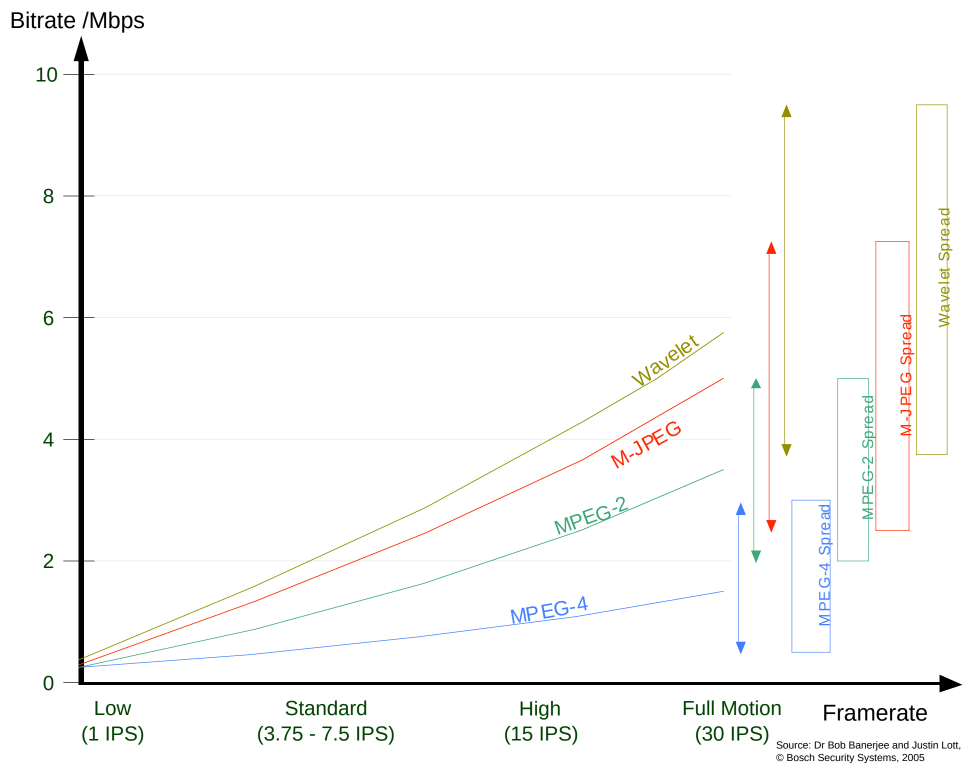
White Paper: Video Compression for CCTV

Over the last few years, the analog video signal from surveillance cameras has been increasingly digitized for archiving and transmission purposes. This article provides an overview of the methods currently available for digitization and data compression, and considers the advantages and disadvantages of their use in CCTV systems.

A digital video signal provides many advantages over analog signals: digital signals can be processed using computers or signal processors and transmitted via digital networks. Transmission via existing (phone) lines and non-wearing recording on hard disks are promoting the trend toward digital video signals. Because a digitized video signal has a very high data transfer rate of 216 Mbit/s (according to CCIR 601), suitable data compression methods must be used for processing and transmission. Imagine, an entire DVD, capable of storing about 8 GB of data, would store less than five minutes of a movie.

Compression algorithms have been developed for various applications at an international level. These algorithms have been tested and standardized by several organizations. The methods are alternatives to a certain extent, because they were developed with different aims and are tailored to specific applications. Overall, the range of available standardized video compression methods covers practically all applications.

Comparison of the Performance of Different Video Compression
Algorithms at Various Framerates



Some compression algorithms permit a degree of variation within the standard to meet technology constraints. The most common video compression methods use the statistical similarity of adjacent pixels. The image is converted to a frequency space block by block using the DCT (Discrete Cosine Transformation). In video coding, it is also possible to use the strong time correlation of consecutive frames – the objects shown are usually static or move in a more or less constant manner. If the difference between the images is calculated with motion estimation, it is only necessary to transfer the changes from frame to frame.

