

Video Analytics for Tropos Metro-Scale Video Surveillance

A TROPOS NETWORKS WHITE PAPER | APRIL 2008

Introduction

Video analytics is the automatic analysis of video images to identify suspicious events and behaviors in real-time. It takes motion detection to a new level, utilizing local or centralized processing power to characterize moving objects based on their size and shape, and to analyze their behavior based on predefined rules.

Typical activities that can be identified by the latest video analytics software include persons or vehicles entering or leaving an area, illegal loitering or parking, the removal (stealing) of objects, or tailgating through a security entrance.

This paper describes the latest video analytics technology. It looks at the solution architecture components and explains how that architecture works extremely effectively over a Tropos metro-scale wireless IP infrastructure.



Existing Methods of Remote Automated Image Processing

Over the last few years, significant research effort has been focused on the ability to extract meaningful objects and their behaviors from video surveillance images. The result is a large number of proven algorithms for both real-time and offline applications that are implemented on platforms ranging from pure software to pure hardware. These platforms, however, are generally designed to deal with a relatively small number (usually no more than one) of simultaneous image inputs. They are normally designed in one of two main architectures: server or local processing.

Server Processing Architecture

Initial video analytics solutions were most often based on a "Server Processing" system architecture (Figure 1) in which all the image processing tasks are put in one location on a powerful server bank that supports many cameras. The central servers can theoretically support several cameras, because there are only a small percentage of "interesting" occurrences at each camera that actually require processing power. In practice, however, processing multiple algorithms per camera requires a significant image computation load on each server, resulting in normal practical processing limits of around four cameras per server processor.

A 500 camera, city-wide deployment, for example, would require over one hundred quad-processor servers just to perform video content analytics for the system.

Video surveillance networks always require careful design to accommodate local and aggregate bandwidth requirements, no matter which type of IP network is deployed. Of course, network

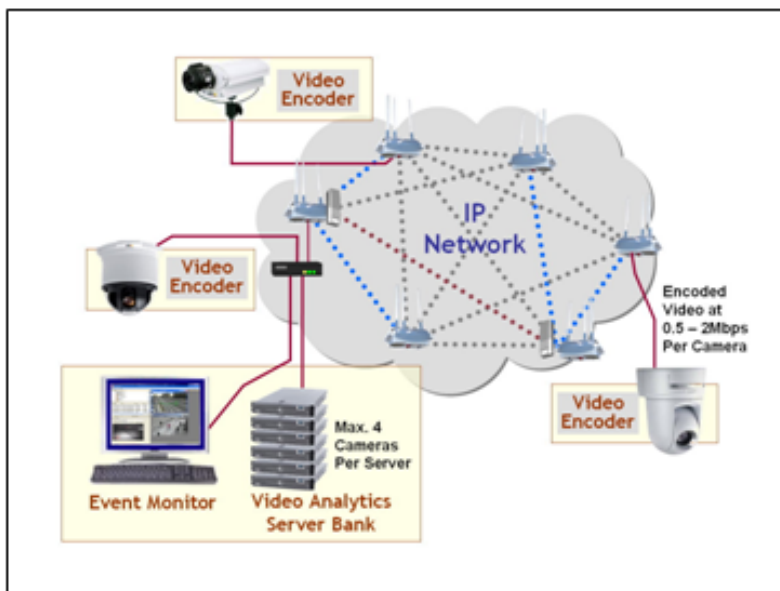


Figure 1: Server Processing Architecture

capacity management is even more important for a wireless infrastructure. A typical Tropos surveillance deployment for example, must take into account the data load at all layers of the network design, from delivering maximum capacity in the mesh by increasing node density, to choosing a high throughput capacity injection layer, to selecting a very high performance backhaul layer. Please see the companion Tropos White Paper "Wireless Video Surveillance with Tropos MetroMesh™" for more details.

Video analytics using a server processing architecture normally increases network throughput over standard NVR requirements, as continuous transmission of high quality images at a high frame rate (10-20 fps) is needed from every camera for reliable analysis. Typical per camera throughput requirements are 0.5-2Mbps or higher depending on the compression codecs used.

