics and Image Processing(2), 1973, pp. 377-392.

[10] R.O. Duda, P.E. Hart, D.G. Stork, “Pattern Classification”, 2nd Edition, 2000, published by John Wiley and Sons.

[11] D. McCullagh, “Call it Super Bowl Face Scan I”, Wired, Feb 2, 2001 (http://www.wired.com/news/politics/0,1283,41571,00.html)

[12] H. Schneiderman, T. Kanade. "Probabilistic Modeling of Local Appearance and Spatial Relationships for Object Recognition." IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 45-51. 1998. Santa Barbara, CA

[13] H. Fujiyoshi and A. Lipton, "Real-time human motion analysis by image skeletonization", in *IEEE WACV*, 1998

[14] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *Proc. 7th Int. Joint Conf. on Art. Intell.*, 1981.