WHITE PAPER

Digital Video Compression

Reviewing the Methodologies and Standards to Use for Video Transmission and Storage

Created: June 2004

TABLE OF CONTENTS

1 The Changing World of Digital Video Compression

Now that digital video recording and network video have entered the mainstream, attention has turned to methodologies for transmitting and storing digital data in the most effective manner to achieve lower costs and desired performance. This White Paper will encompass a wide-ranging discussion of digital video compression methodologies—particularly Motion JPEG (M-JPEG), MPEG-2, and MPEG-4—and the most appropriate applications for each. The paper will also focus in on MPEG-4, as this compression standard has recently gained in popularity, yet remains largely misunderstood. For this reason, we will explore the hype and reality of MPEG-4 and try to better understand what it actually offers to users. Finally, this paper will conclude with recommendations that answer the questions: Has a single existing methodology proven so effective that the industry should move in this direction going forward? Or is there a combination of standards that are effective and therefore appropriate for the wide range of existing transmission and storage applications?

In a nutshell: Does one compression standard fit all?

When thinking about this question and when designing a network video surveillance application, the following issues should be considered:

What frame rate is required?
Is the same frame rate needed at all times?

- Is recording/monitoring needed at all times, or only on motion/event?
- What length of time must the video be stored?
- What resolution is required?
- What image quality is required?
- What level of latency (total time for encoding and decoding) is acceptable?
- How robust/secure must the system be?
- What is the available network bandwidth?
- What is the budget for the system?

As we attempt to answer if one compression standard fits all network video applications, some background on different compression techniques will also be provided in this paper.

2 A Network Video System Explained

Before we explore the question of which compression method is most suitable for a given application, a short detour is required to better understand the process of digital recording and storage. There are multiple processes at work in a network video system. We will focus on just a few of those most relevant to the subject of compression.

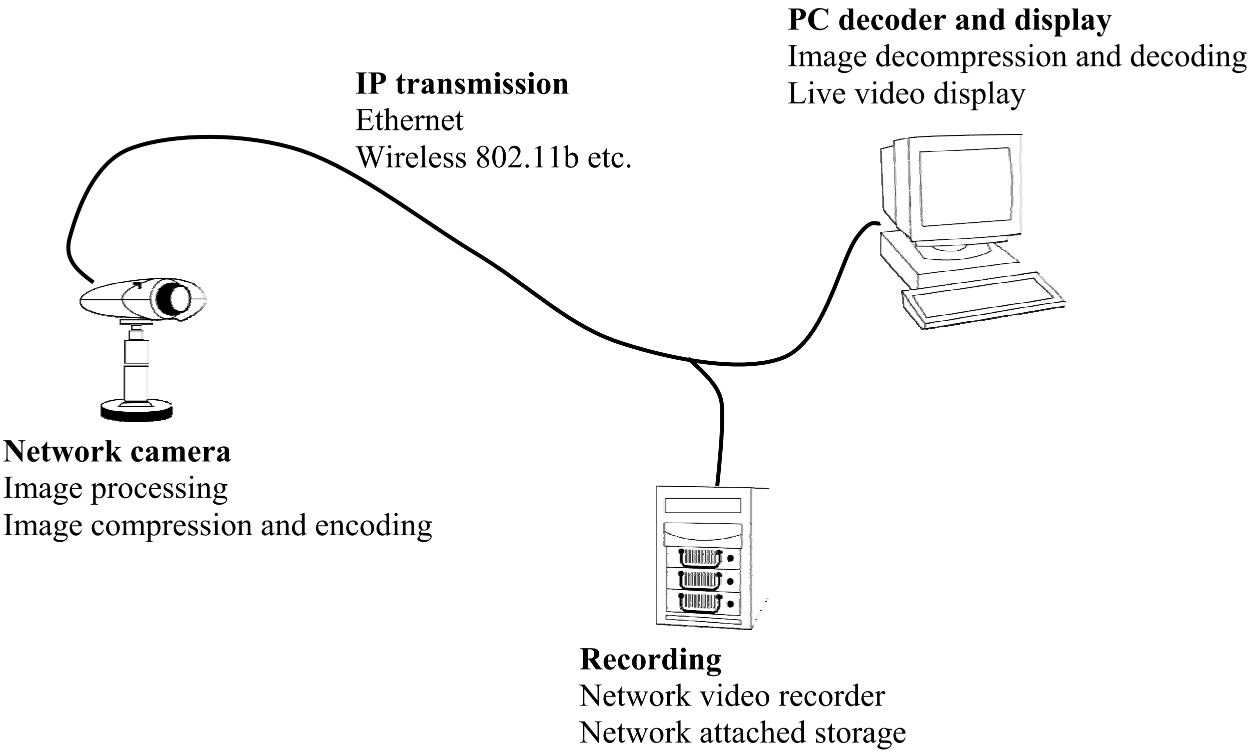
- **Encoding**: The process performed in the network camera or video server, which encodes (digitize and compress) the analog video signal so that it can be transmitted over a network.