Quick Comparison Table of Compression Algorithms

Compression	Strengths	Weakness	Ideal Applications
JPEG	An industry standard, meaning that a camera they can generate JPEGs is likely to be 'viewable' by the largest range of video management systems. No noticeable degradation at 10-20x compression. Effective with very active scenes. Particularly easy to implement trick-play (play/ pause /rewind / forward wind/ step fwd/rev etc.)	Under high compression ratios the video appears blocky. Inefficient at compressing quiet-moderate activity scenes. Bitrates of 8 Mbps are very common, which makes it hard to transmit but almost unaffordable to store	Storing photographs or snapshots from a video Very low framrate CCTV (<5 IPS) but with high scene-activity level. E.g. Bosch's DiBos v7 Hybrid DVR.
M-JPEG	Same as JPEG, except the JPEG images are payed in rapid succession.	As per JPEG.	As per JPEG, but particularly common in CCTV due to its simplicity. It's popularity has caused the CCTV market to accept low framerates (3.75 IPS) as the 'norm'.
H.320 H.261	Almost the same as MPEG, but optimized for 2-way low latency transmission via ISDN. At 128 kbps (2 ISDN B channels) image quality and image refresh rate are acceptable.	Low performance when higher bandwidths are available.	Video conferencing and videophones. Because of its large range of bandwidth (from 64 to 1920 kbit/s), it can be used in almost all other media (LAN, WAN)
H.263	Optimized for low data transfer rates (below 64 kbps).	Low performance when higher bandwidths are available.	Connections via GSM mobile network, modem and analog telephone lines.
MPEG-1	VHS quality video on extremely cheap CD media.	VHS quality is considered by many as an unreasonable quality restriction.	Low cost video CDs (VCDs). The audio component (so-called Audio Layer 3) of MPEG-1 has become known as MP3.
MPEG-2	The best quality video modern CCTV systems can buy.	Bandwidth-intensive, typically 2.5-15 Mbps per camera.	Broadcast quality video, DVDs, LAN TV, high-fidelity stereo audio. E.g. Bosch's VideoJet 8000 and VIP 1000 encoder and decoder.
MPEG-4 (Part 2)	Very efficient at high framerates when the difference between subsequent frames diminishes. Bandwidth is typically 100-1,000 kbps per camera.	Low efficiency at very low framerates or extremely high scene activity. When the bitrate is limited the video artifacts are speckling and a blocky effect.	CCTV, especially when a high framerate is used, or when the majority of scene activity is low to medium. E.g. Bosch's VIP X1 and VIP XD encoder and decoder. E.g. Bosch's DiBos v8 Hybrid DVR.

Compression	Strengths	Weakness	Ideal Applications
H.264 (a.k.a. MPEG-4 Part 10)	Promising broadcast-oriented technology that is more efficient than current MPEG-4 Part 2	High power processing hardware required and higher lag-times	Next generation MPEG-4, to be used on HD DVDs, HDTV and pay-TV. Microsoft's new XBOX 360 games console is expected to use HD DVD, offering 15 GB DVDs.
Wavelet	Video looks great even under high compression because the human eye accepts a fuzzy picture more than discrete blocks.	Under extreme compression the video becomes fuzzy and diffused, not blocky. Imcompatibility between differerent manufacturers.	CCTV recording. E.g. Bosch's Divar.
Fractal	High compression ratios for images that have large areas of repetition, e.g. sky, water etc.	Slow to compress, requiring intensive computation. Imcompatibility between differerent manufacturers.	Niche photographic compression, e.g. compressing a photo on a smart-card identity card, where compression time is unimportant.

Table summarizing the pros and cons of each of the mainstream compression algorithms

Storage Space vs. Transmission Bandwidth

There are two primary reasons to compress video:

- To reduce the amount of of storage space required to store the video so that more hours of video can be stored on various media including hard disk drives, solid state memory, CDs and DVDs plus many more. Storage is typically measured in bytes (B), Kilobytes (kB), Megabytes (MB), Gigabytes (GB) and Terabytes (TB).
- To reduce the bandwidth required to transmit the video across digital connections including local and wide-area IP networks, the Internet, SCSI connections and USB connections plus many more. Bandwidth is typically measured in bits per seond (bps or baud), Kilobits per second (kbps), Megabits per second (Mbps) and Gigabits per second (Gbps).

