



Boundless Security Systems, Inc.

sharper images with better access and easier installation

Are Manufacturers of Digital Video Surveillance Equipment Responsible for Forensic Problems with Recorded Video ?

Steve Morton, CEO and CTO, Boundless Security Systems, Inc., Newtown, CT

Two articles on forensic video analysis in the September-October 2005 issue of *Evidence Technology Magazine* stated that standards are needed for the compression and recording of surveillance video so that forensic needs for high image quality and interoperability are met. The articles recommended recording video at 25 Mbps using the same digital compression used by consumer camcorders. The articles identified manufacturers of digital video surveillance systems as the cause of a proliferation of compression algorithms and poor quality recorded video.

I wholeheartedly agree with ensuring that recorded images are worth having, and with interoperability. However, as a developer and manufacturer of ultra low bandwidth, multi-stream, digital video surveillance systems, I disagree with the compression format recommended, and with the notion that manufacturers are the primary cause of forensic problems with video.

To be commercially viable, digital video surveillance systems must be affordable by a large market. Many complex, inter-related, system-level factors affect the design of a modern digital video surveillance system, and the amount and cost of the video storage required. Video surveillance systems are no longer a few cameras connected to an analog video recorder. Modern digital video surveillance systems must now co-exist harmoniously with not only computer networks but also the Internet. Simply recording video continuously with high resolution, high frame rate and high clarity can be extremely expensive.

In addition, manufacturers need to consider that the needs for surveillance video by the various law enforcement and security teams conflict with one another. The manufacturers must provide a system that can meet the diverse imaging, communications and video display needs economically.

In this article I try to look at these complex, inter-related, system-level factors from a forensic science point of view rather than a computer science point of view. Let's start with the most basic question of all...

What do You Want to Use the Video for ?

There are more needs for surveillance video than forensic investigations.

I often start a conversation with a prospective customer by asking what they want the video for. I'm not talking about the video technology itself, I'm asking a very basic question: *Why do you want the video at all ?*

If I'm talking to police investigators, they tell me they want the clearest images possible for performing investigations. This investigation is after the incident occurred, possibly days or

weeks later. The video files can be large because time is NOT of the essence, at least not on the scale of seconds.

If I'm talking to tactical police officers, they have a completely different perspective. They are sent into a situation that's in the process of unfolding. Victims' lives and their own lives are at stake. They must assess the situation before engaging to prepare for it, and then intervene in the situation to try to save lives or property, and to neutralize threats. The video files must be small so they can be transported over slow speed networks, and because time IS of the essence.

In between these two requirements are the needs of security staffs that want a replacement for an analog CCTV system. They want to view multiple cameras simultaneously, with reasonable clarity and frame rate, from a variety of locations. The video files can be moderate in size.

There can be a 100:1 difference in data rate between the video used by investigators and the video used by tactical police officers. Differences in resolution, clarity, frame rate and compression parameters can make such a large difference in data rate.

The Price of Video Storage

Let me make what may seem an outlandish statement:

The continuous recording of surveillance video at 25 Mbps (million bits per second) per camera has a purchase price of \$25,000 per month of online storage duration for the disk storage equipment alone.

If this is true, how many customers would be able to afford such a compression and recording format ? This is the purchase price *per camera* for the storage equipment. The purchase price for the storage equipment for continuous online disk storage for a 16-camera system would be \$400,000 per month of storage duration, or \$1.2 million for three months' storage duration that many customers say they want.

How can this be ? Everyone knows that the price of disk storage has fallen dramatically in the last decade.

Indeed, the price of disk storage has fallen phenomenally over the past decade. Ten years ago, I was glad when the storage capacity of disk drives reached 1 GB and the price – just for the disk drives alone, not including the servers to use them – had fallen to \$1,000 apiece. Currently, the least expensive, 250 GB disk drives (just the hard drives themselves) sell on the Internet for a price per gigabyte of less than 50 cents, a staggering 2,000:1 reduction over ten years.

