1

Information Management and Video Analytics: the Future of Intelligent Video Surveillance

Bennie Coetzer, Jaco van der Merwe and Bradley Josephs
Protoclea Advanced Image Engineering, South Africa

1. Introduction

The need to monitor exists for many reasons. We use it as a mechanism to protect ourselves and our property, we use it to manage large numbers such as traffic information, we use it to monitor behaviour as in crowd surveillance, we use it to monitor production lines and operations and so on. While recognition of events or alarms may assist in reacting to it, a major objective of systems should be to be proactive, in other words, to prevent events.

Video Surveillance has been with us for a long time. Traditionally it was used to display images on monitors, manned by guards or operators. This allowed us to view a number of places using less people and we could also perform patrolling duties from the safety of a control room. It satisfied the goals of safe patrolling and reducing manpower while performing the role of watchdog or guard. When video recording was introduced we found that we could create evidence of events that would be useful in prosecution, analysis, etc. As it became less expensive, more cameras were placed and of course more monitors. We could watch more areas with less people but very soon it became apparent that human beings have limitations. We also found that recording is an expensive exercise as video information is vast. At this point machine intelligence was introduced to assist with detection and also to reduce recording to be event driven, which of course made it less expensive. Initial techniques were crude with many false alarms but image analysis grew and became more sophisticated, resulting in better detection and even object recognition. Image quality improved, storage cost went down, less compression could be used and overall efficiency in terms of human intervention and prosecution success improved.

This article will limit its scope to Information Management as it pertains to the security and traffic arenas, although much of this is clearly applicable in other areas as well. The chapter concentrates on the use of Video Analytics to achieve the various objectives as defined, but, as will be seen in the paragraph on Intelligent Information management, Video Analytics is merely a part of the complete system.

2. Video analytics

2.1 Basics of video analytics

The role of Video Analytics can be described in a number of ways and consist primarily of the following:

2.1.1 Video enhancement
In this role images are manipulated automatically or by user intervention to assist a human or machine to detect or identify objects better. The processes involved could range from noise reduction, image sharpening, edge detection or various others. Such functionality should be part of any system that uses humans to interpret images.

2.1.2 Video reconstruction
In many instances sensors deliver distorted images. This could be because of poor quality lenses, atmospheric distortion, reflections or moving (vibrating) cameras or subjects. Video reconstruction tools such as stabilisation, anti-blurring and so on could be used to reduce such noise to assist users (humans and machines) to 'see better. These techniques could also be applied during post-event analysis and could assist in reconstruction of distorted images due to tape stretching, very high compression ratios and so on.

2.1.3 Video analysis
Video analysis has primarily to do with intelligence extraction from the visual scene and the rest of this chapter will concentrate on this aspect. This is not to downplay the importance of the other aspects but merely to serve the title of the chapter which is about information management.

2.2 Event detection
The initial objective of current systems is to recognise an event. This could of course mean many things. It could be detection of movement or detection of presence or absence of an object.
The basic mechanism employed is difference detection between prior and current views.

2.2.1 Motion detection and tracking
We have heard a lot about motion detection in the past, but it was usually very expensive and the results were not very accurate. With some further developments in the processing capabilities of modern processors, more advanced techniques could be used to give better motion detection results. It is now not only possible to monitor regions of images, but the entire image on a per pixel basis, and since each pixel can be monitored, a tracking algorithm can be applied to mark and follow the moving pixels in an image. Moving pixels of objects are usually grouped together into what some in the image processing community refer to as blobs.
The basic principle behind the detection of motion is to simply detect changes/differences between consecutive frames in an image sequence. Many detection algorithms are based on the subtraction of a "learnt" background model/image from the current image and applying a threshold value to separate the apparent "moving" foreground from the "static" background. This process of separation is also known as segmentation.
The definition of "background" may differ greatly depending on the application; i.e. average human traffic at an airport might be considered as background if we were detecting unattended baggage, or periodic motion, like swaying trees on a windy morning versus completely still trees on a windless afternoon. This very difference in definition and the pure randomness found in the statistical data of video makes modelling an accurate background

very tricky. Therefore a good motion detection system should get this right in order to do accurate detection with as little false positives or negatives as possible.
We will briefly look at a few current theoretical and practical methods in use and some of their advantages and disadvantages.