In the first generation digital systems DVRs replaced VCRs, and so the market became familiar with units that reflect storage needs, kB and MB etc. However in the second generation systems transmission across networks became so important that units like kbps and Mbps also became important.

Thinking about the 'flow rate of data' requires a small shift in thinking and some people get hung up about the mathematical differences. In order to understand the rest of this White Paper we should make sure we are completely comfortable with a few basics.

Bits and Bytes

As far as CCTV is concerned a byte is simply 8 bits. So, a 1 MB compact flash card can store 1 MB (one Megabyte) or 8 Mb (eight Megabits) of data. They are exactly the same. Similarly a 100 Megabit Ethernet connection, also known as fast Ethernet, can carry either 100 Mbps (one hundred Megabits per second) or 12.75 MBps (12.75 Megabytes per second) of data.

The example below shows that you will arrive at the same result whichever way you prefer. Bitrates are more commonly used when discussing networks, and bytes are used when referring to storage capacities, and it is critical to understand both.

Worked example

If a camera produces 10 images per second where each image is 15 kB, how many cameras can be shared across a 100 Mbps Ethernet connection, and how long would a 200 GB hard disk drive at the other end last before it is filled up and begins to record over itself?

Amount of data transmitted per second is

10 images * 15 kB per image = 150 kB per second

150 kBps * 8 bits per byte = 1200 kbps

1200 kbps / 1024 = 1.17 Mbps

A 100 Mbps connection can carry 100 Mbps / 1.17 Mbps = 85

Therefore, 85 cameras can be carried by a 100 Mbps connection. Note that you will not want to run your network at 100% utilization, so in reality you should select a more realistic bandwidth target.

Method 1 (bitrate-oriented):

A 200 GB hard drive is the same as 200 GB * 1024 = 204,800 MB

204,800 MB * 8 bits per byte = 1,638,400 Mb

The drive is being filled at the rate of 1.17 Mbps so

1,638,400 Mb / 1.17 Mbps = 1,400,340 seconds

1,400,430 / 60 / 60 / 24 = 16.2 days

So the 200 GB drive would offer just over 16 days of recording.

Method 2 (byte-oriented):

A 200 GB hard drive is the same as 200 GB * 1024 = 204,800 MB

204,800 * 1,024 = 209,715,200 kB

The drive is being filled at the rate of 150 kBps so

209,715,200 kB / 150 kBps = 1,398,101 seconds

1,398,101 / 60 / 60 / 24 = 16.2 days.

So the 200 GB drive would offer just over 16 days of recording.

Note: the difference of 38 mins (1,400,430 - 1,398,101) is due to the rounding of 150 kB per second to 1.17 Mbps.

Spatial vs. Temporal Compression

The world of compression techniques revolves around two basic concepts: Spatial compression and Temporal compression.

Spatial compression is used on a single image as a completely independent entity with no relation to other frames, and removes data unuseful for describing the image.

Figure 1: *A Five-picture JPEG video sequence.*

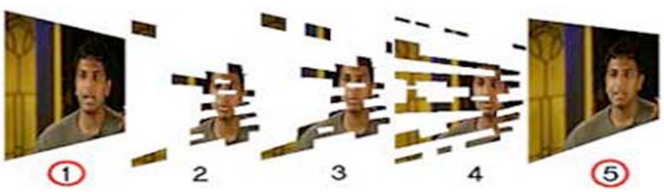


Figure 2: *A Five-picture MPEG video sequence.*

Temporal compression focuses on storing the changes between subsequent images (frames) rather than the entire frame in isolation. If very little changes between subsequent frames then very little storage space is consumed.

Some algorithms such as MPEG, are a careful blend of both – compressing individual frames, but also recording only the changes between frames.

Lossy vs. Non-Lossy Compression

Compression techniques are also divided into those that lose nothing (non-lossy or lossless) of the original image and those that do lose some of the original information (lossy).

While it would appear that a non-lossy algorithm would be the obvious choice, it is important to realize that they consume far for space than lossy algorithms. Also lossy algorithms are smart enough that you can adjust the amount that is lost. This means that while some information may be lost, it can be controlled to be indescernable to the eye.