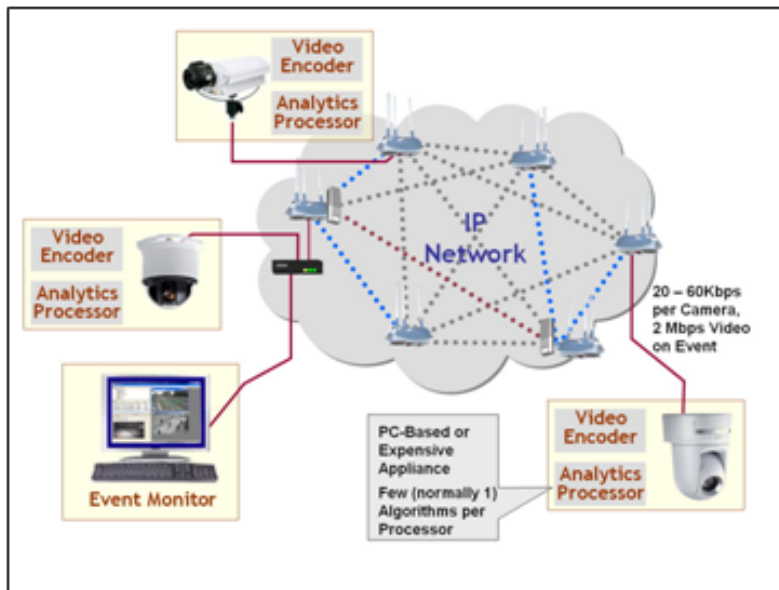


Local Processing Architecture

A common alternative image processing system architecture is "Local Processing (Figure 2)," in which all the processing is done at the camera location at the edge of the network, with the results transmitted through a network connection to the monitoring station. The local processing unit is typically PC-based for more complex solutions, although the recent trend is to move the processing to standalone DSP-based or even ASIC-based appliances.

In this design, the local processing unit performs the entire image-processing task and then transmits an appropriate message to the network when an event is detected.

A video encoder at the camera location (usually built into the camera itself) is used for remote video viewing and recording through the IP network. It is configured to transmit the video using standard video compression techniques such as MJPEG, MPEG-4, etc., at various qualities depending on the application and the available bandwidth.



The key advantage of this architecture is the reduction of network bandwidth required. Since event detection is performed at the edge of the network, non-event data traffic is reduced to a periodic status "ping" from the camera, typically using only 20-60kbps of bandwidth. For smaller, very distributed installations running a limited number of algorithms (usually one per camera), the high cost of the local processors can sometimes be offset by the network bandwidth reduction. However, when the number of cameras increases and a more robust solution is needed, the local processing solution has serious limitations:

Figure 2: Local Processing Network

- Each camera, or small camera cluster, requires its own dedicated processing resources, causing the system cost to scale linearly with the number of cameras. Large-scale systems using this architecture have proven to be prohibitively expensive to install and to maintain.
- Each additional algorithm requires additional processing resources, and the necessary integration between various algorithms can be expensive. PC-based products are difficult to integrate into outdoor camera environments due to space limitations, power requirements, and lack of environmental robustness.
- DSP-based solutions require greater development efforts, due to limited resources and inferior development tools. The result is a very limited (and therefore expensive) choice of compatible cameras or encoder appliances for any given analytics solution.



Requirements for a Viable, Next Generation Analytics Solution

The limitations of these conventional image processing architectures indicate the requirements for a technically viable and cost-effective solution:

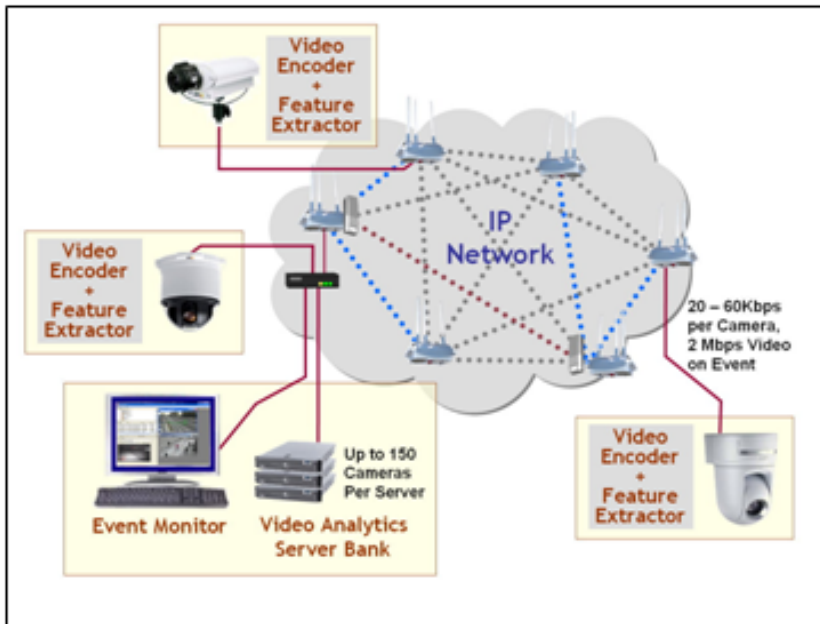
- **Scalability up to metro-scale systems** – The system must be able to simultaneously handle deployments ranging from a few dozen up to thousands of cameras.
- **Cost-effectiveness** – No matter what the scale of the deployment is, the system must provide a cost-effective solution.
- **Minimal network bandwidth requirements** – Very large scale deployments require the use of limited bandwidth wireless networks. As much processing as is economically possible should be performed at the edge of the network to minimize network traffic.
- **Distributed event monitoring and viewing** - A network-connected monitoring station must have the ability for remote viewing of each camera.
- **Multiple algorithm support** - One or more image processing algorithms must be applicable on each camera at any given moment. The outputs of these algorithms must be collected in a central database and should also be viewable on the monitoring station.
- **Multi-camera event support** - The solution must be able to detect both single-camera and multi-camera events. Multi-camera events fuse the information from several sensors to create a higher-level incident perspective.
- **Easy to deploy** – Component power and mounting requirements need to be manageable in order to maintain low deployment and maintenance costs in outdoor, remote environments. PC-based solutions are not an option for these installations.
- **Upgradeability** - The solution should have the flexibility to easily add new algorithms, or customize existing ones, without requiring a massive system upgrade. A Tropos MetroMesh Wi-Fi network uses standard wireless (802.11g, 802.11a, or 4.9GHz) as the communications medium to wireless clients (cameras, laptops or PDAs). For example, there are over 500 million Wi-Fi devices deployed world-wide, and Wi-Fi connectivity is now built in as standard to most new laptops. Many devices such as PDAs and video cameras have been enabled for its use.





IPoIP Architecture

The next generation of video analytics is based on a partially distributed IPoIP (Image Processing over IP) architecture (Figure 3). IPoIP is the core architecture behind the system provided by Tropos' Solution Partner, Agent VI. This system has also been integrated as On-Net Surveillance System's Video Content Analytics module and combines the best of previous designs. The system specifically fills the needs defined above with the following key goals:



- Provides a cost-effective solution for image processing applications over a large number of cameras without sacrificing detection probability or increasing false alarm rate (FAR).
- Enables the application of any algorithm on any camera, even those in geographically remote locations with limited supporting facilities.
- Simultaneously applies a wide range of algorithms to any camera without limiting the solution to a single application at any given time.

Figure 3: IPoIP Architecture

Rather than performing the image-processing task at either the camera or the server, the unique IPoIP architecture distributes the image processing between these locations. The algorithms are segmented into two parts and divided between the IP camera or video encoder hardware and the central image-processing server. Thus, IPoIP retains the strengths of both the local and server architectures, while avoiding their limitations.

This segmented processing leverages the power of the video encoder which is embedded in or deployed near each camera. Used to compress the video signal, the encoder has its own low-cost, fixed-point processor, which is highly suitable for performing several image processing tasks, enabling the unit to send a relatively small amount of information to the central server for the main analysis. In this way, the IPoIP architecture utilizes both the high resolution of the original video and the computing strength and flexibility of the central server, and minimizes the load on the expensive network infrastructure.