- **IP-Transmission:** Transmission over an IP-based network, wired or wireless, of data from a source to various recording and or monitoring hardware (e.g. PC server).

- **Recording:** Transferring data to standard hard disks connected to a storage device such as a server, NAS (Network Attached Server) or SAN (Storage Area Network).

- **Decoding**: Video encoded must then be translated, or decoded, in order to be viewed/monitored. This process is done on PC or other decoder system that is used for viewing the video.

Example of a Network Video System.



3 A Compression Primer

When an ordinary analog video sequence is digitized according to the standard ITU-R BT.601 (CCIR 601), it requires as much as 116 Mbit/s, or 116 million bits every second. Because most networks are only 100 Mbit, it is not practical or desirable to transmit video sequences without some modification. To overcome this problem, a series of techniques – called image and video ***compression techniques*** – have been developed to reduce the high bit rate required for transmission and storage.

Image compression is applied on an individual image by making use of similarities between ***neighboring pixels*** in the image and the limitations of the human visual system. JPEG is an example of such an image compression technique.

Video compression is applied on a series of consecutive images in a video stream, making use of similarities between ***neighboring images***. One example of such a video compression technique is MPEG.

The effectiveness of an image compression technique is given by the ***compression ratio,*** calculated as the original (uncompressed) image file size divided by the resulting (compressed) image file size. At a higher compression ratio, less bandwidth is consumed at a given frame rate. Or, if the bandwidth is kept the same, the frame rate is increased. At the same time, a higher compression ratio results in lower image quality for each individual image.

The trade-offs from compression compound: the more sophisticated the compression technique used, the more complex and expensive the system. What you save on bandwidth and storage, you pay for in system complexity, latency and decoding costs. One must also factor the licensing terms and fees associated with a number of compression standards. These factors generally make sophisticated compression restrictive in terms of keeping the system robust while also achieving and/or maintaining low system costs.

3.1 Image Compression – JPEG

JPEG, a well-known image compression method, was originally standardized in the mid-1980s in a process started by the *Joint Photographic Experts Group*.

JPEG compression can be done at different user defined compression levels, which determine how much an image is to be compressed. The compression level selected has a direct relation to the image quality requested. Besides the compression level, the image scene itself also has an impact on the resulting compression ratio. While a white wall, for example, may produce a relatively small image file (and a higher compression ratio), the same compression level applied on a very complex and patterned scene will produce a larger file size, with a lower compression ratio.

The two images below illustrate compression ratio vs. image quality for a given scene at two different compression levels.

Compression level "low"
Compression ratio 1:16
6 % of original file size
No visible image quality degradation

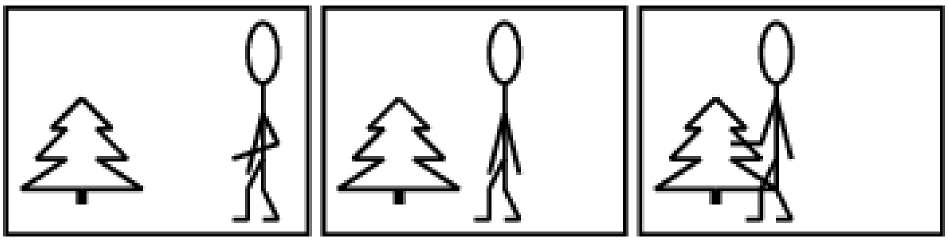
Compression level "high"
Compression ratio 1:96
1 % of original file size
Image quality clearly degraded

3.2 Video as a Sequence of JPEG Images – Motion JPEG (M-JPEG)

Similar to a digital still picture camera, a network camera captures individual images and compresses them into a JPEG format. The network camera can capture and compress, for example, 30 such individual images per second (30 fps), and then make them available as a continuous flow of images over a network to a viewing station. At a frame rate of about 16 fps and above, the viewer will perceive full motion video. We refer to this method as Motion JPEG or M-JPEG.