However, the price is still high if we want high data rates, continuous recording and long storage duration.

A Simple Way to Calculate the Amount of Storage Needed

Let's look at just how much disk storage is needed for surveillance video.

I found a simple way to do a back-of-the-envelope calculation of how much storage is needed. I call it “**The Rule of Ones**.”

$$1 \times 1 \times 1 = 1$$

What this means is:

1 camera recorded continuously at **1 Mbps** for **1-quarter of a year** requires **1 trillion bytes** (1 TB = 1,000 GB) of storage

The numbers just happen to work out this way. **The Rule of Ones** is generic; it doesn't make any assumptions about the type of compression being used or the type of data being stored.

If we want to record continuously at some other data rate, such as 25 Mbps recommended in the magazine article, **The Rule of Ones** gives us:

$$1 \text{ camera} \times 25 \text{ Mbps} \times \frac{1}{4} \text{ of a year} = 25 \text{ TB, or } 8+ \text{ TB per month per camera}$$

The Price of Storage

Now, let's look at the price of online disk storage. For large amounts of disk storage, we need server-based network-attached storage. I'm not talking about the price of a single disk drive on an Internet-based price comparison service; I'm talking about a ready-to-use, mass-storage system that can handle a large amount of information.

We can look at the price of enterprise-class, network-attached storage, with server, on Dell's website. In early November 2005, the price of such a mass-storage system made by EMC and sold by Dell is about \$3,000 per terabyte. This price is far more than the price of buying four of the least-expensive, lowest performance, 250 GB hard drives, but Dell is selling a turn-key, networked storage solution with server, not just a few hard drives as components.

Thus 8+ TB per month per camera times \$3,000 per TB gives us \$25,000 per month per camera for continuous recording at 25 Mbps.

Ouch ! What can be done to reduce the amount of storage needed ?

Factors that Affect the Amount of Storage Needed

Reducing the data rate has many benefits. It:

- 1) reduces the amount of storage required, reducing the cost of the storage system
- 2) requires less computation power to compress the video
- 3) requires less computation power to communicate the video

4) requires less communications bandwidth to move the images over local area networks and the Internet

5) reduces the amount of computation power required by display devices to acquire the compressed video, decode it, scale it for presentation, and present it.

Some of the variables that affect data rate are:

- 1) type of digital video compression
- 2) type of image scanning
- 3) spatial resolution with which the video is digitized
- 4) method by which brightness and color are represented
- 5) frame rate with which the video is compressed
- 6) amount of compression applied for a given resolution
- 7) frequency of key frames when inter-frame compression is used
- 8) protocols used to transport compressed video
- 9) choice of continuous or event-based recording

Let's look at these factors in some detail...

1) type of digital video compression

There are two types of video compression, lossy and lossless. Lossy compression can never completely recreate the original image, whereas lossless compression can. I assume that lossy compression is used because it gives far more compression than lossless compression.

There are two classes of lossy compression. Intra-frame compression, such as wavelets and JPEG, compress each frame in isolation. Inter-frame compression, such as MPEG-2 and MPEG-4, takes into account changes between frames, reducing the amount of data when parts of the frames don't change. Motion-JPEG (MJPEG) is a misnomer and falls into the first group not the second, since only the file header information is common across multiple frames.

For round numbers, inter-frame compression gives a 3:1 to 5:1 improvement in compression over intra-frame compression when the scene is static. If the inter-frame compression allows the data rate to vary depending upon the amount of change from frame to frame, popularly called "motion," then even greater benefits can accrue, reducing storage requirements further.

2) type of image scanning

The vast majority of CCTV cameras use interlaced scanning, like broadcast TV cameras. Each frame is digitized as two inter-digitated fields, one with the even scan lines, the other with the odd scan lines. This gives twice as many half-image snapshots per second, tricking the eye into seeing smoother motion. However, any object that's moving is photographed at two separate times, giving motion blur. This is becoming more important as the resolution of digital video

surveillance systems is generally increasing from a resolution that is only 25% of the native camera resolution, using only the odd or even scan lines and horizontally scaling them by 50%, to 100% of the native camera resolution, using all scan lines.