Modelling the background using a codebook [1]

This method "learns" a background model by monitoring each pixel in a scene over a set period and listing all values "captured" during that time as a list of "background" pixel values. Each new video frame is compared to these values and all pixel values that fall outside the codebook values are considered to be foreground/movement.
With this method it is possible to "train" the system to recognise swaying trees, for instance, as moving background. This however requires the monitored scene to be free of "foreground" movement during the training process, which is usually difficult in real world applications. This method would however be well suited in an indoors environment where light changes are minimal and some movement, such as a moving fan or a moving escalator, is present.
Since calculation consists of only simple value comparisons after the codebook has been learnt, this method is computationally speaking, very inexpensive and therefore requires very little processing power.

Mixture of Gaussians Background Modelling [2]

This algorithm uses an adaptive background; this simply means that the background model is continuously updated to allow slow background changes (such as gradual light condition changes or slow moving shadows cast by the sun) to be factored in, by updating the background model gradually as the scene background changes. Each pixel in the background is modelled by a mixture of K Gaussian probability distributions (where K is a number between 3 and 5), each representing different colour occurrences. The background is then modelled as the top X highest probable colours. Probable colours are the ones which stay longer and more static. In order to keep this model "adapting" each new value is compared and matched to the existing model and if no match is found a new Gaussian component is added and reordered with the existing components to create an updated background model.
This method is very resilient to periodic motion such as swaying trees and gets "better" with age. It can also be tuned to be very sensitive to minute colour changes, thus allowing the algorithm to be used on other image types such as thermal imagers. Another addition to this algorithm is the ability to distinguish shadows cast by moving objects versus static objects by comparing both the chromatic and brightness differences of each new pixel and the current background model to some thresholds.
Even though this model does seem to work well for most indoors and outdoors cases it is still sometimes difficult to balance the predefined threshold perfectly. Clouds passing before the sun in an outdoors scene, does change the total brightness values of the scene, causing the sudden colour change to be detected as movement. This can however be "handled" by monitoring the total scene brightness and adjusting the model parameters accordingly.
In general this detection algorithm does really well when very accurate detection is necessary, and performs well in both colour and monochrome video. False positives (objects detected as foreground that was supposed to be background) do occur rather frequently if

the parameters for the specific scene aren't set properly, which makes this algorithm rather difficult to set up for a generic scene. But in certain detection situations, more false positives can be tolerated for the sake of accuracy of detection.

Foreground Object Detection from Videos Containing Complex Background

This algorithm, developed by Liyuan Li et al. [3], deals with the detection of movement in scenes with complex backgrounds; both stationary and moving background objects, and undergoes both gradual and sudden "once-off" changes. In many shopping malls you have the situation where there are flickering screens, opening/revolving doors, indoor water fountains, high gloss reflective floors, switching lights and so on causing plenty false positive detections. One option is to simply mask these areas as non-detection areas but in doing this you have created a detection black spot. This algorithm is able to deal with scenarios of this kind with much success without "hiding" possible detection areas.

The algorithm employs a Bayes decision rule formulated to classify background and foreground using selected feature vectors. The stationary background object is described by the colour feature, and the moving background object is represented by the colour co-occurrence feature. Foreground objects are extracted by fusing the classification results from both stationary and moving pixels.

The author presented the following block diagram to explain the slightly more complex algorithm:

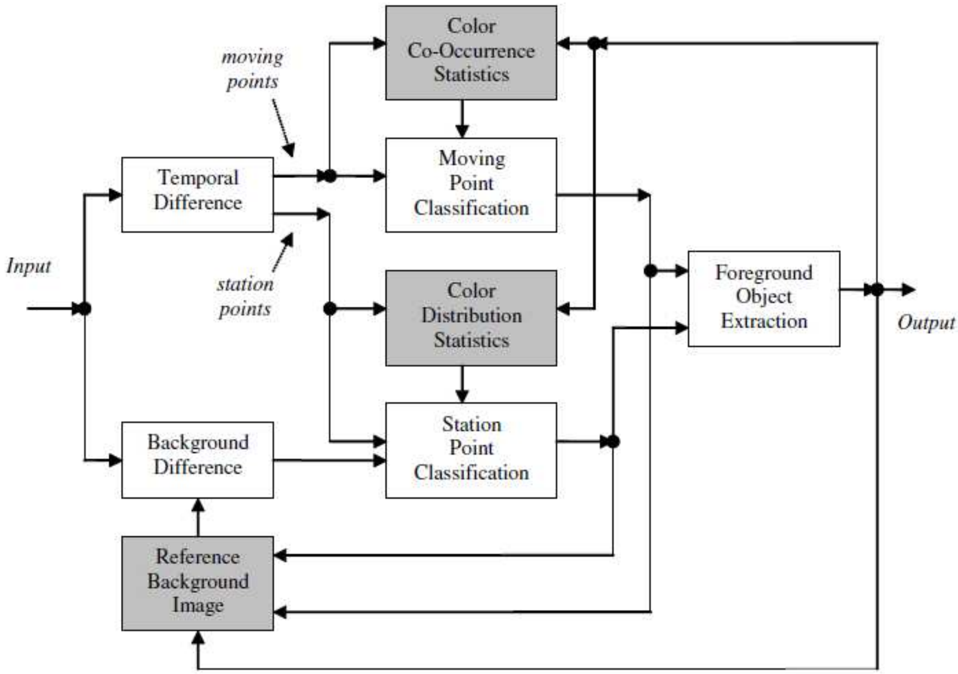


Fig. 1. Algorithm block diagram for the Foreground Object Detection from Videos Containing Complex Background

The algorithm consists of four parts: change detection, change classification, foreground object segmentation, and background learning and maintenance. A block diagram of the Liyuan Li et al. [3] algorithm is shown in Fig. 1. The light blocks from left to right illustrate the first three steps, and the grey blocks illustrate the adaptive background modelling step. In the first step, non-change pixels in the image stream are filtered out by using simple background and temporal differencing (i.e. subtracting two sequential frames). The detected changes are then separated as pixels belonging to stationary and moving objects according to inter-frame changes. In the second step, the pixels associated with stationary or moving objects are further classified as background or foreground based on the learned statistics of colours and colour co-occurrences respectively by using the Bayes decision rule. In the third step, foreground objects are segmented by combining the classification results from both stationary and moving parts. In the fourth and final step, the background models are updated. Gradual and "once-off" learning strategies are applied to learn the statistics of the feature vectors. At each step a reference background image is maintained to make the background difference accurate and adaptive to the changing background. The detail of the algorithm can be obtained from [3].

This algorithm performs well under varying backgrounds with very little false positives, it does however suffer quite a bit from false-negative detections (i.e. not detecting objects it should have) especially in monochrome video such a thermal images. But used in a less critical general monitoring environment this algorithm performs very well with constant detection results.

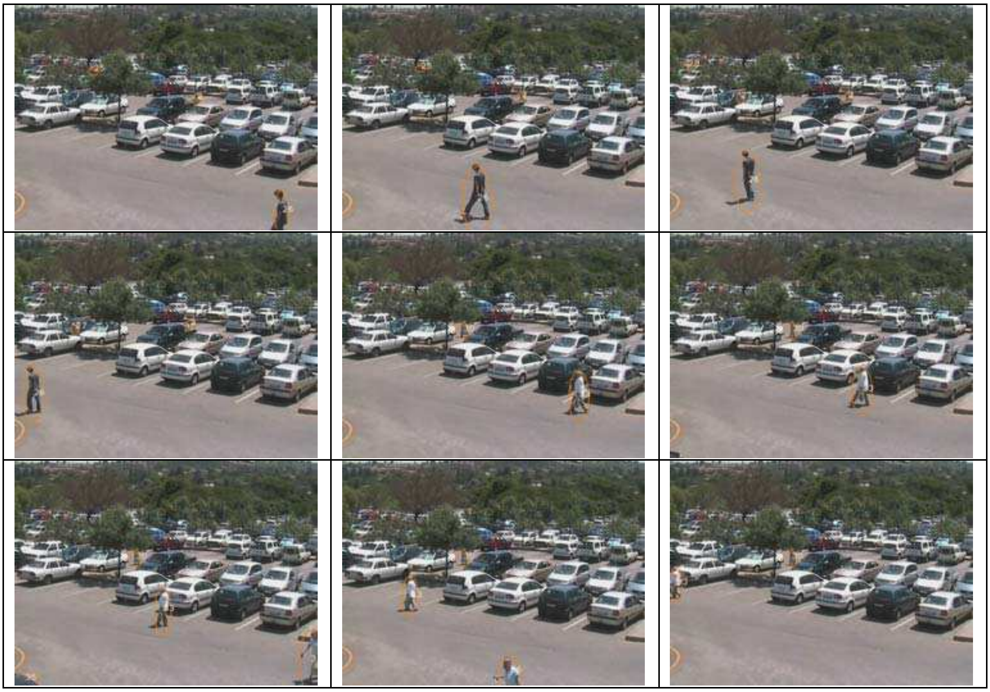


Fig. 2. Video sequence containing a few tracked objects

In comparison the Mixture of Gaussian Modelling and the Complex background both have good merits for use and are even combined in some instances to achieve user specific requirements. However just motion detection is not enough for a good events detection system. The objects detected have to be tracked to allow for further intelligence extraction. The following paragraph will discuss foreground object tracking.

