2015

Safe City Standoff Video Analytics Based Biometric Technologies

Homeland Security Research Corp.

Safe City Standoff Video Analytics Based Biometric Technologies

August 2015

Homeland Security Research Corp. (HSRC) *is an international market and technology research firm specializing in the Homeland Security (HLS) & Public Safety (PS) Industry. HSRC provides premium market reports on present and emerging technologies and industry expertise, enabling global clients to gain time-critical insight into business opportunities. HSRC's clients include U.S. Congress, DHS, U.S. Army, U.S. Navy, NATO, DOD, DOT, GAO, and EU, among others; as well as HLS & PS government agencies in Japan, Korea, Taiwan, Israel, Canada, UK, Germany, Australia, Sweden, Finland, Singapore. With over 750 private sector clients (72% repeat customers), including major defense and security contractors, and Fortune 500 companies. HSRC earned the reputation as the industry's Gold Standard for HLS & PS market reports.*

Washington D.C. 20004, 601 Pennsylvania Ave., NW Suite 900,
Tel: 202-455-0966, info@hsrc.biz, www.homelandsecurityresearch.com

Table of Contents

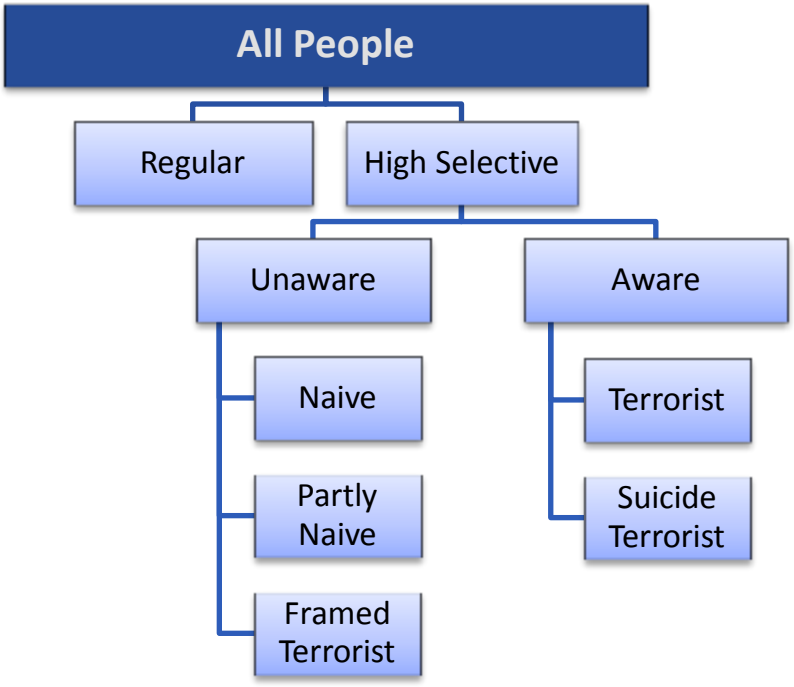
List of Figures

1 Standoff Video Analytics Based Biometrics

1.1 Introduction

Screened people profiling and behavioral tracking are part of the tools used in an attempt to significantly reduce security threats (and problems such as drug trafficking) without inconveniencing individuals being screened en mass and consequently causing delays that increase costs borne by checkpoint operators.

Figure 1 - Profiling and Behavior Tracking – Principles of Operation



Profiling and behavior tracking is no longer a single activity, low tech effort. It has evolved over the years (since 1994) to include a variety of mitigating activities including the following:

- **Computer-Assisted People Pre-Screening (CAPPS) –** This requires screened people to provide personal information when they make reservations. CAPPS and CAPPSII were programs that never took off.
- **Screening Screened People by Observation Techniques (SPOT) –** This is based on behavioral pattern recognition. In this method, security officers use a technique to flag screened people who appear to be acting suspiciously. While it is behavior and not race or ethnicity that is being tagged, there are concerns that behavioral pattern recognition can turn into racial profiling or subject innocent people to illegal searches without sufficient cause.

- **Electronic Behavior Tracking –** This method uses a combination of sensors such as infrared cameras, biometrics, together with computer-based software designed to detect and examine people whose behavior is atypical. These techniques, most of which are now in the development and testing phase, use "triggers" to catalyze a response in people and then follow up on the manifestations of these "triggers".
- Profiling and behavior tracking have three goals:
  - Locate and possibly mitigate the threat of individuals who might pose a danger to facilities, other people or to populations (e.g., airplane hijackers, terrorists, criminals and anti-government activists).
  - Identify and neutralize individuals who might be carrying contraband.
  - Supplement the far-from-perfect capabilities of today's functioning screening technologies.
- There are two specific modes of operation:
  - **Active**: In this mode, the profiler screens people and actively engages those who appear to attract attention. The engagement should be circumspect, friendly and "with a smile". But the purpose will be to collect additional information about the screened person and within ten seconds make a decision as to the need to further inquire into certain aspects of the screened people's behavior.
  - **Passive**: In this mode, the screeners remotely monitor the behavior of multiple screened subjects.
- The main trigger for further inquiry into a person's intentions is an impression (by the profiler) that the screening subject is either lying or that he/she is hiding something about the purpose, itinerary or any other element relating to the trip. Screeners are trained to observe micro-facial expressions and to make split second decisions about the need for additional impressions.
- The responses are not really as important as the way they are delivered. When the stress normally associated with conversing with a security official is discounted, innocent people will behave much more openly than someone who has something to hide. Demonstrated stress, fear and/or anxiety are significant indicators that more in-depth questioning may be warranted.
- Similarly, the passive profile looks for such indicators as presented above, to become much more pronounced when a potential perpetrator approaches a security portal or a checkpoint
- Passive remote behavior tracking is defined as behavioral pattern

recognition (without stimuli) so that anomalies are automatically flagged. It includes:

- Computer recognition of "normal" human behavior patterns from various sensor feeds
- Recognizing behavioral anomalies and variations
- Models of behavior building the database
- CCTV Systems that deliver a composite picture – giving operators situational awareness so that incidents can be dealt with proactively rather than reactively
- Linkages to tracking software and anomalous event/behaviour detection
- Automatic alarming
- Predictive modelling – human behavior
- Need for highly accurate map overlays and predictive software for the tracking of key devices and people.
- Interface with various devices or packages of software that can provide mapping quickly
- Ability to both track and predict paths based on behavior (real-time)
- Needs to be combined with 3D rendered models, where necessary.

While CCTV cameras are inexpensive and used extensively at security sensitive locations, the work force required to monitor and analyze them is expensive. Moreover, differentiating between normal human activity and aberrant behavior via CCTV cameras is difficult. Vigilance must remain constant, yet attention reduces dramatically over time. Consequently, the videos from these cameras are usually monitored randomly or not at all; they are often used merely as archival material for future reference. CCTV can be far more useful, detecting scenarios and taking action in real time. Video surveillance is a topic of considerable interest for computer vision and homeland security applications. The framework of a video surveillance system includes the following stages:

- Modeling of environments
- Detection of motion
- Classification of moving objects
- Tracking, behavior understanding and description, and
- Fusion of information from multiple cameras and sensors

Despite recent progress, there are still major technical challenges to be overcome before reliable, automated, real-time video remote behavioral

surveillance can be realized. Low level intent detection which can be used to provide an initial evaluation of actionable hostile behaviors will rely upon tracking software that will function in an environment having variable conditions. A resulting confidence will be computed to provide useful monitoring benchmarks of human actions and may provide an early warning mechanism for terrorism-related activities.

1.1.1 Video Surveillance Based Behavioral Profiling

This type of profiling relies on the observer's skill in detecting telltale indications that something out of the ordinary is taking place. Such indications may be avoidance of eye contact, sweat, general uneasiness, shifting gaze, easily aroused nervousness and defensive responses.

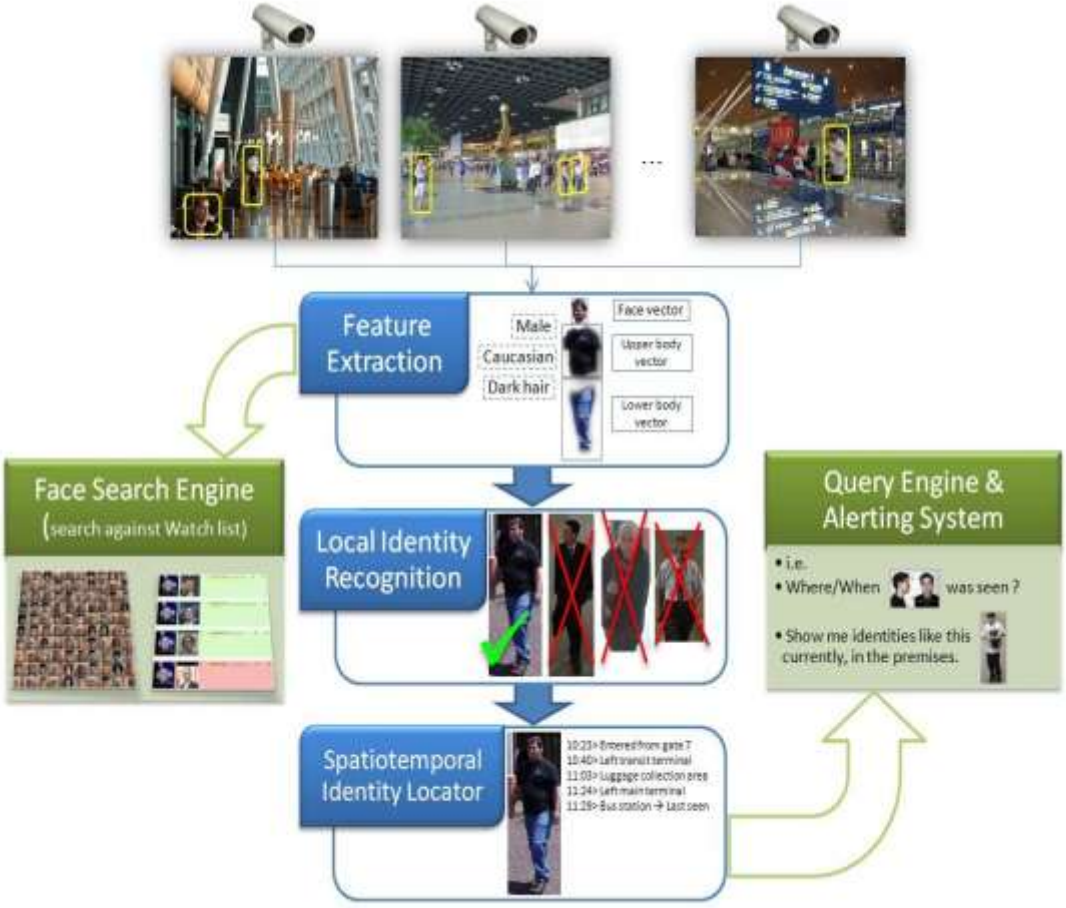
Of course, considering the fact that many of the stress indicators manifest themselves naturally in people who are not used to travel under stress etc., the possibility of false positives is quite high. The only remedy for this is an observer skilled enough to not only reading the indications, but to gently tease out the reasons for the indicators.