Many CCTV cameras specify an exposure period as short as 1/50,000 second or less, but this factor only affects the camera's ability to capture bright images, not to freeze motion. The total time to expose an entire frame is still 1/30 second for NTSC and 1/25 second for PAL – longer than one would be comfortable hand-holding a 35 mm camera to take a snapshot.

Progressive-scan video cameras, like a 35 mm camera or a digital still camera, take a single photograph of the entire scene and are superior for forensic investigations. They are still not the equivalent of a flash bulb, which freezes the entire image in an instant, but there is much less motion blur than with interlaced CCTV cameras. Motion appears jerkier for a given frame rate.

When a progressive scan camera is used, compressing an image with motion requires about 10% less data than when an interlaced scan camera is used. The reason is that nearby pixels were digitized at about the same time, increasing their correlation, rather than 1/60 second apart.

3) spatial resolution with which the video is digitized

It has been stated in the magazine that video should always be digitized at a spatial resolution of 720 pixels per line so that objects photographed have their true shape in the digital image, simplifying forensic analysis.

The problem is that few CCTV cameras have sufficient resolution to justify digitizing so many pixels per line, wasting data. Many inexpensive, current model, color CCTV cameras have only 330 TV lines of resolution, equivalent to merely 439 pixels per line. Many “high resolution” color cameras claim 450 to 520 TV lines of resolution, equivalent to 600 to 690 pixels per line.

Note that “high resolution” has nothing to do with HDTV, it only refers to the fact that the camera produces higher resolution than the least-expensive model. All of the cameras' video signals are still contained in a “standard definition television” (SDTV) signal.

Since lenses, especially small ones, create barrel and other distortions, it may be advisable to video photograph a test chart with the cameras of interest and use the results to adjust the scaling of the recorded video.

The amount of compressed data is roughly proportional to the spatial resolution of the video, so reducing the resolution from 720x480 to 360x240 roughly reduces the amount of compressed data by 75%. This reduced level is inadequate for forensic investigations, ideal for monitoring with a few cameras on screen, but overkill for situation assessment, where screen resolution and communications speeds are very limited.

4) method by which brightness and color are represented

There are two considerations:

- 1) how much brightness and color information are provided by the video signal from the camera
- 2) how much brightness and color information are provided by the digital compressed image file format

Broadcast-quality video cameras have separate image sensors for each of three colors and produce the same spatial resolution for each color. The output video signal is not necessarily restricted by the bandwidth limits of an over-the-air, NTSC composite (brightness plus color in a single signal) video signal.

For economy, CCTV cameras use a single image sensor that has an array of tiny color filter dots, one for each pixel in the sensor. They use analog or digital signal processing to convert information from the sensor pixels into brightness and color information for the output video signal. The least-expensive cameras use image sensors with a relatively small number of pixels, while more expensive cameras use image sensors with a larger number of pixels. Note that there's no direct connection between a pixel in the sensor and a pixel in the digitized version of the analog video signal produced by the sensor.

The human eye is much more sensitive to brightness than color, so the NTSC system reduces the amount of color information in a video signal to squeeze more visible information into a given amount of bandwidth. The amount of color information in a line of a NTSC composite video signal is only half as much as the amount of brightness information. A very few cameras have S-Video outputs which output the color and brightness information separately, producing higher quality images for two reasons: (1) the color and brightness signals are separate, and (2) the bandwidth limits of NTSC don't apply.

If a composite video signal is used, which is the overwhelming choice, then the spatial resolution of the color information is only half as much as the brightness information. Thus a pair of pixels in a given scan line can have different brightnesses, within bandwidth limits, while those same two pixels have exactly the same color information.

