Sony Intelligence Video Analytics - Distributed Enhanced Processing Architecture (DEPA)

Contents
Overview.................................................................2
1. Background ..........................................................3
2. Benefits of Intelligent Video Analytics .........................4
3. Video Analytics Today ................................................5
4. Sony’s New Method:
   “Distributed & Enhanced Processing Architecture” .........6
5. DEPA Technology Infrastructure ....................................9
6. Conclusion ............................................................13
Overview

All security systems perform the same fundamental task – sift through everyday events, identify those that are unusual because they pose a risk to either health or security, and then bring these events to someone’s attention. This editing process has traditionally taken place manually. Human operators observe events as they unfold through “live” monitoring and/or watch recordings of the events. A fundamental factor in the success or failure of this approach is the human’s limited attention span. Typically, long periods of uneventful monitoring are punctuated by the occasional false alarm – this period often exceeds the attention span of a human observer. Even the most dedicated, best-trained security professional loses focus after less than an hour on duty.

To counter this problem and to aid human attention, effective security systems must provide automated solutions to assist this sifting process. The arrival of end-to-end digital IP-based systems offers a fundamental advance in the development of event-based recording techniques through intelligent analytic capabilities. However, the implementation of such solutions has been hampered by an inefficient workflow dictated by legacy system architecture. By mimicking age-old human security organizations, these systems consolidate intelligence on the back-end. This creates workflow bottlenecks that limit capabilities, drive up costs, and create systems that exceed the limits of a human’s attention span.

Sony’s Distributed Enhanced Processing Architecture (DEPA) is a fundamental advance in system design that divides up processing tasks by placing analytic capabilities in third-generation network cameras (SNC-RX550, SNC-RZ50, SNC-CS50 series *) matched with back-end processing capabilities. This workflow innovation has produced a new generation of systems that are inherently more efficient and effective, and also respect the boundaries of human attention. The arrival of DEPA marks a new stage in the development of security systems – a stage akin to the shift from centralized MIS-focused, mainframe-based corporate computing to today’s decentralized, networked environment.

* In the following text, “SNC-RX550,” “SNC-RZ50,” and “SNC-CS50” refer to both NTSC and PAL models (i.e. SNC-RX550N/SNC-RX550P, SNC-RZ50N/SNC-RZ50P, and SNC-CS50N/SNC-CS50P).
1. Background

1.1 The Development of Event-Based Recording
To operate effectively, video-based security system producers have developed numerous strategies to focus on alarm events. Historically, these have been based on targeting when and where to record/monitor live events and then searching through the images by playing back the recording. Crucial developments fall into three areas:

- Time-lapse VCR
The first-generation technology combined long-play VCRs triggered by physical motion detection sensors. However, this method could only achieve rudimentary filtering. The best results with this technique produced lengthy recordings and demanded constant live monitoring. Operators had to watch enormous amounts of video punctuated by false alarms. Searching through these extensive video recordings manually was time-consuming and exhausting.

- Digital Video Recorder (DVR)
Recording on a DVR’s hard disk drastically improves accessibility compared to linear tape recordings. Random access capabilities enhanced with advanced features such as event search help to reduce the time and effort required to locate alarm events. Additional improvements include motion-based recording at higher frame rates. However, these systems are still plagued by an unacceptable number of false alarms.

- Network Video Monitoring System
With the arrival of IP-based components all the signal and control elements of these systems became fully digital. With this change, the digitalization process moves from recorders to cameras equipped with high performance processors. The recorder’s functionality therefore shifts to managing and filtering digital data.

1.2 Camera Image as Sensor
Video systems have utilized various sensor technologies to implement event-based recording strategies that successfully reduce false alarms. Traditionally, physical sensors including infrared detectors, human operators, or integrated CCTV systems with access control systems linked to doors, have been used to trigger recording/monitoring. Over the last five years, advances in DVR and camera-based video motion detection have made these the preferred methods for triggering recordings. Video motion detection offers numerous advantages. It allows for simpler, less costly systems because physical sensor devices are not needed. Physical devices are also limited to the specific area that they cover, while video motion detection covers the entire camera frame. Event-based recording will continue to improve as image-based detection strategies advance.
2. Benefits of Intelligent Video Analytics

2.1 Human Attention & Security System Design

Despite the arrival of IP-based systems, alarm events missed due to operator error remain a persistent problem. The root cause of this problem is the inordinate demands that systems make on the human attention span. False alerts and constantly observing several monitors simultaneously, often push operators beyond the limits of even the most capable security professional. According to scientific studies, the effectiveness of a well-trained operator to detect events drops to just 50% when monitoring nine cameras. Failure to appreciate the limitations of an operator’s attention span can lead to paradoxical systems being designed. For example, increasing the number of cameras to improve security will have the exact opposite effect!