Since this type of profiling uses behavior rather than ethnicity or religion, it stands a better chance of avoiding deliberate or “blind” discrimination of one group or another. Further, it stands a better chance of selecting a perpetrator rather than an innocent person. This method relies heavily on a screener’s skills and alertness and will require talented personnel to properly perform their task.

Experience and history show that behavioral profiling is much more substantial than ethnic profiling. Considering the adaptability of terrorists, it is unlikely to assume that they will not resort to using “Western” looking colleagues (witting and unwitting) to carry out attacks particularly on the aviation industry.

Biometrics is a term that applies to the many ways in which human beings can be identified by unique aspects of the body. Fingerprints are the most commonly known biometric identifier. Other biometric identifiers include hand prints, vein dimensions, iris (eye) designs, the pattern of blood vessels in the retina, body odors, characteristic and unique movements, individual voices, and of course, DNA. Countries around the world are implementing biometric monitoring procedures. For example, Brazil has begun a national fingerprint system to track recipients of unemployment benefits and healthcare entitlements.

Figure 2 - Standoff Biometric People Screening Using Biometric Intelligent Video Surveillance



(**Source:** http://www.nicta.com.au)

International usage and interest in surveillance of public spaces is growing at an unprecedented pace in response to crime and global terrorism. However, whilst it is relatively easy, albeit expensive to install increasing numbers of cameras, it is quite another issue to adequately monitor the video feeds with security personnel.

A pressing need is emerging to monitor all surveillance cameras in an attempt to detect events and persons-of-interest. The problem is that human monitoring requires a large number of personnel, resulting in high ongoing costs and questionable reliability as the attention span of humans decreases rapidly when performing such mundane tasks. A solution is found in advanced computer surveillance systems to monitor all video feeds and deliver alerts to human responders for triage -- a well-designed computer system is never caught "off guard".

A fast and robust CCTV-based face matching and search framework can be used to detect persons of interest. Unlike many existing solutions, such systems are designed to work with low-resolution images of 'uncooperative' subjects (people who are not posing for the camera). It is robust to variations in environmental conditions, quality and pose. Additionally, it has a comparatively small computational footprint and can be parallelized, thus making it easily scalable.

As it is often difficult to acquire good face images, this technology is being combined with further research in person tracking so that a better estimate of identity can be inferred through combining several CCTV-based biometric techniques over time.

Who Will Benefit? This tool can assist in forensic examination of video after an incident or potentially proactively alert the authorities at the outset. The techniques can also be applied outside of security and surveillance such as facial recognition for photo-tagging in consumer products.

The technical definition of a biometric is “any measurable, robust, distinctive, physical characteristic or personal trait of an individual that can be used to identify or verify the claimed identity of that individual.” Every biometric system contains three components:

- Enrollment – the process of collecting biometric samples.
- Templates – the data that represents the enrollee’s biometric located in a database.
- Matching – the process of comparing a submitted sample against one (verifying) or many (identifying) templates in the database.

1.1.2 Video Analytics Based Biometric Face Recognition Identification vs. Verification

Biometric identification or verification systems are distinct from each other. Facial recognition identification systems are being combined with video monitoring to identify suspected terrorists in airports and at border crossings. Combined biometric verification systems and videos are used to control access to computers, secured areas and to verify passport information or citizenship status. When biometric face recognition technology is used to identify an individual, the system attempts to answer the question “Who is John Doe?” by reading the information or sample provided and comparing it to many templates in the database. It then reports or estimates the person’s identity. When the technology is asked to verify someone, the system is asked “Is this John Doe?” The system compares the biometric information presented to the template in the database identified as John Doe and either accepts or rejects the claim.

Templates in a biometric face recognition database typically are composed of complex, programmed knowledge rules, statistical decision rules, neural networks and software.

This means that the database is built using certain assumptions that induce a potential for errors. In other words, biometric face recognition database templates do not contain exact likeness of individuals but rather complex statistical and mathematical estimates of digitized images.