JPEG image compression supports both color and brightness formats above: (1) every pixel having its own brightness and color, as well as (2) the NTSC format where pairs of pixels in a line share color information but have separate brightness information.

However, MPEG-2 and MPEG-4 support only a lower resolution of color information. Every pixel in a 2x2 block can have its own brightness (within bandwidth limits), but all four share the same color information. JPEG also supports this lower level of color information and one should not assume that a JPEG system provides the higher level of color information.

Having two pixels share the same color information reduces the amount of data to be compressed by 25%. Having four pixels share the same color information reduces the amount of data to be compressed by 50%.

5) frame rate with which the video is compressed

The simplest way to reduce the amount of compressed data to process, store, communicate and display is to throw away video frames before compressing them. The key question is how rapidly is the scene changing? Is one trying to capture rapid hand movements in a casino, make sure that no one walks past an open doorway without being photographed, or increase the chance that someone is photographed looking directly at a camera?

The amount of compressed data is roughly proportional to the frame rate of the video, so reducing the frame rate for 30 frames per second (fps) to 3 roughly reduces the amount of compressed data by 90%.

6) amount of compression applied for a given resolution

Spatial resolution is not the only factor that affects the clarity of an image. Another key factor in image clarity is the amount of compression applied to the images.

Digital still image cameras have two key parameters that affect the file size of recorded images: (1) number of pixels, and (2) amount of compression. A relatively small number of pixels with low compression gives an image that is comparable to a relatively large number of pixels with a large amount of compression.

Digital video recorders work similarly. One selects the resolution with which images are digitized, and one selects a “quality” setting. Many digital video recorders have four or more settings for image quality, which select the amount of compression provided. The “standard” setting is often selected because it requires much less data than the “highest” quality setting.

However, images become very blocky as the amount of compression is increased. This is radically different than the behavior of analog video recorders, which apply low pass filtering, smoothing the images rather than creating non-linear compression artifacts. Such non-linear artifacts make image enhancement much more difficult.

Assuming a broadcast quality video camera, with 8 bits of data for each of three colors per pixel, and 480 lines, the amount of raw video data is 211 Mbps for 640x480, and 249 Mbps for 720x480. JPEG compression of 10:1 gives 21 to 25 Mbps, a low level of compression that generally creates sharp recorded images. This is the low level of compression recommended in the prior articles.

However, it is impractical to record such a data rate for long periods of time, to communicate such data streams via the Internet, or to decode and display multiple such images simultaneously.

7) frequency of key frames when inter-frame compression is used

In JPEG and wavelets, each frame is compressed in isolation and is a key frame – a complete frame. MPEG-2 and MPEG-4 achieve more compression than JPEG, but with comparable image quality, by removing similar information between successive frames. From time to time, an

entire frame is compressed in isolation, forming a key frame, and then only changes to that key frame, and to accumulated updates to that key frame, are coded.

Thus the less frequent the key frame, the less data is needed in the compressed data stream. But how often should a key frame be coded ? If one can guarantee that no digital data is lost anywhere in the system, then the less frequent the key frames, the less the amount of data, with no impact whatsoever on the quality of the recorded images.

Issues that affect the frequency of key frames are:

- 1) whether or not the digital system, including the communications network, loses any data
- 2) how long a digital video display system has to wait for a key frame to start showing a new video stream after switching from one digital video stream to another

One might think that digital systems don't lose any data, but that's not correct. Digital communications systems lose data. This leads into a discussion of communications protocols in the next section.

8) protocols used to transport compressed video

Compressed digital images must be moved from one place to another for primary or backup storage, depending on the system, and display. All communications systems, including local area networks and wide area networks (the Internet) have protocols that require overhead data. The overhead in these protocols can range from a few percent for wired networks, where the network itself rarely loses data, to nearly 50% for wireless networks, where data is lost constantly.