JPEG

JPEG is an extremely popular spatial compression algorithm focusing on individual photographs or very low framerate video.

The JPEG standard was developed by the Joint Photographic Expert Group (part of ISO) for efficient storage of individual frames. With compression factors between 10 and 20, no visible loss of quality can be perceived in individual frames. With greater compression factors up to 40, quantization artifacts are clearly visible. The resolution is not defined in the standard itself. Based on standard formats and in an effort to reach a good compromise between quality and storage requirements, the CIF resolution of 352x288 pixels has become common in security technology.

M-JPEG

M-JPEG is a spatial compression algorithm, where JPEGs are played in rapid succession to give the illusion of motion video.

A video sequence consists of many individual frames. By repeated application of the JPEG method described above to the frames of a video sequence, it is possible to reduce the data volume of camera signals.

This popular method is called Motion JPEG or M-JPEG. Contrary to common belief it is **not** a new or separate method, and it is **not** a temporal compression algorithm.

Because the relation between individual frames is not taken into account for M-JPEG, this method only yields relatively low compression rates compared with the H.320/H.261 or MPEG methods described below. However, MJPEG is widespread in security technology – especially for archiving video sequences – because of its easy access to individual frames and the relatively low hardware cost. The problem is that the M-JPEG method is not internationally standardized, and JPEG does not include a transmission standard. The implementations of different manufacturers are therefore incompatible.

In a rare variation, the difference between consecutive images are sometimes coded with the JPEG method to further reduce the data volume. This differential frame method is not standardized, so the decoder of the same manufacturer is required.

H.320/H.261

H.320 is a temporal compression algorithm similar to MPEG.

The H.320 standard is an ITU-T recommendation (International Telecommunication Union) and was finalized some time ago. H.320 consists of a series of sub-standards that deal with individual aspects of a complete system. For example, H.261 describes video coding and H.221 is responsible for multiplexing audio, video, data, and control information. The H.320 recommendation is mainly intended for video conferencing systems and videophones. It offers an acceptable image quality and framerate, optimized for transmission via ISDN. At 128 kbit/s (2 ISDN B channels) the quality improves substantially. Because of its large range of bandwidth (from 64 to 1920 kbit/s), it can be used in almost all other media (LAN, WAN).

Since H.320 was developed for two-way video communication between humans, this standard is useful for real-time transmission in security technology. In human-to-human communication, it is important that delays remain below the tenth of a second limit, because otherwise natural conversation is difficult. Another important feature of the H.320 is level of control over the image quality. The user can compromise between acuity-optimized or motion-optimized transmission.

H.261 normally transmits images in CIF resolution (352x288) mentioned above, but can also transmit with a quarter of this resolution in QCIF (176x144) as well as 4 CIF. H.320 is not limited to image coding; it also standardizes all other components of a complete transmission system.

The most significant advantage of H.320 is that it is a standard implementation, which means compatibility between different manufacturers terminals. For example, an ISDN videophone from one manufacturer can communicate audio-visually with an ISDN video conferencing system or an ISDN video transmitter of another manufacturer – as long as both support the H.320 standard.

H.263

H.263 is a further development of the H.261 method. It has been optimized for low data transfer rates (below 64 kbit/s) within the H.324 standard – for example, connections via modem and analog telephone lines. H.320 assumes the use of H.263 as an alternative to H.261 if both terminals support this standard. Especially for transmission in the mobile radio network GSM (9600 bit/s) or in the analog telephone network, use of H.263 improves both image quality and image refresh rate. At higher data rates, the quality is comparable with H.261.

MPEG

MPEG (Moving Pictures Expert Group) is a temporal compression standard developed by an expert group within ISO (International Organization for Standardization).

MPEG-1

MPEG-1 is the standard temporal compression algorithm for Video CDs, extremely popular in the Far East for the distribution of VHS-quality movies.