1.1.3 Video Based Biometric Recognition Technologies

- Since identification and verification systems are different, so too are the performance procedures and protocols used to evaluate the efficacy of each type of system. For identification systems, the principal measure "equals the rate of queries in which the correct answer can be found in the top few matches." In other words, the higher the score of a correct match contained within the top matches, the better the system.
- Thus, if used to attempt to capture terrorists in public places, the data input would be an image captured on video and the output from the database would be a list of top matches. The sheriff or airport security guard would make a subjective decision to further search or detain the individual.
- Two error statistics, false-reject rate and false-alarm rate are used to measure the ability of a verification system. "A false reject occurs when a system rejects a valid identity (i.e., the real Michelle Kwan is denied access to the Olympic skating rink); a false alarm occurs when a system incorrectly accepts an identity."
- In general, experts and researchers report that face recognition software are sensitive to changes such as shifting sunlight during the day and changes in facial positions. A system's performance will drop significantly if the softwares are not corrected to address lighting variations and moving faces.
- The recent use of biometric face recognition technology at the Super Bowl is a good example of law enforcement's use of these emerging technologies and the debate over potential misuse. The faces of over 100,000 fans entering the stadium to watch the Super Bowl were recorded by the local law enforcement on video cameras. The facial images were then digitized by sophisticated software and checked electronically against a watch list database. Fans were not aware that this had occurred until it was reported in the media. Law enforcement officials maintained that they were using the latest available security tool and that it was no more intrusive than a video camera in a convenience store.

- The narrow accuracy range of the technology also raises concerns about false identification. A recent study by the National Institute of Standards and Technology found that when digital photos of the same person taken 18 months apart were compared, they triggered a false rejection by computers 43% of the time. With such a large potential error, the law enforcement relying solely on these technologies to identify individuals might often stop and question an innocent person instead of a possible terror suspect.

1.2 Video Based Face Recognition

- Face recognition provides for a more accurate identification, but requires a face image with good resolution and in a proper position (facing forward as much as possible). Most systems will tolerate a face rotation. Many face encoding and comparison software have been proposed and are based primarily on the extraction of mathematical descriptors or topological points on the face.
- Initially developed based on photographs, face recognition applications have existed for a few years on video footage. Depending on the technique used, recognition can be achieved by changing appearance, for example the addition of glasses or a beard, facial expressions and illumination. Advanced 3D face recognition software appear to improve identification performance and robustness in comparison to 2D software. However, face recognition in an uncontrolled background, for example a crowd, continues to be a problem that has not yet been satisfactorily resolved by current video analytics systems.
- Finally, commercial facial recognition software solutions are available to satisfy and meet a wide range of requirements and applications; “from gender and age range identification for real-time retail and marketing activities, to biometric facial recognition technologies able to capture and compare” real time video stream images for applications in airports, large critical power and energy plants, oil refineries and law enforcement government buildings. In this framework, filtering and de-noising techniques are systematically implemented in embedded hardware subsystems for real–time or near real-time surveillance applications using embedded Fast Signal Processing techniques.
- From a forensic perspective, the technology of video surveillance system is handled by other units. Modern facial portrait recognition systems are widely available to law enforcement agencies in the commercial market. The key question is the legal and ethical use of biometric databases. When a specific crime has been committed, use of the facial portrait data is very similar to utilization of other biometric information. The problems of privacy intrusion and authorized use arise when searches are made to

determine who was in a geographic area (tracking) at a particular time without him being considered a suspect involved in a specific crime.

1.2.1 Remote Biometric Identification Technologies

There has been considerable work done in remote face-recognition and other biometrics including for example, a DARPA program in human identification at a distance. The program has evaluated face recognition under extreme lighting and at over 50m, gait recognition at similar distances and iris recognition at 2-3 meters. However, these results were obtained under controlled daylight conditions with cooperative subjects.

Achieving 24/7 face and iris recognition requires significant improvements in the operational field-of-view to account for uncooperative subjects, handling wide ranges of poses and illumination, advances in stabilization, handling of motion blur, and addressing atmospheric disturbances.

To summarize, the overall goal of most RDT&E programs in this field is to develop a system that can provide personal biometric recognition with high accuracy in 24/7 operations. In addition, the systems will need to support several operations:

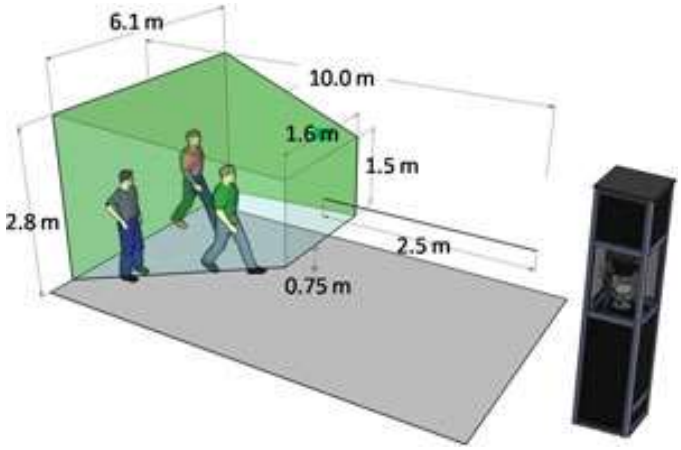
- High volume, low security identification quickly scanning the area of interest
- High security, low volume identification in a smaller area.
- Authenticate an identity from a distance without slowing people down as they go about their business.
- Covertly capture biometrics and track through an infrastructure
- Find a face/person in a crowd

Recognition of individuals at a distance is a challenging problem that requires advancement in a variety of biometric technologies. Some researchers believe that iris recognition is the most promising technology and in combination with other biometrics such as facial recognition and situational awareness analytics of video scene images, promises to deliver the highest performance remote identification. Multi-modal systems that definitively identify individuals within randomly moving crowds of people at variable distances are the key.
(**Sources:** US Congress, DHS, CBT, FBI, HSRC)

1.2.2 Fused Intelligent Video Surveillance & Watch Lists

Pre-screening citizens and checking their personal information against watch lists is a controversial subject. Issues of privacy as well as erroneous diversion of flights and the misidentification of passengers have raised concerns about the practice. Additionally, system failures have also taken the practice into question.