Current motion detection analytics can only tell operators when and where to look at images, but cannot identify specific behaviors. With intelligent video analytics, a system can be tuned to send an alert based on conditions that are highly specific to a camera’s individual role within an installation. A system designer can therefore create highly customized installations – each camera can deliver a range of alarms that signal as various events occur. The operators then respond appropriately. Such a fine-tuned system with its high degree of filtering makes efficient use of the operator’s attention. They do not have to watch multiple screens simultaneously, and images that are selected and presented can be looked at carefully. Intelligent video analytics makes better use of operators’ attention spans and produces greater accuracy and faster response times while they are monitoring events. Consequently, the number of operators remains the same, but the system’s overall effectiveness increases.

2.2 Quick Search

Searching through recordings for events remains the primary means of accessing information captured in security systems, yet this continues to be a difficult and time-consuming task. Playing back recordings to find crucial images, for example, is painstaking and takes operatives a long time. However, intelligent processing can filter footage by searching for particular factors – or combinations of factors – such as the position and size of objects, and their speed and direction. This allows for workflow innovation and improves the overall effectiveness of the security system. The time required to search for events is reduced and operators can spend more time monitoring live proceedings.

2.3 Efficient Recording Improves Storage Capacity Management

Current motion-based recording methodologies make inefficient use of storage capacity. The captured and stored critical information makes up only a small percentage of what is actually recorded. To make up for this inefficient arrangement, most DVRs are forced to compromise image quality to gain additional storage space. But intelligent video analytics-based recording can focus solely on the important information. This greatly reduces wasted storage and, compared to current DVR systems, improves image quality.
3. Video Analytics Today

3.1 Shortcomings of “Back-End Heavy” Consolidated Image Analytics
Using video image processing as a sensor to trigger the monitoring and recording of events is an emerging trend that has been gathering momentum over the past few years. Companies at the forefront of developing this approach have introduced new analytic capabilities powered by centralized, dedicated processors located at the back-end of the system. Placing centralized processing capabilities between an analog camera and a DVR recorder creates a back-heavy system that causes numerous issues and performance bottlenecks.

3.2 Issues Inherent to Centralized, Back-End Heavy Processing

3.2.1 High Cost
Using dedicated CPUs for video analytics is inherently costly. To meet the heavy demands of advanced image analytics, centralized processing requires the highest performance and most expensive processors. The typical cost of a single video processing unit is relatively high. In addition, limitations to processor capacity mean that the number of units required increases quickly as the system expands. Each processing unit can only support very limited number of cameras.

3.2.2 Limitations to Post Recording Search
Most intelligent imaging products are designed for live monitoring. Real-time processing to carry out searches through recorded images also utilizes expensive, high-performance centralized processors. Unfortunately, this means that automated searches are no faster than the time required for real-time playback. Consequently, compared to live monitoring, no time is saved when conducting these archival searches. This falls short of the real goal when conducting searches – to sift through multiple recordings from many cameras in a greatly compressed time period.

Sony’s new IP-based security architecture is a fundamental departure from today’s back-heavy, centralized processor systems. As its name suggests, DEPA distributes processing power throughout the system to avoid the performance bottlenecks inherent to previous designs. A DEPA system assigns specific and appropriate processing tasks to each component while sharing processing power between like components. This opens new opportunities for system flexibility and scalability. Overall system costs are greatly reduced by closely integrating pre-processing analytics and metadata production into third-generation IP cameras having the back-end processing in network video recorders. In addition, the front-end analytics operate across all the cameras simultaneously. This significantly raises the overall security capabilities.