Two dominant Internet protocols (IP) for moving data over networks are TCP/IP and UDP/IP. These protocols not only affect whether or not data is lost while it passes through a network, they also affect how long it takes data to pass through a network.

Streaming video, like Voice-over-IP (VoIP) uses UDP/IP. Think of UDP/IP as broadcasting. A TV station broadcasts a video signal continuously, and only once, regardless of whether or not there's static and whether or not you put a tape in your VCR to record it. UDP/IP has a relatively low overhead and a relatively short end-to-end delay. The problem is that it allows blocks of data, or packets, to be lost. Video compression and communications systems that use UDP/IP and MPEG can have errors that ripple through multiple frames.

Digital video systems that use UDP/IP deal with the data loss in two ways:

- 1) error concealment, in other words, they try to hide the loss of parts of the images
- 2) increased frequency of key frames, generally several per second, at the expense of increased data rates for a given image quality

Data is lost even in the best local area networks. The problem is that network traffic and processor activity depend upon random variables. Temporary overloads and thus data loss can occur in the packet buffers in:

- 1) the network interface card in the computer or video encoder that sends data
- 2) ethernet switches
- 3) the network interface cards in computers that receive the data
- 4) in the buffers in the memory of the computers that receive the data

Please note what I'm saying:

Digital video cameras that stream their data using UDP/IP lose data before it's recorded. Display systems for these cameras try to conceal the loss of data.

Digital video surveillance systems that use TCP/IP don't lose data while communicating it. The TCP/IP protocol ensures that information is correctly sent from one place to another, even if it takes extra time to send it. In contrast, one would not want to send one's bank statement using UDP/IP and have the computer system lose data and try to conceal the loss of it.

9) choice of continuous or event-based recording

A common technique for reducing the amount of data storage is to detect an event – usually motion – and only record a brief segment of video before and after the event. The problem is how do you know if the event detection algorithm worked perfectly ? If the algorithm malfunctioned and no information was stored, you will never know it. There are often many parameters that affect the operation of a motion detection system – how can you go into court and assure everyone that nothing significant was missed ?

Some image-understanding, or image analytic, “smart” video surveillance systems try to reduce storage requirements by only recording video when a very complex event occurs, such as a piece of luggage is left in a hall for too long. Problems with this are that you can't later change your mind about what you were looking for because the video was not recorded, and that no algorithm is perfect, so it may not record what you were looking for.

Conclusion

Many complex factors affect the design of modern digital video surveillance systems and the amount and cost of the video storage required. Video surveillance systems are no longer a few cameras connected to an analog video recorder. Digital video surveillance systems must now co-exist harmoniously with not only computer networks but also the Internet. Simply recording video continuously with high resolution, high frame rate and high clarity can be extremely expensive.

One needs to consider the multiple, different uses for video by the various law enforcement and security teams, and provide a system that can meet the conflicting needs economically.

Boundless Security Systems, Inc., makes the ***Boundless Security System™***, an ultra low data rate, multi-stream, digital video surveillance system that simultaneously satisfies this wide range in needs. Its ***Storage Operating System™*** enables it to handle a vast amount of storage.

About the author

Steve Morton is the CEO and CTO of Boundless Security Systems, Inc. (www.BoundlessS.com – three S's) in Newtown, CT. He is the architect of the ***Boundless Security System™***, an ultra low bandwidth, multi-stream digital video surveillance system that makes city-scale, Wi-Fi wireless video surveillance with large numbers of cameras practical. He has a BSEE '71 and MSEE '72 from MIT. He has experience building Air Force communications spacecraft, which require the utmost reliability and security, and central office telephone switches, which require large scale and high reliability. He has been awarded more than 20 patents. He has been a computer systems architect for more than 35 years and has been involved in digital imaging for more than 20 years. He has a special interest in the use of digital imaging for public safety. In response to the 9-11 Commission Report in August 2004, he released a White Paper detailing a dozen hidden vulnerabilities in conventional digital video surveillance systems.

###