As each individual image is a complete JPEG compressed image, they will all have the same guaranteed quality, determined by the compression level as defined for the network camera or network video server.

Example of a sequence of three complete JPEG images

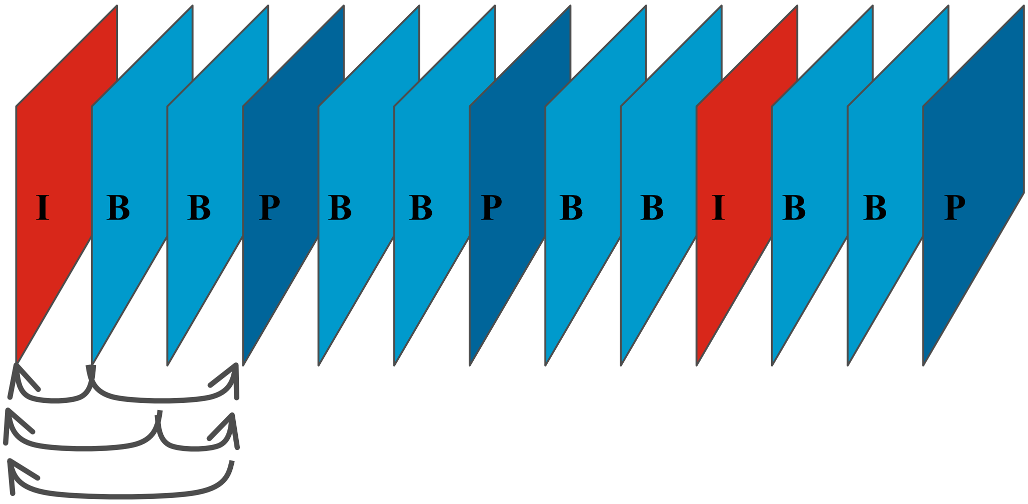


3.3 Video Compression – MPEG

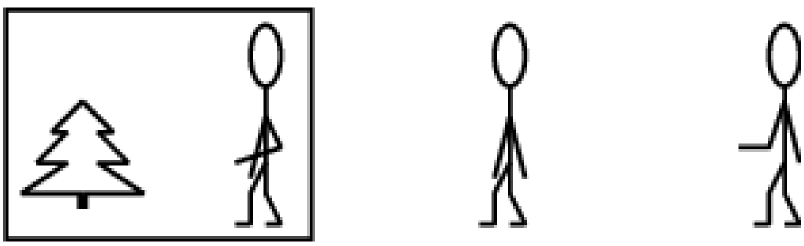
One of the best-known audio and video streaming techniques is the standard called MPEG (initiated by the *Motion Picture Experts Group* in the late 1980s). This paper focuses on the video part of the MPEG video standards.

Simply described, MPEG's basic principle is to compare two compressed images to be transmitted over the network, and using the first compressed image as a reference frame (called an I-frame), only sending the parts of following images (B- and P-frames) that differ from the reference image. The network viewing station will then reconstruct all images based on the reference image and the "difference data" contained in the B- and P-frames.

A typical sequence of I-, B-, and P-frames may look as below. Note that a P-frame may only reference a foregoing I- or P-frame, while a B-frame may reference both foregoing and coming I- and P-frames:



At the cost of higher complexity, the result of applying MPEG video compression is that the amount of data transmitted across the network is less than that of Motion JPEG. This is illustrated below where only information about the differences in the second and third frames is transmitted



MPEG is in fact far more complex than indicated above, and parameters such as prediction of motion in a scene and identifying objects are additional techniques or tools used within MPEG. Also, different applications can make use of different tools, for example comparing a real-time surveillance application with an animated movie. There are a number of different MPEG standards, MPEG-1, MPEG-2 and MPEG-4, which will be introduced below.