4.1 What is DEPA?

Sony’s distributed video analytics is named DEPA, Distributed Enhanced Processing Architecture. The DEPA design divides traditional processing into two separate tasks. Front-end processing is distributed to the endpoints of systems within cameras while back-end processing takes place at the recorder.

4.1.1 DEPA’s Front-End Processing
- Distinguishes objects from environmental noise.
- Detects moving and/or stationary objects.
- Object information is converted into metadata and then transferred over the network separately from the digital video stream.

4.1.2 DEPA’s Back-End Processing
- Receives and stores pre-processed object data from cameras.
- Extracts objects that match the filtering condition set in recorder.
- Displays information; creates alarm responses appropriate to specific conditions.
4.2 Advantage of DEPA

Sony’s DEPA platform enjoys the following advantages:

4.2.1 High Performance With Minimum Investment

Compared to standard back-end heavy consolidated processing designs, DEPA shares the analytic tasks between the cameras and recorders. With numerous inexpensive processors focused on specific tasks, there is no need for significant investment in high performance processors. DEPA achieves greater overall system capabilities, increases scalability, and requires lower hardware investment while simultaneously avoiding performance bottlenecks.

4.2.2 Pre-Compression Image Processing Yields Higher Accuracy

Compression artifacts that generate digital noise and trigger false alarms are an inherent shortcoming of digital security systems that consolidate analytic functions at the back-end. DEPA’s workflow eliminates this design issue by carrying out front-end processing inside the cameras. These analytics take place before video data is compressed for transfer over the network. Consequently, the object data concerned is not affected by digital noise regardless of the compression type and ratio used later.
4.2.3 Suitable for a Broad Range of Analytics & Recorded Searches
DEPA post-processing does not need to carry out any image processing. Instead, it focuses on analyzing object data. This makes a DEPA system ideal for applications such as multiple channel searches that require heavy back-end analysis. In addition, since the object data is recorded separately from the video streams, this data can be filtered in various ways to speed-up searches through archived video recordings. With consolidated, centralized processing designs, analytics and filtering occur concurrently. This is problematic because the filters cannot be altered to facilitate searches through any recordings. Instead, the search must be started again in real-time, reenacting live monitoring – a painstaking, and time-consuming process.

4.2.4 Conserves Bandwidth
Conserving network resources and optimizing remote location monitoring are key concerns for the owners of IP-based security systems. With DEPA, the object information metadata does not occupy much bandwidth. This information asks the back-end processing to decide which images are needed. Through this highly focused approach, the amount of video data transmitted from the camera is therefore minimized. This is in stark contrast to consolidated, centralized processing designs where the back-end heavy processing requires full time video streaming. To make matters worse, to improve the accuracy of analytics, frame rates exceed those needed for monitoring and recording.

Bandwidth Requirement

![Bandwidth Requirement Diagram]

SNC-RZ50 at 10 fps JPEG Level 5, Number of object in the scene = 10
5. DEPA Technology Infrastructure

5.1 Architecture

Image processing for surveillance systems typically breaks down into two stages. First, pre-processing performs tasks that include image conversion and object information extraction. Secondly, post-processing utilizes this information to carry out analytic functions like decision-making and alert creation. These different processes place different demands on systems. Pre-processing to extract object data is usually processor and memory intensive and the information generated can be shared with several applications. Post-processing does not make such demands but, to select the correct object information that is specific to particular applications, increased flexibility is required. With detection systems that process several objects, an architecture that separates these contrasting tasks into a two-step process offers fundamental advantages. Locating pre-processing in the front-end components and post-processing with the back-end equipment is a logical distribution of tasks. By understanding this concept and reconsidering the respective roles played by cameras and recorders, Sony has reinforced its leadership in advancing IP-based security. Now, the innovative DEPA platform with its closely integrated intelligent cameras and recorders creates a new dimension to security technology – it achieves what was impossible with the legacy systems designed around consolidated back-end processing. DEPA designs yield systems that are more effective, scalable and economical.

5.2 Pre-Processing

DEPA’s pre-processing is capable of detecting both moving and stationary objects.