Tracking

Tracking is the process of following the movement of an object over time. In video analysis this would translate to the following of a detected object in between successive frames in video or in more advanced instances between different videos.
Before we jump into an explanation of how tracking works let us look at an example; Fig. 2 shows a sequence of images containing some tracked objects. The images are a few frames taken from video sequence, showing orange ovals drawn around the moving “objects”, each with a tracking number attached to it.
In order to track an object it should first be detected as an object of interest of some nature. In the case of movement detection this would be “foreground” objects detected using any of the previously mentioned motion detectors. The generic output of these methods/algorithms is a series of masks showing groups (blobs) of moving pixels in each frame. Many methods exist to track blobs but the basic principle stays the same; first, detect the “tracking blob” and assign some identifier to it, then detect its position in the next frame and the following ones until it has left the scene or cannot be found anymore.
We have developed our own novel method of tracking blobs based on their contours; to detect a blob, a small buffer of newly untracked blobs is kept and updated with each new frame. If a blob satisfies certain “tracking” criteria, such as size, speed and direction it is added to the list of tracked blobs. A matching blob is then searched for in each new frame. The simplest matching algorithm simply checks whether any of the new blobs found overlaps a currently tracked blob; this however is not always as effective as some blobs may move so fast that they don’t overlap in two consecutive frames. In this instance the historical “track” information of the object is used to predict where the object “should” be and then searches for it in within the predicted parameters. This method is fast and able to track large number of multiples objects in real time (i.e. several individuals entering a building during rush hour). To even further enhance the tracking the object shape can be utilised, this allows us to track objects that may be temporarily occluded or partly hidden from view.
The information from these trackers can be fed in real time from the processing unit to any device or service that can make use of the information. Furthermore, a sub-image of each of these tracked objects can be used in an object recognition algorithm to determine the type of object. This is a valuable capability, since it’s now possible to “see” what it is, where it's heading, and possibly also what it's doing, all automatically and in real time. The amount of intelligence information that can be gathered in a relatively short space of time is enormous. [4]

2.2.2 Intelligence extraction

Once an event has been detected it has to be analysed to determine whether the event is benign or a threat. Traditionally this task was left to humans but, modern video analytic tools promise automation of this. The event is thus analysed such that benign movement such as scene clutter, movement outside regions of interest (ROI), moving trees, busy roads, etc. are ignored and those movements or events that matter are considered.

Fig. 3. Examples of man-made object detection

Intelligence from Typed Text

A very common form of "object" recognition is optical character recognition (OCR), currently widely used in license plate recognition systems. These systems has accuracies of nearly 100% just showing how much this technology have matured. The technology is also harnessed in other situations such as reading shipping crates and ship identification numbers. Wherever characters are written in a typed font this technology can be utilised.

Fig. 4. Examples of object recognition from a video sequence (Face Recognition)

Intelligence from recognising and detecting objects

Object detection and recognition sounds like something reserved for academic research papers, but the truth is that this technology is gaining rapid popularity and the accuracy of detection and recognition is getting better at an ever increasing rate.
One kind of detection that is rather common in the military and rural security applications is the detection of man-made objects. The examples in Fig. 3 shows how a very simple algorithm that uses texture and edge information can be used to detect man-made "regions" in an image. This at a first glance does not look very useful but imagine having to go through many images or video trying to find scenes containing only farmhouses? This technology could certainly speed up the process if the most likely images could be filtered.

Finally there is also object recognition. The images in Fig. 4 present the recognition of a human- face "object" in a video sequence. This was done in real time, which show that the speed at which this can be done has increased dramatically. In a similar fashion to which the face "object" was recognised, any rigid or regular feature object can be recognised by training the algorithm with the features of the object to be recognised. Objects could be the frontal or side view of a vehicle, or the shape of a certain building, or even different weapon kinds and makes. The possibilities are vast and certainly possible.

3. Information management

While we are convinced that Video Analytics will play the dominant role in intelligence extraction, as described above, this is only a part of the overall requirement as seen in figure 5. From this point onward data analysis plays the major role and databases and analysis techniques are dominant.