3.4 MPEG-1

The MPEG-1 standard was released in 1993 with the target application of storing digital video onto CDs. Therefore, most MPEG-1 encoders and decoders are designed for a target bit-rate of about 1.5Mbit/s at CIF resolution. For MPEG-1, the focus is on keeping the bit-rate relatively constant at the expense of a varying image quality, typically comparable to VHS video quality. The frame rate in MPEG-1 is locked at 25 (PAL)/30 (NTSC) fps.

3.5 MPEG-2

MPEG-2 was approved in 1994 as a standard and was designed for high quality digital video (DVD), digital high-definition TV (HDTV), interactive storage media (ISM), digital broadcast video (DBV), and cable TV (CATV). The MPEG-2 project focused on extending the MPEG-1 compression technique to cover larger pictures and higher quality at the expense of a lower compression ratio and higher bit-rate. MPEG-2 also provides additional tools to enhance the video quality at the same bit-rate; thus producing very high image quality when compared to other compression technologies. The frame rate is locked at 25 (PAL)/30 (NTSC) fps, just as in MPEG-1.

3.6 MPEG-4

The MPEG-4 standard was approved in 2000 and is a major development from MPEG-2. In this section we'll take a close look MPEG-4 to better understand terms and aspects such as:

- MPEG-4 profiles
- MPEG-4 short header and MPEG-4 long header
- MPEG-4 and MPEG-4 AVC

MPEG-4 constant bit-rate (CBR) and MPEG-4 variable bit-rate (VBR)

3.6.1 MPEG-4 part 2 (MPEG-4 Visual)

When people talk about "MPEG-4," it's usually MPEG-4 part 2 that they mean. This is the "classic" MPEG-4 video streaming standard, a.k.a. MPEG-4 Visual.

As a major development from MPEG-2, there are many more tools in MPEG-4 to lower the bit-rate needed to accomplish a certain image quality for a certain application or image scene. Furthermore, the frame rate is not locked at 25/30 frames per second.

It's important to note however, that most of the tools used to lower the bit-rate are relevant today only for non-real time applications. This is because some of the new tools need so much processing power that the total time for encoding and decoding (i.e. the latency) makes them impractical for applications other than studio movie encoding, animated movie encoding, and the like. In fact, most of the tools in MPEG-4 that can be used in a real time application are the same tools that are available in MPEG-1 and MPEG-2.

Another enhancement with MPEG-4 is the larger number of profiles and profile levels (explained below) that cover a wider range of applications, everything from low bandwidth streaming to mobile devices, to applications with extremely high quality and almost unlimited bandwidth demands. The making of studio movies is one such example.

3.6.2 MPEG-4 Profiles

At one end of the system, the encoding of the camera images into the MPEG streaming format takes place. Obviously at the other end of the system, this MPEG stream needs to be decoded and then displayed as video on the viewing station.

Because of the large number of techniques (tools) available in MPEG (especially MPEG-4) to reduce the bit-rate, the varying complexity of these tools, and the fact that not all tools are applicable to all applications, it would have been unrealistic and unnecessary to specify that all MPEG encoders and decoders should support all available tools. Therefore subsets of these tools for different image formats and target bit-rates have been defined.

There are a number of different subsets defined for each of the MPEG versions. Such a subset of tools is called an ***MPEG Profile***. A specific MPEG Profile specifies exactly which tools the MPEG decoder shall support. In fact the requirements are in the decoder and the encoder does not have to make use of all available tools.

Furthermore, each profile exists at different *Levels*. The Level specifies parameters such as maximum bit-rate and supported resolutions. By specifying the MPEG Profile and Level, it's possible to design a system that only uses the tools in MPEG that are applicable to the target application.

MPEG-4 has a number of different profiles. Among them Simple Profile and Advanced Simple Profile are the most commonly used in security applications. While many tools are used by both of these profiles, there are some differences. For example, Simple Profile supports I-, and P-VOPs (frames), while I-, B-, and P-VOPs (frames) need to be supported by Advanced Simple Profile.

Another difference between Simple Profile and Advanced Simple Profile is the supported range of resolutions and bit-rates, denoted by the Level. While Simple Profile goes up to CIF resolution and 384 kbit/s (at the L3 level), Advanced Simple Profile goes up to 4CIF resolution and 8000 kbit/s (at the L5 level).

3.6.3 MPEG-4 short header and long header

Some network video streaming systems specify support for “MPEG-4 short header,” so it’s important to understand this term. In fact it’s nothing more than an H.263 video stream encapsulated with MPEG-4 video stream headers.