5.2.1 Moving Object Detection

The core challenge for any motion detection system is eliminating false alarms. The primary cause of false alarms is environmental noise such as wind moving leaves and trees, ripples on water or camera vibration. AGC noise can also be introduced as objects move in and out of shadows thereby triggering an increase in gain and overall scene brightness.
Inaccurate detection is caused either by an object size, an object blending in with other objects of similar color within the frame, or by an object’s speed. Tracking motion vector information is an increasingly popular method of dealing with this problem although some systems do rely on a basic two-frame comparison.

Sony’s third-generation IPELA IP cameras (SNC-RX550, SNC-RZ50, SNC-CS50) enhance the methodology by analyzing as many as 15 continuous frames. A new algorithm drastically reduces environmental noise by identifying any differences between the frames – it distinguishes different shadows and filters out redundant, overlapping patterns and movements. This produces a more robust detection system.

5.2.2 Stationary Object Detection

In addition to superior moving object detection capabilities, Sony’s third-generation IP cameras can also detect stationary objects. Objects that appear, disappear, or are left in the camera frame, are recognized and this information is sent to the back-end analytics so that an appropriate response can be generated.

Object disappears from background

Object becomes a part of background
Sony’s DEPA enabled cameras achieve these capabilities by continuously analyzing scenes and extracting multiple background patterns. The challenge is - images that are virtually identical have minute differences. For example, any scene exposed to daylight or even reflected daylight will experience continual lighting changes as the sun’s position and cloud cover vary. Sony’s new algorithm achieves high detection accuracy by minimizing such environment changes. It identifies objects by comparing the background scene with several reference background models. End users can adjust the camera’s sensitivity (40 seconds to 12 hours) to the length of time that an object remains stationary.

*2: Patent pending

5.3 Post-Processing

5.3.1 Data Processing Performance
DEPA’s post-processing is dedicated to data processing only. This is far less processor-intensive than image processing. Basic post-processing tasks such as filtering object data is a light task that can be carried out in the recorder without the need for a dedicated processor.

5.3.2 DEPA Motion Filters
DEPA’s six post-processing motion filters can be combined and overlaid to select highly specific criteria.

1) Appearance
   When objects appear in a selected area, the system creates an alarm.

2) Disappearance
   When objects disappear from a selected area, the system creates an alarm.

3) Existing
   When objects stay in a selected area beyond a specified time period, the system creates an alarm.

4) Passing
   When objects pass a virtual line drawn in a scene, the system creates an alarm.

5) Capacity
   When objects in a selected area are counted and their number exceeds a specified capacity, the system creates an alarm.

6) Unattended / Removed Objects
   When objects are left in a selected area or when they disappear from an area, the system creates an alarm.

By assessing various pieces of object data including its size, it can be identified with each object ID and other traits such as direction and speed calculated. Basic and calculated data can then be combined to generate additional properties of the object.
Applications for such inferential analysis include:

1) Overlay multiple object filters on the same scene to concentrate on specific objects.
2) Record only those entering, not exiting, a building entrance.
3) Track customer traffic in different retail store departments and then allocate salesclerks to service them. These trends can be recorded by time and date to detect patterns tied to seasonal events or sales promotions.
6. Conclusion

According to IMS research, analytics will remain the greatest growth segment within the security industry for the foreseeable future. New software and hardware products are constantly being developed to serve this expanding market. But profitability within this market segment has proven elusive because of technological limitations compounded by system performance and pricing issues. Whether such intelligent analytics will become broadly accepted in the market depends on ease of use and affordability. Sony’s DEPA platform has now created the infrastructure for broad-based, economical deployment of these advanced analytic capabilities. DEPA’s arrival should make every security system professional reconsider their existing strategy and to begin planning the implementation of the innovative workflow this new architecture makes possible.

Sony’s DEPA is an open development platform. DEPA-enabled products are available from Sony as well as other leading recording solution developers and contents analysis providers. A growing number of DEPA-enabled applications are coming to market ranging from highly specific deep data analysis packages to general-purpose security suites. DEPA enables a growth in the market that allows intelligent analytics to be deployed in a variety of settings by a large number of security operations. This increase in the quality and functionality of IP-based security systems will quickly establish a new standard for safety. This heightened expectation for security will drive demand for these systems across many industries, schools, transportation hubs and municipalities.