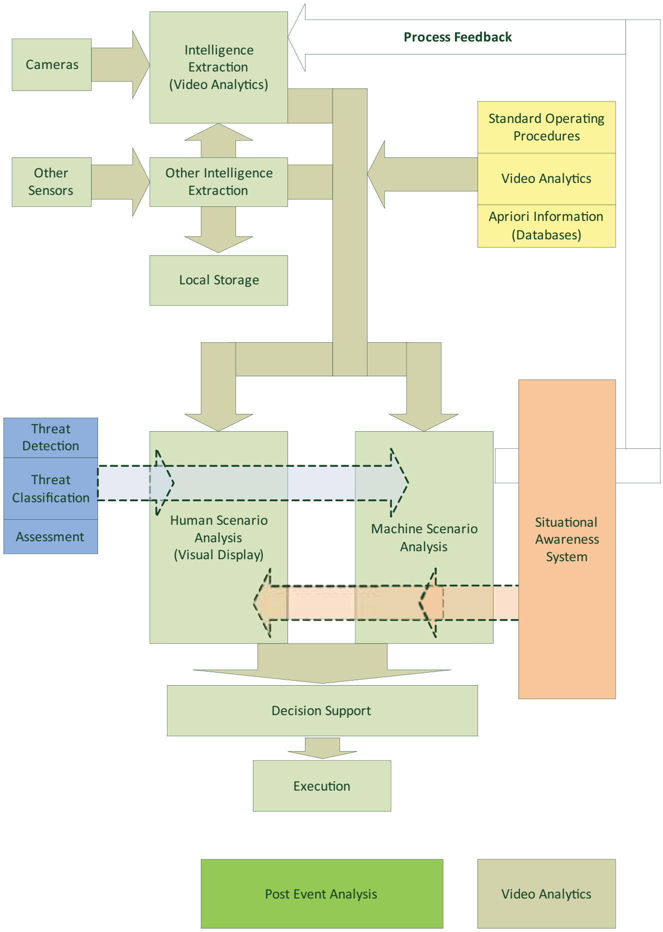


Fig. 5. Information Management Process

Up to this point video analytics played a pivotal role and, improvement in analytic techniques will assist in this. However, the solution to our original problem does not lie solely in our awareness of what is happening around us but also on our ability to recognise the intent of the object, classifying its potential and to be able to put counter effects in motion to prevent unacceptable events. For this we need to add intelligence or predictive ability to the solution.

3.1 Threat detection

After isolating relevant events (intelligence) these need to be classified. At this point the analysis takes a different tack and changes from detection to recognition. A threat could be identified by comparison to a set of known threats, which would be the initial task. This process could be done by **content analysis.**

In addition to recognising threats, a major output of the video analytic system is the ability to provide 'tracks' or a history of the path that an object has taken. This task is achieved by tracking algorithms.

In identifying a threat, a number of parameters are important. Naturally the first would be to identify the object but classifying it as a threat involves more than simply recognising it. Parameters such as direction of movement, speed of movement, linearity (meandering vs purposeful) are important as well. The detection of this is fraught with difficulties such as what to do with multiple objects, ie multiple tracks. Unlike radar images the reality of low angle vision virtually guarantees that objects will pass behind one another (occlusion) resulting in difficulties to attach the track to a specific object. Special algorithms, predictive and otherwise are needed to be able to manage such tracks.

In addition to this, the difficulty of reducing false alarms while at the same time maintaining a high probability of detection is increased dramatically. It is also clear that human intervention at this level will probably always form part of any solution but it is our view that Video Analytic solutions will continually improve and replace human decision making, wherever possible.

Motion-image intelligence extraction

Extraction of intelligence from moving images/video gets a lot more interesting. Once objects can be detected and recognized in each frame, aspects like their movement and behaviour can be analysed which brings a whole new set of automatically extracted information to the table to work with. Following an object's current location can not only give you current behaviour information, but also allow the ability to predict. Behavioural information can be matched up with archived patterns to provide early warning of possible behavioural threats. Proactive decisions, such as pointing a camera, or sending security personnel to the right location in time, can be made, saving precious minutes or even seconds that could just give the upper hand.