Figure 3 - Field of View of a Fused Video Surveillance and Biometrics



(**Source:** Retica Systems, Inc.)

Systems like the Eagle-Eyes by Retica couple video surveillance software with a biometric acquisition system. Eagle-Eyes is a long-range multi-biometric system that improves on existing iris acquisition approaches in terms of stand-off distance, capture volume and subject motion. It is capable of acquiring face and iris images from multiple subjects present within a scene. The system uses multiple cameras with hierarchically-ordered fields of views, a highly precise pan-tilt unit (PTU) and a long focal length zoom lens.

System specification includes:

- Day/night operation.
- Screened population: Cooperative/non-cooperative operation.
- Real-time iris & face capture.
- Identifying and tracking up to 40 subjects per minute.

1.2.3 Crowd and Riot Surveillance

- A domain of video analytics development is crowd and riot analysis and monitoring. With the quick escalation of the global population, the densification of large urban centers, and the growing issue of providing security during large gatherings, Intelligent Video Surveillance has become an interesting avenue. Video cameras are actually already set up to monitor large events (e.g., Public events, conventions), but the analytical efficiency of video systems in this context remains debatable.
- The different video analytics steps discussed in the previous sections are applied to it. However, analyzing a crowd has its own set of clear complications because crowds are made up of numerous individuals.

- Estimating density is a fundamental step in crowd monitoring and management requiring an approximate assessment of the number of individuals. The numerous obstructions and juxtapositions make pedestrian segmentation for counting applications difficult. Certain software assumes a proportional relationship between the number of pixels associated with the foreground following segmentation and the number of people. Other methods are based on image texture in order to characterize crowd density. Certain software detect head contour, while others use the histograms of several signatures to infer the approximate number of people.
- Although item tracking is one of the most popular topics in video surveillance, most of the software developed applies to a small number of people. Tracking pedestrians in crowds presents key difficulties, especially with respect to the large number of individuals to be followed from one frame to the next and the numerous obstructions present. Certain methods track key features which are less subject to being disturbed by obstructions than contours. Others model the human body or its parts. Probabilistic matching methods or clear filters are used to follow these models during video footage. A body of research exists on tracking using several video cameras: M2 Tracker works with synchronized video cameras, while others use a combination of fixed and PTZ video cameras.
- It is important to follow a crowd's movements for security applications. For example, the crowd's trajectory and flow will be analyzed. When modeling the crowd and its behaviors, certain researchers consider the crowd as a whole and interpret the movements of the different parts. Software such as optical flow and hidden Markov models are used to model movements. Certain models will combine individual and crowd analysis. In this context, two types of approaches are proposed to describe a crowd's activity: the application of physical models, for example the kinematics of gas particles and fluid dynamics and agent-based software.

1.2.4 Wireless Video Analytics

(**Source**: Eeneywireless)

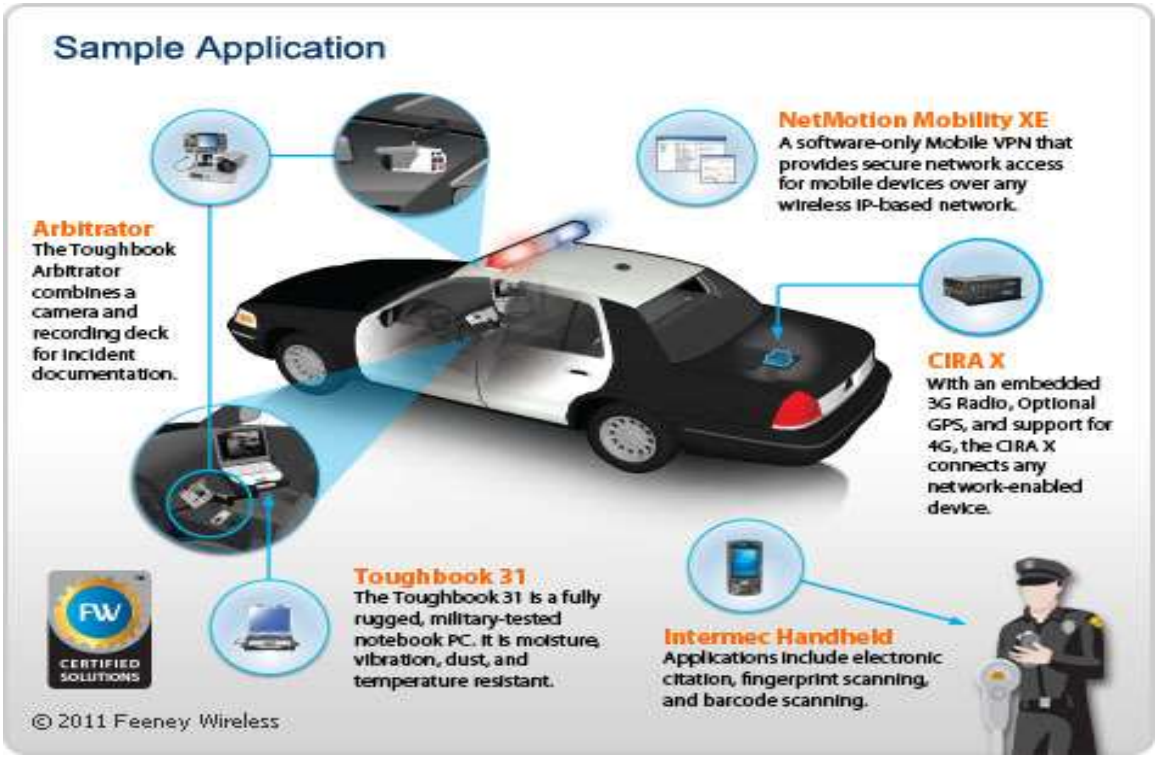
Wireless Video enable users to remotely view and record live video from anywhere in the world. They use standard IP networks such as local area networks (LANs) and the internet for transporting information rather than dedicated point-to-point cabling such as that used in analog video systems.