MPEG was originally developed as MPEG-1 for storage of videos on CD-ROM and was intended to provide a quality comparable to VHS. MPEG-1 is defined as an SIF image format with 352x288 pixels and a typical data transfer rate of 1.5 Mbit/s. The method has a similar structure to H.320. Because of its orientation toward video signal storage, however, delays have not been considered. On the contrary, to support video recorder functions such as fast forward and reverse, I frames (JPEG-coded single frames) have been distributed throughout the data stream to permit random access within a recording. This reduced efficiency is normally compensated for by two bi-directionally interpolated frames (B frames) per coded differential frame (P frame). Since there is no direct relation to real time in a recording, this “trick” has no further consequences. However, in a real-time system, the non-causal forward relation causes delays, because the next coded image has to be transmitted before it is possible to start interpolation of the permitted image. Apart from the storage requirement for several images, remote control of cameras is not possible using this method. If MPEG is used without the B frames and the scattered I frames, it is practically identical to H.320.

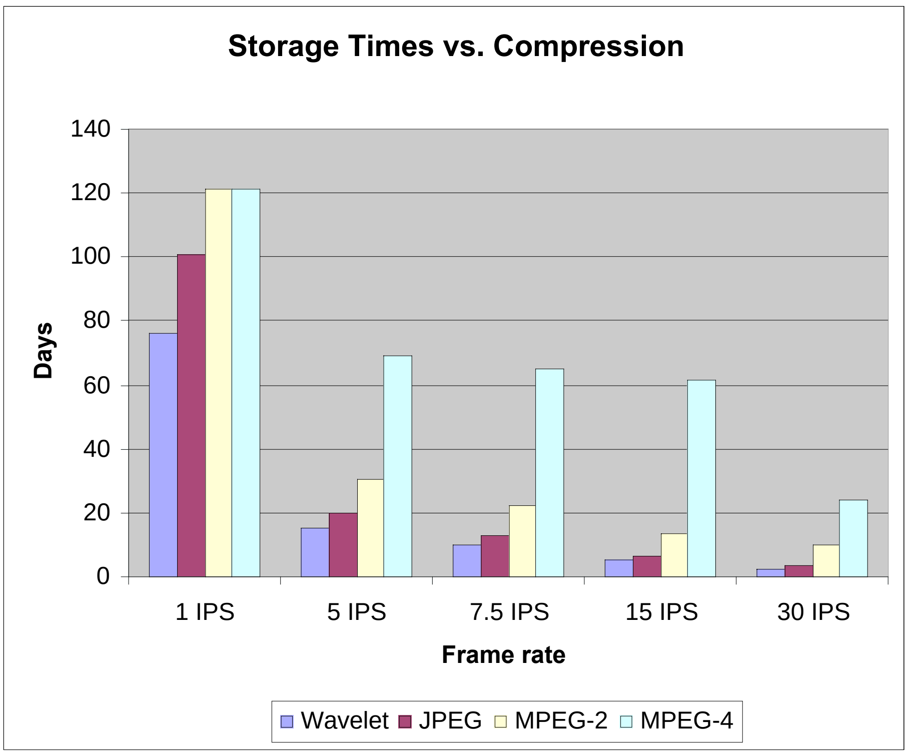


Chart is based on a 250 GB hard drive with 1 camera recording 24 hours per day, 2 CIF and a medium activity scene.

MPEG-2

MPEG-2 is the standard temporal compression algorithm for broadcast quality video.

An extension of the MPEG-1 standard, MPEG-2 was specially developed for digital TV. MPEG-2 takes account of the line interlacing used for standard TV signals. It also provides a number of quality levels and options (called profiles and levels) that permit professional video image transmission. A data transfer rate of up to 15 Mbit/s is defined.

MPEG-4 (Part 2)

MPEG-4 (Part 2) has become the standard temporal compression algorithm for CCTV. It is used in the Bosch Video over IP encoders and decoders.

MPEG-4 was originally created to develop video coding for extremely low data transfer rates (<<64 kbit/s). However, the first results were not very convincing – even compared with H.324. So attention increasingly turned to tools for interaction with the image content and the description of scenes composed of natural and computer-generated contents, hence its popularity in computer video games. MPEG- 4 is a tool-kit permitting description of any bordered sub-areas of an image, and is valuable for editing images and videos. The standardization of MPEG-4 has not yet been completed.