With the advancement of computing technology, the speed at which these processes can be performed can be increased considerably, and by adding "machine learning" to regular intelligence questions this extraction can be automated to provide immediate decision support. Imagine deploying several UAV's or autonomous ground vehicles into a disaster or emergency situation and having immediate intelligence information streaming directly back to security and safety headquarters. Intelligence information that can range from something as simple as the number of humans in danger, to a complete situational analysis. A complete situational analysis that could contain a comprehensive breakdown, from type of vehicle, their number plates, their drivers, the identified criminals, the weapons they are wielding to specific threat identification such as fire, explosion hazards and other dangerous situations. Yes, it does sound like something from a science fiction novel, but why not? The technology is there, we should harness it. [4]

3.2 Intelligent analysis

For this context a limited definition of intelligence is the ability to learn about, learn from, understand, and interact with the environment. This general ability consists of a number of specific abilities, which include the following:

- Adaptability to a new environment or to changes in the current environment
- Capacity for knowledge and the ability to acquire it
- Capacity for reason
- Ability to comprehend relationships
- Ability to evaluate and judge

Environment in this definition includes the immediate surroundings, including all objects, reactive capacities and other effects that may influence the judgement.

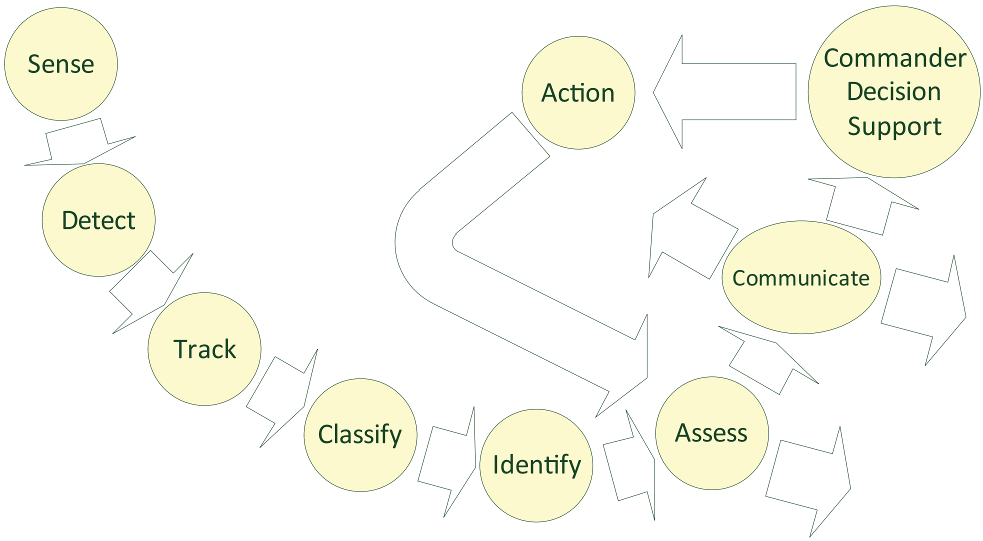


Fig. 6. Decision making process

3.2.1 Intelligence sources

The sources for this information would obviously be the real time sources such as the video information from the cameras. But it should also include other real time sources such as perimeter alarms, information from guards, the news, etc. as well as historical information such as previously recorded footage, faces of suspects and so on.

3.3 Detection of intent

A very important parameter to determine in our problem is the detection of the intent of the identified threat. While any person may be walking in an area, it is the one with malicious intent that is the threat, even though he does not differ physically from the one with benign intent.

A number of algorithms have been demonstrated that attempt to identify this. In this regard algorithms to detect behaviour, especially human behaviour would be those that can contribute most. These algorithms would include relative easy ones such as detection of running, loitering but more sophisticated algorithms can identify aggressive behaviour and possibly recognise specific weapons such as handguns.

The major benefit will come when an object's movement tracks are identified and prediction algorithms are applied to such movement. Thus someone walking along a fence and suddenly turning towards the fence could be identified as having a different intent, possibly a threat.

3.4 Context analysis

Naturally, the analysis of intent is dependent on a clear picture of the current situation, or situational awareness. This aspect would consider not only recognition of and predicting movements but also estimating the threat level and the possible response to such threats. Such decisions clearly require the ability to understand one's own ability to respond and the available options. While we are far from having this kind of engine, at least in practical applications, one can go far by using adequate automatic Standard Operating Procedures (SOP). In this regard analysis of event, behaviour and intent could be a process of applying the pre-determined procedure.

The increasingly sophisticated nature of crime demands a comprehensive approach to solve the problem. Some intelligent video surveillance platforms typically stem from the expansion of Building Management or Access Control systems. What is required is a unified front end that sees and controls all systems on a single user interface. The system should provide a platform that fully integrates DVR's / NVR's, Video Analytics, access control, perimeter alarm systems, fire systems, time and attendance systems and other components. The future has to be **Intelligent Information Management**.