Feature Extraction near the Camera

The initial part of the image processing is called the Universal Feature Extraction (UFE). Here, the video encoder processes the part of the algorithm that works at the pixel level, extracting condensed information (features) from the image pixels. This processing is performed on the incoming images, when they are at their highest quality and no data has been lost due to video image compression. When a suitable feature is identified, the encoder sends the data over the IP network to the central server for further analysis. Since the feature data is very compact, it uses a negligible amount of network bandwidth – typically 20-60 Kbps per camera.

The UFE can identify and extract several types of features, including:

- Segmentation of foreground and background.
- Motion vectors – generated by tracking areas of the image between successive frames.
- Histograms.
- Specific color value range in a specified space (RGB, YUV, HSV).
- Edge information.
- Problems with the video, such as image saturation, noise, etc.

All these features share critical attributes: they can be efficiently implemented on fixed-point DSP processors on one hand; and provide excellent building blocks for a wide variety of algorithms on the other.

Feature Analysis at the Central Server

The IPoIP server performs the main part of the processing. The server dynamically requests specific features from each camera, according to the requirements of the algorithms then in force.

The server analyzes the feature data collected from each camera, and dynamically allocates computational resources as needed. In this way, the server is able to utilise large-scale system statistics to perform very complex tasks as needed, without requiring extensive network support.

The part of each algorithm running on the server performs these main tasks:

1. Requests specific features from the remote UFE.
2. Analyzes the incoming features over time and extracts meaningful objects from the scene.
3. Tracks all moving objects in terms of size, position and speed, and calibrates all this data into real-world coordinates. Using various techniques, the calibration process converts two-dimensional data from the sensors into three-dimensional data. Multiple techniques can be implemented for each specific scene.



4. Classifies these objects into several major categories, such as vehicles, people, animals, and static objects, using parameters such as size, speed and shape (pattern recognition).
5. Obtains additional information about "objects of interest," such as color or subclassification (type of vehicle, etc.).
6. Optionally extracts unique identifying features for an object, such as license plate information.
7. Using all the collected information vis-a-vis active detection rules, determines whether to generate an event and notify the system operator.
8. Receives and analyzes information from other algorithms running simultaneously on the server – a powerful capability that easily enables tasks such as inter-camera tracking, which accurately tracks a specific moving object (person or vehicle) from one camera view to the next, while continuously providing the correct image to the system operator. This capability also enables rule sequencing, where a rule on one camera is activated (or deactivated) when a rule on another camera detects an event.

The algorithms at the server are constantly gathering information about the scene, even though, most of the time, there are no events being generated. This information can be stored as meta-data along with the video recording and later used for rapid, efficient searches on large amounts of recorded video content.

The Combined End-Product

Utilizing these methods, the IPoIP Architecture provides a combination of algorithm complexity and low costs that is unrivaled by any other method.

| Feature | Local Architecture | Server Architecture | IPoIP Architecture |
|--|--------------------------------|--------------------------------|---------------------------------|
| Cost for medium to large installations | High (due to many processors) | High (due to network) | Low |
| Suitability for large installations | Network - Yes Hardware - No | Network - No Hardware - Yes | Network - Yes Hardware - Yes |
| Scalability of applications (adding new algorithms and features) | No | Yes | Yes |

The end result is a dual-mode system that delivers the best of both worlds:

- A light-weight UFE module that is small enough to be embedded in the latest generation of IP cameras from Commercial Off-The-Shelf (COTS) manufacturers such as Sony and Axis.
- A centralized rule management and monitoring server that, with much of the video pixel processing distributed out to the intelligent cameras, can now handle up to 100 cameras per server, and apply multi-camera and multiple, logically inter-dependent rules to create real analytical intelligence.



FLIR Camera Support



Unattended Object



Boat Moving over Water



All combined to deliver a next generation, scalable, affordable video analytics system with a vastly superior feature set.

Real World Video Content Analytics Applications

Agent VI and ONSSI have been selected as Tropos Video Solution Partners Their IPoIP-based next generation video analytics system is widely regarded as one of the best-of-breed video analytics components on the market today. It is particularly suitable for the very large scale video deployments that are facilitated by Tropos' MetroMesh city-wide networks.