This ensures a cost-effective, flexible and scalable video surveillance solution that can be easily expanded as needs evolve. Wireless Video enables:

- Flexible and adaptable computing platform for video surveillance systems which connect video and sensing devices over wireless networks.

- Low-bandwidth/low-power consumption for transmitting high-resolution imaging and data to the control room, desktop, laptop, cell phone, PDA, and storage media.
- Integrates/expands legacy analog systems with next generation IP surveillance/security applications.
- Ability to add new camera to a network and control it on the Web.

Figure 4 - Police Application of Intelligent Video Surveillance



1.2.5 Cloud Video Analytics

Cloud infrastructure as a service is an important deployment model for modern Intelligent Video Surveillance application development as it forces clean, modular designs to take advantage of the scaling opportunities available, and as a consequence creates more testable and robust software. The scalability potentials available with the cloud model makes it particularly attractive to data and compute intensive projects, either research or commercial applications, particularly those which are long term and may require scalability over time.

The primary advantages of Cloud Video analytics are:

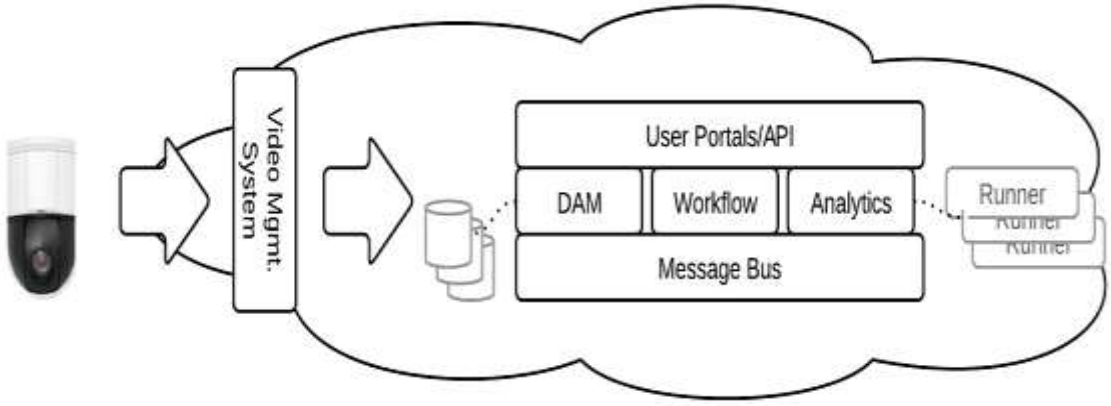
- Scalability – easy provisioning of newly computed and storage infrastructure;

- Cost structure only paying for resources used and less complex administration
- Maintenance, as there is no hardware to maintain.

The Core Platform identified storage and the execution of analytical applications as the major scaling points and kept the digital asset management, workflow management, and analytics managements as the main discrete components. The Video Management System (VMS) performs the recordings from the cameras into its own archive. Next, the recordings are extracted from the VMS by the Video Extraction and Processing (VEaP) component and ingested into the Digital Asset Management system (DAM).

The DAM notifies other components of the new data via a message bus, the workflow component acts on these messages and queues the analytics tasks. The analytics service schedules these tasks for execution on scalable analytic runner resources and when the tasks are completed, the results are ingested back into the DAM. User access and external system integration is achieved through the user portals and APIs.

Figure 5 - Cloud Based Video Analytics Architecture



1.2.6 Online Video Analytics

- Online video analytics (also known as web video analytics) is a way of measuring how viewers get to an online video and what they do when they watch it. A video is any length of security and commercial video stream such as a VTR review, movie trailer, television show or full-length video. Web video analytics aim to answer such questions as:
  - How long did the viewer watch a particular video?
  - Did they pass it along to a friend? Embed it in their home page?
  - If there was a security related event with the video, did they watch it? If so, for how long?