MPEG-4 short header does not take advantage of any of the additional tools specified in the MPEG-4 standard. MPEG-4 short header is only specified to allow backwards compatibility with the H.263 recommendation for video conferencing over ISDN and LAN. For practical purposes, MPEG-4 short header is identical to H.263 encoding/decoding, which gives a lower quality level than both MPEG-2 and MPEG-4 at a given bit-rate.

The image and video quality in “short header” is not close to that of true MPEG-4, since it does not make use of techniques that allow it to filter out picture information that is not visible to the human eye. Nor does it use methods like DC and AC prediction, which can significantly decrease bandwidth demands.

To clarify a network video streaming system specification, support for MPEG-4 is sometimes denoted “MPEG-4 long header,” which in other words is the mode where the MPEG-4 compression tools are being used.

3.6.4 MPEG-4 part 10 (Advanced Video Control)

MPEG-4 AVC, also referred to as H.264, is a further development in which MPEG has a completely new set of tools that add more advanced compression techniques to further reduce the bit rate at a given image quality. Being more complex also adds performance requirements and cost, especially for the encoder, to the network video streaming system. MPEG-4 AVC will not be covered in this paper.

3.6.5 Constant bit-rate (CBR) and Variable bit-Rate (VBR)

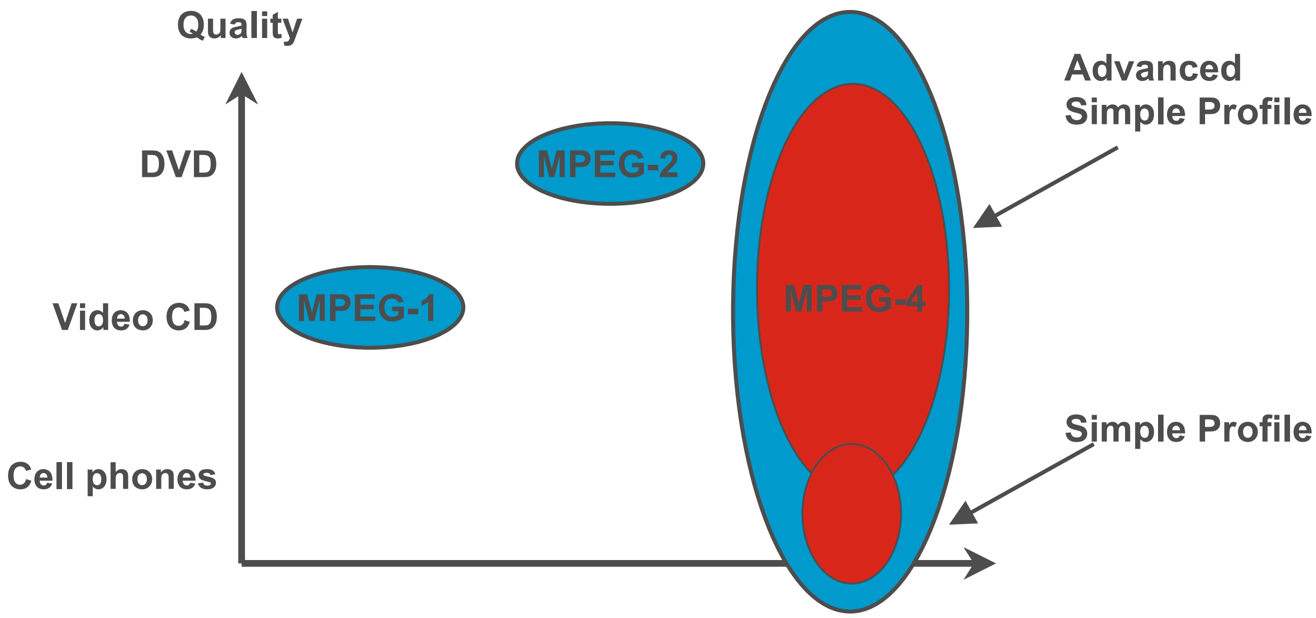
Another important aspect of MPEG is the bit-rate mode that is used. In most MPEG systems, it’s possible to select if the bit rate should run in CBR mode or VBR mode. The optimal selection depends on the application and available network infrastructure.