3.5 Data fusion

Proper contextual or scenario analysis requires the ability to evaluate information from different sources. This effort is maintained by a Data Fusion system which generally provides the following functions.

The main functions of the system would include the ability to

- Filter information for relevant intelligence
- Classify the intelligence in the context of the situation
- Be able to predict activity
- Be able to present potential solutions

and finally

DECISION SUPPORT

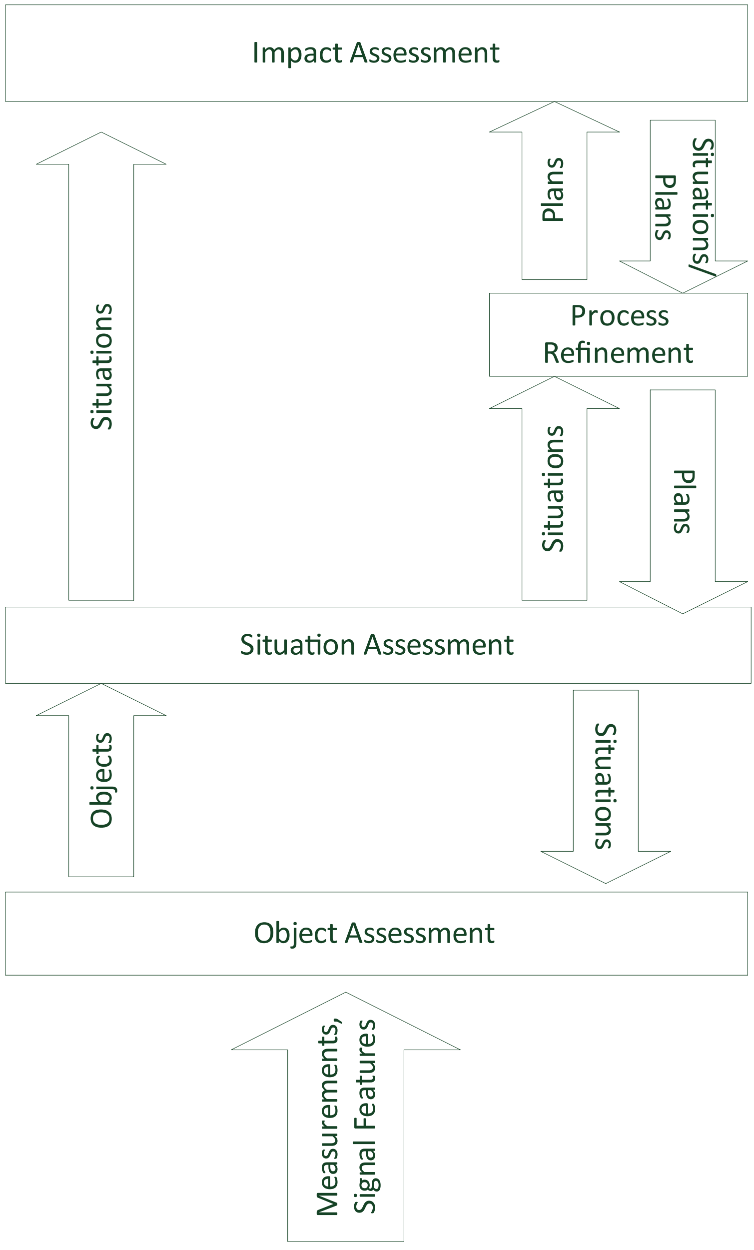


Fig. 7. Data Fusion Process

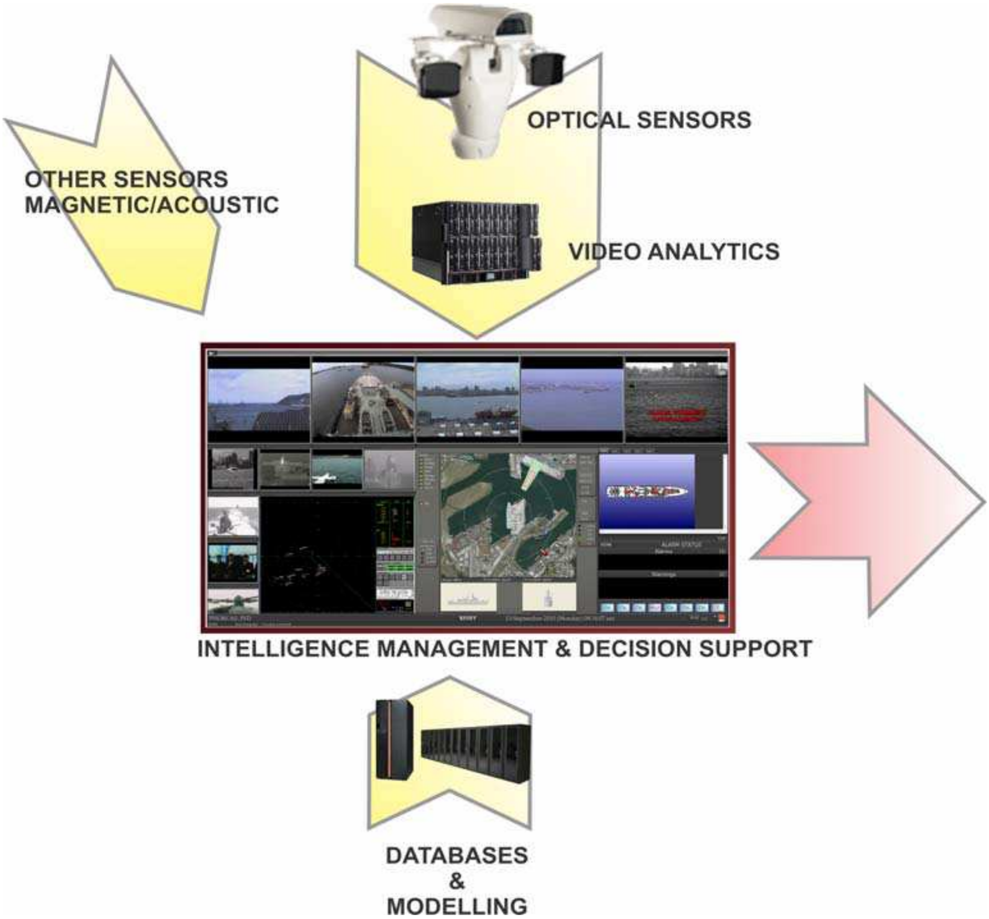


Fig. 8. Unified Decision Support User Interface

4. Conclusion

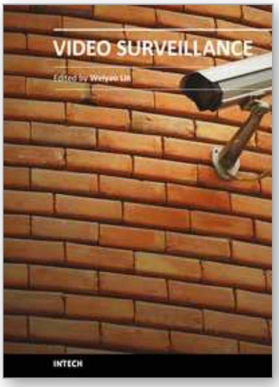
Simplistic approaches to security are just not good enough. This chapter identifies sophisticated detection (Video analytics) and Intelligent analysis as the key factors for future Video Surveillance.

5. References

[1] G. Bradski, A. Koehler, *Learning OPENCV*, Sebastopol, CA: O'Reilly Media, 2008.
[2] P. KaewTraKulPong, R. Bowden, "An Improved Adaptive Background Mixture Model for Realtime Tracking with Shadow Detection," *Proceedings of the 2nd European Workshop on Advanced Video Based Surveillance Systems*, AVBS01. Sept, 2001

[3] L. Li, W. Huang, I. Y. Gu, Q. Tian,."Foreground object detection from videos containing complex background," *Proceedings of the Eleventh ACM international Conference on Multimedia*, MULTIMEDIA '03. ACM, Nov 2003.

[4] B.H. Coetzer, J.S. van der Merwe, "Interoperability in Visual Command & Control Systems" *Proceedings of the 4th Military Information and Communications Symposium of South Africa*, MICSSA 2009, July, 2009



Video Surveillance
Edited by Prof. Weiyao Lin

ISBN 978-953-307-436-8
Hard cover, 486 pages
Publisher InTech
Published online 03, February, 2011
Published in print edition February, 2011

This book presents the latest achievements and developments in the field of video surveillance. The chapters selected for this book comprise a cross-section of topics that reflect a variety of perspectives and disciplinary backgrounds. Besides the introduction of new achievements in video surveillance, this book also presents some good overviews of the state-of-the-art technologies as well as some interesting advanced topics related to video surveillance. Summing up the wide range of issues presented in the book, it can be addressed to a quite broad audience, including both academic researchers and practitioners in halls of industries interested in scheduling theory and its applications. I believe this book can provide a clear picture of the current research status in the area of video surveillance and can also encourage the development of new achievements in this field.