H.264, MPEG-4 (Part 10), Advanced Video Coding (AVC)

This standard is designed to be 50% or more efficient that MPEG-2 and MPEG-4 Part 2 while offering the same video quality. Also designed to fit many applications – both high and low resolutions, high and low bandwidths and many kinds of media ranging from IP networks to DVDs to telephony systems.

H.264 or MPEG-4 Part 10 promises to be the standard for the broadcast industry, but because of the higher latency or time lag it cannot be used directly for surveillance where low latency is important, especially for PTZ camera control. In the broadcast industry a high latency of 1-2 seconds is acceptable.

The compression algorithm is more demanding that MPEG-4 Part 2 which means that while existing hardware can run it there will be a drop in frame rate performance. For full 30 IPS at high resolutions new hardware is required.

The security industry can use select parts of H.264, to benefit from 20-30% improved compression ratios without compromising latency.

One cautionary note, while H.264 is highly fashionable in the security industry it's purpose is to improve the quality of video at lower bitrates. When selecting a system based on H.264 take care to carefully measure the bitrates at the resolution and frame rate you have selected. Some manufacturers have not improved performance but are using H.264 to compress I-frames, which is effectively an M-JPEG approach and offers no advantage except a marketing differentiator.

Wavelet

Wavelet is a spatial compression algorithm that results in a smoother and fuzzier image than JPEG.

Like JPEG, the wavelet method was also developed for coding individual frames. In the JPEG method, the image is first divided up into 8x8 blocks that are then coded individually. But the wavelet method also includes filtering the complete image with the intention of dividing it up into various resolution levels. The image is therefore successively divided up into low-pass and high-pass components (resolution pyramid) that are also important for human perception. Waves of various shapes, called "wavelets," are used for filtering and their coefficients describe the image content.

The advantage of wavelet transformation is its more compact description compared to JPEG (at least theoretically). Moreover, the artifacts that are unavoidable at higher compression rates are usually subjectively felt to be more acceptable with the wavelet method. At a higher compression, wavelet-coded images become fuzzy and the edges are smeared by the filter (overshoot) – whereas with JPEG, the especially annoying "blocking" effect occurs because of the independent image block coding. In practice, the wavelet and JPEG methods do not exhibit any great differences in performance, despite marketing statements to the contrary. The large number of wavelet-based functions used is an impediment to standardization, since there is no compatibility between different encoders and decoders.

Fractal image compression

Like JPEG, this is not a lossless compression method and is based on fractal geometry. It is based on the observation that natural objects exhibit “self-similarity” and obey a fractal geometry in which coarse structures look exactly like fine structures (i.e. they repeat). The task of coding is to find such similarities within digital images and to describe these fractals and their repetitions efficiently. This method is very slow in compression because of the complex analysis, and it has not been standardized.

Conclusion

For security and surveillance, image compression is primarily used for storage and real-time transmission. As technology progresses, the market is demanding higher framerates as well as higher resolution video, primarily intended for transmission over LANs and relatively high-speed WANs.

JPEG and M-JPEG is suitable for the lower framerates, and wavelet is fine for high-bandwidth yet high-compression situations where the video needs to appear smooth but in fact much of the detail has been lost. MPEG-2 offers the best video money can buy but the MPEG-4 compression algorithm still offers the best video quality for the most common networks available today.

JPEG will remain popular because it is fairly well standardized and interoperates with different manufacturers. MPEG-4 is a standardized tool-set, which means that different manufacturers can and have implemented it differently. This means that it is a pre-requisite to carefully select your MPEG-4 source (IP camera, video encoder/server or Network Video Recorder) and your MPEG-4 destination (web browser, video management software or hardware decoder) so that they are compatible.

By the end of the decade it appears inevitable – the standard CCTV system will be pure IP-based. IP cameras will deliver video at high framerates (7.5 - 30 IPS) and at high resolutions (2 or 4 CIF) over LANs and WANs to video management systems and storage systems. In order to take into account the different bandwidth capabilities of LANs and WANs, and different storage capacities of local and centralized storage systems, it will be critical that the video is dual-streamed, one running at a high bandwidth, the other at a lower one – both destined for different purposes.