With only limited bandwidth available, the preferred mode is normally CBR, since this mode generates a constant and predefined bit-rate. The disadvantage is that image quality will vary and while it will remain relatively high when there is no motion in the image scene, the quality will significantly decrease with increased motion.

The VBR mode on the other hand will maintain a high, if so defined, image quality regardless of motion or no motion in the image scene. This is often desirable in security and surveillance applications when there is a need for high quality, especially if there is motion in the picture. Since the bit-rate in VBR may vary, even though an average target bit-rate is defined, the network infrastructure (available bandwidth) for such a system needs to have this capacity.

3.7 Positioning MPEG-1, MPEG-2, and MPEG-4

The illustration below shows MPEG-4's much wider scope relative to MPEG-1 and MPEG-2, which were developed for more specific applications. While MPEG-1 was developed for digital video on CD-ROM, MPEG-2 was developed with DVD and high-definition television in mind. MPEG-4 on the other hand does not have such specific target applications and may be appropriate from the studio down to cellular phone applications.



4 Advantages and Disadvantages for M-JPEG, MPEG-2 and MPEG-4

Due to its simplicity, M-JPEG is a good choice for use in many applications. JPEG is a widely available standard in many systems often by default.

It's a simple compression/decompression technique, which means the cost, in both system time and money, for encoding and decoding is kept low. The time aspect means that there is limited delay between image capturing in a camera, encoding, transfer over the network, decoding, and finally display at the viewing station. In other words, M-JPEG provides low latency due to its simplicity (image compression and complete individual images), and for this reason it's also well suited for when image processing is to be performed, for example video motion detection or object tracking.

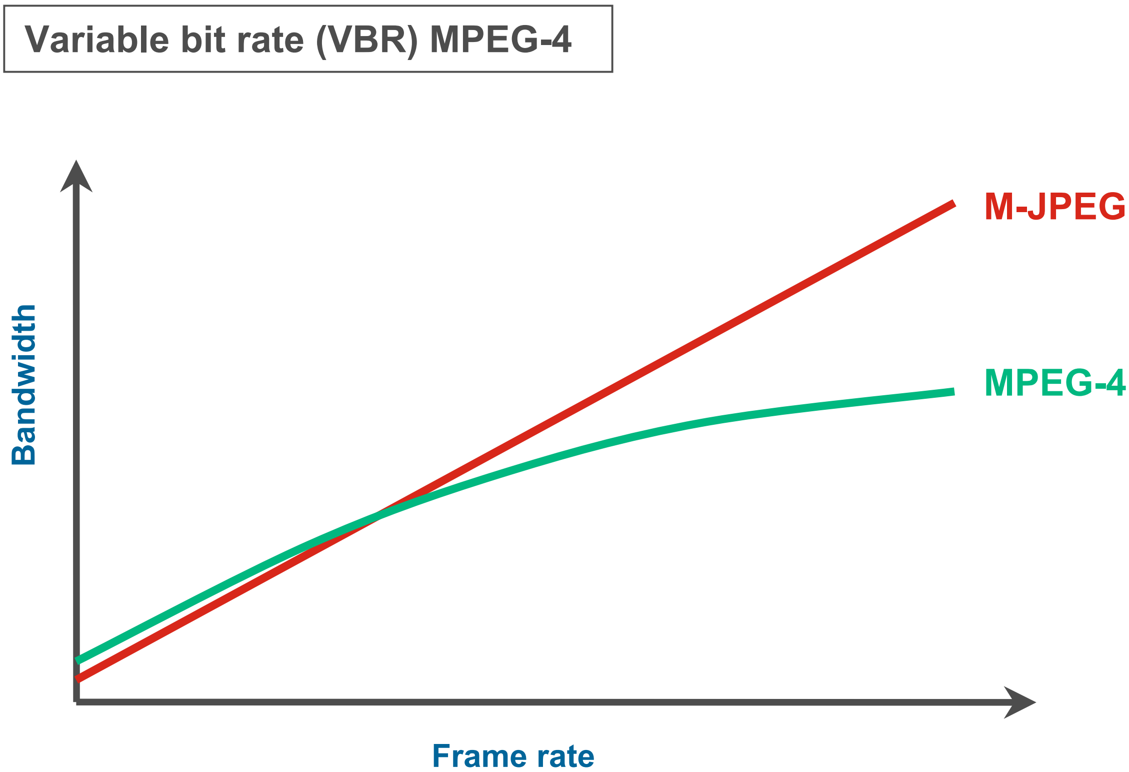
Any practical image resolution, from mobile phone display size up to full video (4CIF) image size, is available in M-JPEG. It also gives a guaranteed image quality regardless of movement or complexity of the image scenes. It still offers the flexibility to select either high image quality (low compression) or lower image quality (high compression) with the benefit of lower image file sizes, thus lower bit-rate and bandwidth usage. At the same time the frame rate can be easily controlled, providing a means to limit bandwidth usage by reducing the frame rate, but still with a guaranteed image quality.