- In the TV industry, online video analytics differs from traditional television analytics because it can be measured using census-based methods instead of panel-based metrics. Every action that a viewer does while watching a video online can be captured and analyzed precisely

1.2.7 Pulse Video Analytics

As video surveillance systems capture and share high amounts of video data, users need easy ways to find videos based on content and to search within videos to find relevant segments. The main tasks of Pulse Video Analytics are as follows:

- Automatically creates a vocabulary of relevant keywords for an organization by analyzing existing documents
- Identifies every place in each video where these keywords are spoken and where each speaker speaks
- Displays a list of speakers and keywords that appear beside each video: Viewers can go directly to relevant segments by selecting a speaker, keyword, or a keyword spoken by a particular speaker.
- Increases the value of the video by making it searchable
- Views hours of video to find a few seconds of relevant content

1.3 Smart Cameras

Smart cameras lack a clear definition. There are no standards established to define a “smart camera”. There is no regulating body or association defining capabilities. So, the name isn’t 100 percent clear and hence has become a generic term. In the same manner, many vendors work hard to proclaim their camera as smart due to its ability to do “something” beyond visual data capture. But the notion of what constitutes a smart camera is certainly not immediately identifiable. One can generally conclude that smart cameras contain some kind of video analytics processing capability along with visual acquisition.

While analog cameras are still used in many embedded vision systems, the Embedded Vision technology definition primarily focuses on digital image sensors—usually either a CCD or CMOS sensor array that operates with visible light. However, this definition shouldn’t constrain the technology analysis since many machine vision systems can also sense other types of energy (IR, sonar, etc.).

The camera housing has become the entire chassis for an embedded vision system leading to the emergence of “smart cameras” with all of the electronics integrated. By most definitions, a smart camera includes computer vision, since the camera is capable of extracting application-specific information. As both

wired and wireless networks get faster and cheaper, there still may be reasons to transmit pixel data to a central location for storage or extra processing.

A classic example is cloud computing by using the camera on a smartphone. The smartphone could be considered a “smart camera” as well, but sending data to a cloud computer may reduce the processing performance required on the mobile device (lowering cost, power, weight, etc.). For a dedicated smart camera, some vendors have created chips that include all of the smart camera features. An example is the Cognivue CV220X which the company calls an “image cognition processor” (ICP). The device stacks up to 16 megabytes of DRAM in the same package as the processing chip which integrates an ARM CPU with an array processor to accelerate computer vision algorithms.

Before Microsoft's Kinect, many people would imagine a camera for computer vision as the outdoor security camera shown in this picture. There are countless vendors supplying these products and many more supplying indoor cameras for industrial applications. It would be easy to ignore simple USB cameras for PCs, since these are arguably not embedded systems. However, that still leaves almost a billion cameras used for the embedded system in the mobile phones of the world. These cameras can’t be ignored, since the speed and quality have risen dramatically—supporting over 10 mega-pixel sensors with sophisticated image processing hardware.

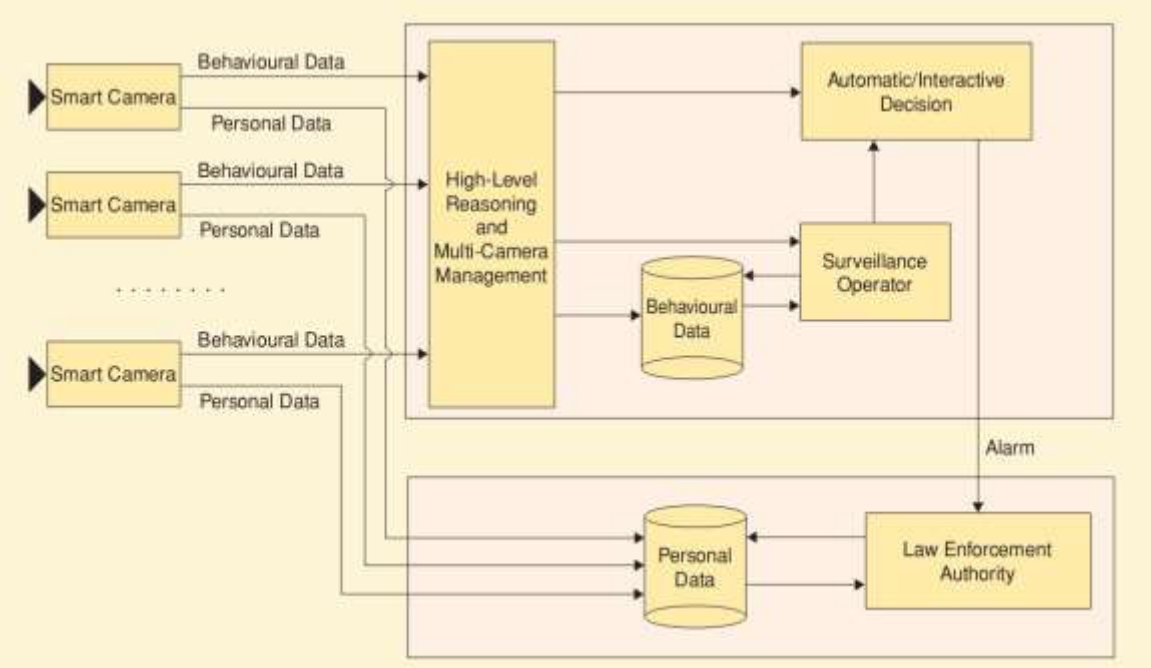
Considering another important factor for cameras—the rapid adoption of 3D imaging with stereo optics. In fact, cell phones now offer this technology. An example is the Sharp Aquos Smartphone incorporating 2 cameras and a pair of 8 megapixel sensors to create 720p 3D video. Look again at the picture of the outdoor camera and consider how much change is about to happen to computer vision markets as this new camera technology becomes pervasive.