These metro-scale networks now enable video surveillance systems to grow from enterprise-wide to municipality-wide, and beyond. Video content analytics is becoming an increasingly important tool in the security professional's arsenal as he wrestles with the requirement to effectively monitor and manage hundreds or thousands of cameras.

The systems currently support identification of the following behaviors:

- Person(s) or vehicle(s) moving into or out of an area
- Loitering
- Unattended Object
- Illegal parking
- Tailgating
- Leaving objects behind
- Removing (stealing) objects
- People counting (queues, crowds)
- Vehicle counting
- Boat Moving over Water
- Camera/network tampering

System Integration

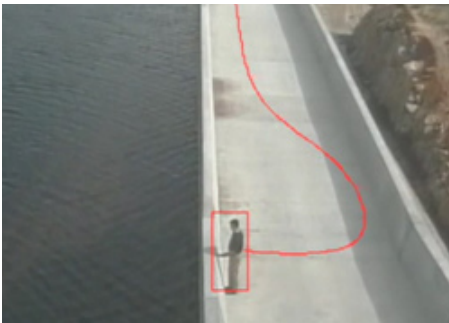
Integration with other system components is a key requirement. Any very large scale video surveillance system is inherently distributed: cameras, multiple command and control sites, multiple fixed and mobile users will all be situated throughout the IP network, which often encompasses an entire city.

ONSSI Video Content Analytics is totally integrated with the other ONSSI system components. Multiple rules can be sequentially processed, and multiple events generated. For example, a fixed camera can be used to monitor a defined security area. A person entering that area (identified by size, shape and speed as a person), can generate an alarm which will cause a nearby PTZ camera to automatically zoom in and track the person, while simultaneously switching the video wall to display the output from both cameras. That same event can also cause video from both cameras to be "pushed" to a remote monitoring station or to a number of mobile users with wireless laptop, PDA or cell phone access.

Conclusions

The convergence of IP-based digital imaging recording and management, with Tropos' industry-standard wireless broadband infrastructure, has created a new class of video surveillance applications on a scale that was previously impractical. It is now possible and economical to deploy video surveillance solutions across entire cities and beyond. Next generation video analytics adds a new level of functionality to metro-scale video surveillance.

Person Tracking



A Tropos MetroMesh Wi-Fi network enables video solutions both in the provisioning of wireless backhaul for remote cameras, and in the distribution of live images and recordings to mobile units in the field. Although video surveillance will always be a high bandwidth application, with care, it is possible to design a MetroMesh network that meets aggregate data bandwidth requirements for high performance cameras and avoids data bottlenecks, while still delivering excellent service to other users of the network. A next generation video analytics system, based on a distributed architecture such as the one described above, greatly improves the system's throughput performance, since cameras need to transmit high frame rate video only when suspicious activity is detected.

Parked Car Detection



While it is generally impractical to monitor hundreds or thousands of cameras with human assets, next generation analytics can provide that monitoring capability for many applications. It is often the only practical way to elevate after-the-fact recording to actionable, real-time security monitoring.

The availability of Tropos metro-scale Wi-Fi networks with the converging technologies of IP video and software-based video monitoring and recording has brought unprecedented economics to large-scale video surveillance applications. Leveraging this digital revolution, next generation video analytics makes these systems easier to install, more efficient to operate, and delivers automatic real-time security monitoring that could only be imagined a few years ago.

About Tropos

Tropos Networks is the market leader in delivering metro-scale wireless mesh network systems with more than 500 customers in 30 countries around the world. The patented Tropos MetroMesh™ architecture delivers the ultimate scalability, high capacity at low cost and great user experience demanded by enterprises and network users. Tropos Networks' unique expertise includes high reliability and performance mesh software development, mesh RF engineering, metro-scale network planning, deployment and optimization, and navigating the municipal approval process. Tropos Networks is headquartered in Sunnyvale, California. For more information, please visit <http://www.tropos.com>, call 408-331-6800 or email info@tropos.com.



555 Del Rey Avenue
Sunnyvale, Ca 94085
408.331.6800 tel
408.331.6801 fax
www.tropos.com
sales@tropos.com