Since M-JPEG doesn't make use of a video compression technique, it generates a relatively large amount of image data that is sent across the network. For this reason, at a given image compression level (defining the image quality of the I-frame and JPEG image respectively), frame rate, and image scene, the amount of data per time unit sent across the network (bit-rate) is less for MPEG compared to M-JPEG, except at low frame rates as described below.

This following neatly summarizes the benefit of MPEG: the ability to give a relatively high image quality at a lower bit-rate (bandwidth usage). This can be especially important if the available network bandwidth is limited, or if video is to be stored (recorded) at a high frame rate and there are storage space restraints. The lower bandwidth demands come at the cost of higher complexity in encoding and decoding, which in turn contributes to a higher latency when compared to M-JPEG.

One other item to keep in mind: Both MPEG-2 and MPEG-4 are subject to licensing fees.

The graph below shows in principle how bandwidth use between M-JPEG and MPEG-4 compares at a given image scene with motion. As can be seen, at lower frame rates, where MPEG-4 compression cannot make use of similarities between neighbouring frames to a high degree, and due to the overhead generated by the MPEG-4 streaming format, the bandwidth consumption is actually higher than M-JPEG.



4.1 M-JPEG Pros and Cons

Pros

- Graceful degradation: if bandwidth is reduced, image quality is maintained at the cost of a lower frame rate
- Constant image quality: quality remains constant regardless of image complexity
- Interoperability: standard compression/decompression available on all PCs
- Low complexity: low-cost coding and decoding. Faster and simpler to perform content searches and do image manipulation
- Less computation-intensive: many channels can be decoded and shown on a PC monitor
- Low latency: encoding and decoding relatively simple and resultant low latency means it's good for live video
- Clear individual images
- Resiliency: fast image stream recovery in the event of packet loss

Cons

- High bandwidth consumption at frame rates > 5 fps
- High storage requirements at frame rates > 5fps

No support for synchronized sound.

4.2 MPEG-2 and MPEG-4

Pros

- Constant frame rate: if bandwidth availability goes down, frame rate is maintained at the cost of image quality (benefits monitoring applications, but not surveillance/recording applications)
- High compression: low bandwidth requirements at frame rates > 5 fps
- Low storage requirements at > 5fps
- Constant bit rate (CBR): simplifies network design and bandwidth provisioning

Cons

- Frame rate fixed to 25/30 fps (only valid for MPEG-2)
- Complex compression: processing requirements on PCs performing decompression is quite high (few channels can be shown live and off-line image analysis is slower)
- Low robustness: if bandwidth goes below a certain threshold all video is lost
- Higher latency: potential problem when monitoring live video, and PTZ control
- Poor transport protocol: designed for real-time viewing, not for recording and analysis
- Less resilient at packet loss: IBP frames need to re-sync and data is lost
- In CBR, image quality is lost with network congestion or movement in image scenes
- Licensing restrictions means no free viewers

5 MPEG-4 : Clearing up the Misunderstandings

As stated at the outset, the MPEG-4 compression standard has generated growing interest in the security industry over the last few years. However, this interest has been accompanied by a considerable amount of misunderstanding and no small amount of hype.

Some of the current misunderstandings about MPEG-4 include questions ranging from the most simple to the most fundamental. Among the questions: What exactly is different and therefore so confusing about the MPEG-4 standard? Why are MPEG-1 or H.263 standards sometimes "rebranded" as MPEG-4? Does a fully functioning MPEG-4 exist as yet? Will MPEG-4 make all other standards obsolete? Finally, What constitutes "good" MPEG-4 and what is "bad" MPEG-4?