Charge-coupled device (CCD) sensors have some advantages over CMOS image sensors mainly because the electronic shutter of CCDs traditionally offers better image quality with higher dynamic range and resolution. However, according to iSuppli, CMOS sensors now account for 90% of the market driven by the technology’s lower cost, better integration and speed. However, the cellphone business skews these market numbers and iSuppli predicts that 25% of machine vision applications still use CCD sensors[1].

CMOSIS is an example of a company offering CMOS sensors for embedded vision and their CMV4000 includes a global shutter and a 2K×2K CMOS sensor array that can deliver 180 frames/second with 10-bit pixels (12 bits at 37 frames/sec).

Smart cameras excel at addressing highly constrained, single-purpose detection needs. A key factor instrumental in the proliferation of smart cameras is that they address tasks that are generalized across physical sites such as counting, simple motion detection, license plate recognition, etc. These kinds of single function tasks have proven to be a sound fit for smart cameras.

Figure 6 - Block Diagram of a Video Surveillance System with Smart Cameras



(**Source:** A. Cavallaro)

Smart cameras are typically included in the high-end category because they often contain video analytic software. This typically drives up the price by another $300 to $2,500 per camera, making them in the higher cost range of above $1,200 and as high as $6,000. Intelligent or "smart" cameras as they are sometimes called, can be used in situations where it is cost-prohibitive to use bulkier DVR type devices; this is one of the key areas driving growth in this category. Another technology segment within the intelligent camera category is to allow the camera to be upgraded with third-party software analytics. This not only increases the overall cost of the camera, but also increases the usage scenarios by just changing the analytic software function. In the future, this could become one of the largest high-end camera segments.

What is clear is many companies that have implemented these solutions are now experiencing serious limitations as they have outgrown their smart camera capabilities. Customers are realizing that their need for faster, more accurate and intelligent detection is somewhat lacking these types of solutions.

Smart cameras are targeted for mass marketing and are typically sold through resellers, integrators or the internet, and that in itself has often become a frequent reason for failed implementations. Because smart cameras normally make up only a portion of the integrator's revenue, specific implementation typically is not treated as core to their business. They may even drop one smart camera line for another, leaving you without a path for change down the road. Even with a large integrator base, re-establishing relationships is never easy.

Keep in mind that if you require customization, sparing or growth – which is extremely common – ensuring that you have a long term reliable supply of product through your supplier is important.

Another drawback with smart cameras is their inability to keep pace with the processing and throughput requirements of security detection scenarios. As bad guys get smarter and come up with new ways to threaten facilities and assets, cameras need to become "smarter" by pulling in more data and applying more algorithms to that data. Unfortunately, an embedded system typically doesn't grow. The very objective of the embedded approach is to enable a set design that can be replicated over and over again to facilitate a small package at a low price. As a result, capacity for computational and input growth is typically a trade-off that is made early on in the design.

Although smart cameras continue to advance technically, they struggle to deal with highly complex performance challenges. Simple object detection is viable; however, most intrusion scenarios are not simple and when challenges require multi-faceted detection scenarios, smart cameras cease to be smart. The ability to embed intensive analytical capability on-board a camera with such features as classification, learning or training, and multiple types of detection is challenging. If you need highly reliable detection, high frame rate and/or large resolution, a smart camera might struggle to be successful. Intensive analytics require raw processing power which isn't possible in these camera form factors.

1.4 Smart Cameras Video Analytics vs. Centralized Video Analytics

The advantage of edge-based embedded analytics is bandwidth preservation and reduction of storage requirements because the need to transmit all captured video for analysis is eliminated. In addition, content analysis on the video of interest can be performed when the video is in its highest quality and before it is compressed for transmission over the network to a recording/storage device.

Cost performance developments in imaging technology such as digital signal processing (DSP) electronics, megapixel technology and increased sensitivity to low light have contributed to the increased adoption of smart camera- embedded video analytics. DSP chips deliver more processing capacity which enables more edge processing, while megapixel sensors offer greater resolution and detail which are necessary for premium image analytic algorithms. Smart cameras can be programmed to transmit at low-resolution rates until a pre-defined event of interest occurs at which point transmission switches to a higher frame rate and resolution. Improvements in compression technologies such as H.264 have also enabled transmission of higher definition resolution over a lower bandwidth.

Cost is another factor influencing the move to smart camera embedded analytics because it is a more effective way of implementing Intelligent Video Surveillance Single unit cameras featuring built-in analytics can be installed where and when

required or the number added to as required. And when more processing is done at the edge, it can help to reduce the cost of analytic processing by eliminating or downsizing server requirements.

Due to the large installed base of analog cameras, one should forecast that in the short term the biggest penetration of Intelligent Video Surveillance will occur within network video recorders (NVRs), digital video recorders (DVRs) and video encoders which convert analog signals to network compatible data. Relying solely on smart camera-embedded analytics is not as efficient when analytics must be performed or metadata correlated on enterprise-wide systems such as casino or educational installations where hundreds of cameras are deployed. Additionally, premium forensic searches (e.g. biometric face recognition, object identification) which are usually performed on recorded video are still beyond the capabilities of most smart cameras in the next few years.

In these instances, the greater processing power and central management capabilities of a server-based or head end solution is generally a better solution. Centralized analytics allows for the configuration of different application sets on different cameras and at differing times.

More information can be found at:
Global Safe City: Industry, Technologies & Market – 2015-2020