Another area of confusion is that in non-real time situations, such has when compressing a DVD movie to MPEG-4, more tools can be deployed to further increase the compression ratio as all information is known and more computing time can be spent. This is why one can see claims of 40% or more in compression gain between MPEG-2 (DVD) and MPEG-4 coding, whereas in real-time situations the differences will be less.

A final area of confusion is that whereas in surveillance there are many sources and few receivers, the MPEG standards were initially designed for broadcasting; that is, one source to many different viewers or the opposite of most security situations. Multicasting technology is one of those features that shows the advantage of both MPEG-2 and MPEG-4, but is not utilized or even desired in video surveillance applications.

It is not the intention of this paper to answer all of these questions or misunderstandings. They are important insofar as they demonstrate that misunderstandings and hype do surround the MPEG-4 standard and that the end user should exercise the classic *caveat emptor* (buyer beware) when considering how appropriate MPEG-4 compression is for their application. End users must check to ensure that "MPEG-4" is supported, but they should also look "under the hood" to determine which level/type of MPEG-4 is specified and what related features, such as latency value and CBR/VBR, exist. For example, in surveillance applications VBR is generally preferred, although it makes network design more complicated.

We come to the most important question: Is MPEG-4 really the best solution considering your needs and budget? Might a M-JPEG based system be a better fit and cost less?

6 Conclusion : One Size MPEG-4 Does Not Fit All

As the somewhat brief comparisons above show, not all of the compression methods we've focused on fit for every installation or application. The most appropriate compression technique for an application largely depends on the trade-offs the user is willing to accept between frame rate, video image quality, latency, system robustness and bandwidth consumption.

In this paper, we've completed a survey of the changing world of digital video compression—the different types, how they compare, and for what applications each is most appropriate. The end user has many questions to ask and many issues to sort through to properly decide which technology is best for their uses and objectives. When choosing a compression methodology, it's important to engage in a very thorough process, and the checklist provided below will help. The end user should also enlist a vendor who has solutions to address all the issues outlined here and one who offers a full range of alternatives.

The Compression Checklist

- ✓ Frame rates less that 5 fps at most times, robust and flexible system, low latency, image quality more important than frame rate or low bandwidth. Image processing to be applied → **M-JPEG**

- ✓ Very high image quality required, (always 25/30 fps), high available bandwidth can be guaranteed, viewing and recording mainly → **MPEG-2**

- ✓ Frame rate more than 10 fps at most times, higher latency OK, limited but guaranteed available bandwidth, viewing and recording mainly → **MPEG-4**

A final caveat: these days many networked systems utilize motion detection so that data is only sent based on triggered events. This functionality, in many cases, has a much higher influence on bandwidth and storage requirements than the choice of compression methodology. Keep this in mind as well.

For the up and coming network video technology, one size most certainly does not fit all. At the current time, while MPEG-4 figures prominently in the solution set, the end user should keep their options open and consider all alternatives before making a final decision.

7 About Axis

Axis increases the value of network solutions. The company is an innovative market leader in network video and print servers. Axis' products and solutions are focused on applications such as security surveillance, remote monitoring and document management. The products are based on in-house developed chip technology, which is also sold to third parties.

Axis was founded in 1984 and is listed on the Stockholmsbörsen (XSSE:AXIS). Axis operates globally with offices in 14 countries and in cooperation with distributors, system integrators and OEM partners in 70 countries. Markets outside Sweden account for more than 95% of sales.

Information about Axis can be found at : www.axis.com

Contact Axis
info@axis.com

Head Office, Lund, Sweden
Axis Communications AB
Emdalavägen 14
SE-223 69 Lund, Sweden
Tel : +46 46 272 18 00
Fax : +46 46 13 61 30

Subsidiaries

BOSTON Phone : +1 978 614 20 00	SHANGAI Phone : +86 21 6431 1690
SAN DIEGO Phone : +1 978 614 20 00	SINGAPORE Phone : +65 6 836 2777
LONDON Phone : +44 870 162 0047	MELBOURNE Phone : +613 9225 5244
MADRID Phone : +34 91 803 46 43	TORINO Phone : +39 011 841 321
MUNICH Phone : +49 811 555 08 30	TAIPEI Phone : +886 2 2546 9668
PARIS Phone : +33 1 49 69 15 50	TOKYO Phone : +81 3 6716 7850
ROTTERDAM Phone : +31 10 444 34 34	SEOUL Phone : +82 2 780 9636