An explanation of video compression techniques.

Table of contents

1. Introduction to compression techniques

JPEG, Motion JPEG and MPEG are three well-used acronyms used to describe different types of image compression format. But what do they mean, and why are they so relevant to today's rapidly expanding surveillance market? This White Paper describes the differences, and aims to provide a few answers as to why they are so important and for which surveillance applications they are suitable.

When designing a networked digital surveillance application, developers need to initially consider the following factors:

> Is a still picture or a video sequence required?
> What is the available network bandwidth?
> What image degradation is allowed due to compression – so called artifacts?
> What is the budget for the system?

When an ordinary analog video sequence is digitized according to the standard CCIR 601, it can consume as much as 165 Mbps, which is 165 million bits every second. With most surveillance applications infrequently having to share the network with other data intensive applications, this is very rarely the bandwidth available. To circumvent this problem, a series of techniques – called picture and video compression techniques – have been derived to reduce this high bit-rate. Their ability to perform this task is quantified by the compression ratio. The higher the compression ratio is, the smaller is the bandwidth consumption. However, there is a price to pay for this compression: increasing compression causes an increasing degradation of the image. This is called artifacts.

2. Standardization organizations

There are two important organizations that develop image and video compression standards: International Telecommunications Union (ITU) and International Organization for Standardization (ISO).

Formally, ITU is not a standardization organization. ITU releases its documents as recommendations, for example "ITU-R Recommendation BT.601" for digital video. ISO is a formal standardization organization, and it further cooperates with International Electrotechnical Commission (IEC) for standards within areas such as IT. The latter organizations are often referred to as a single body using "ISO/IEC".

The fundamental difference is that ITU stems from the telecommunications world, and has chiefly dealt with standards relating to telecommunications whereas ISO is a general standardization organization and IEC is a standardization organization dealing with electronic and electrical standards. Lately however, following the ongoing convergence of communications and media and with terms such as "triple play" being used (meaning Internet, television and telephone services over the same connection), the organizations, and their members – one of which is Axis Communications – have experienced increasing overlap in their standardization efforts.

3. Two basic standards: JPEG and MPEG

The two basic compression standards are JPEG and MPEG. In broad terms, JPEG is associated with still digital pictures, whilst MPEG is dedicated to digital video sequences. But the traditional JPEG (and JPEG 2000) image formats also come in flavors that are appropriate for digital video: Motion JPEG and Motion JPEG 2000.

The group of MPEG standards that include the MPEG 1, MPEG-2, MPEG-4 and H.264 formats have some similarities, as well as some notable differences.

One thing they all have in common is that they are International Standards set by the ISO (International Organization for Standardization) and IEC (International Electrotechnical Commission) – with contributors from the US, Europe and Japan among others. They are also recommendations proposed by the ITU (International Telecommunication Union), which has further helped to establish them as the globally accepted de facto standards for digital still picture and video coding. Within ITU, the Video Coding Experts Group (VCEG) is the sub group that has developed for example the H.261 and H.263 recommendations for video-conferencing over telephone lines.

The foundation of the JPEG and MPEG standards was started in the mid-1980s when a group called the Joint Photographic Experts Group (JPEG) was formed. With a mission to develop a standard for color picture compression, the group's first public contribution was the release of the first part of the JPEG standard, in 1991. Since then the JPEG group has continued to work on both the original JPEG standard and the JPEG 2000 standard.

In the late 1980s the Motion Picture Experts Group (MPEG) was formed with the purpose of deriving a standard for the coding of moving pictures and audio. It has since produced the standards for MPEG 1, MPEG-2, and MPEG-4 as well as standards not concerned with the actual coding of multimedia, such as MPEG-7 and MPEG-21.

4. The next step: H.264

At the end of the 1990s a new group was formed, the Joint Video Team (JVT), which consisted of both VCEG and MPEG. The purpose was to define a standard for the next generation of video coding. When this work was completed in May 2003, the result was simultaneously launched as a recommendation by ITU ("ITU-T Recommendation H.264 Advanced video coding for generic audiovisual services") and as a standard by ISO/IEC ("ISO/IEC 14496-10 Advanced Video Coding").

Sometimes the term "MPEG-4 part 10" is used. This refers to the fact that ISO/IEC standard that is MPEG-4 actually consists of many parts, the current one being MPEG-4 part 2. The new standard developed by JVT was added to MPEG-4 as a somewhat separate part, part 10, called "Advanced Video Coding". This is also where the commonly used abbreviation AVC stems from.

5. The basics of compression

Compression basically means reducing image data. As mentioned previously, a digitized analog video sequence can comprise of up to 165 Mbps of data. To reduce the media overheads for distributing these sequences, the following techniques are commonly employed to achieve desirable reductions in image data:

> Reduce color nuances within the image
> Reduce the color resolution with respect to the prevailing light intensity
> Remove small, invisible parts, of the picture
> Compare adjacent images and remove details that are unchnaged between two images

The first three are image based compression techniques, where only one frame is evaluated and compressed at a time. The last one is or video compression technique where different adjacent frames are compared as a way to further reduced the image data. All of these techniques are based on an accurate understanding of how the human brain and eyes work together to form a complex visual system.

As a result of these subtle reductions, a significant reduction in the resultant file size for the image sequences is achievable with little or no adverse effect in their visual quality. The extent, to which these image modifications are humanly visible, is typically dependent upon the degree to which the chosen compression technique is used. Often 50% to 90% compression can be achieved with no visible difference, and in some scenarios even beyond 95%.

5.1. Lossless and lossy compression

There are two basic categories of compression; lossless and lossy. Lossless compression is a class of algorithms that will allow for the exact original data to be reconstructed from the compressed data. That means that a limited amount of techniques are made available for the data reduction, and the result is limited reduction of data. GIF is an example of lossless images compression, but is because of its limited abilities not relevant in video surveillance. Lossy compression on the contrary means that through the compression data is reduced to an extent where the original information can not be obtained when the video is decompressed. The difference is called the artifacts.

5.2. Latency

Compression involves one or several mathematical algorithms that remove image data. When the video is to be viewed other algorithms are applied to interpret the data and view it on the monitor. Those steps will take a certain amount of time. That delay is called compression latency. The more advanced compression algorithm, the higher the latency. When using video compression and several adjacent frames are being compared in the compression algorithm, more latency is introduced.

For some applications, like compression of studio movies, compression latency is irrelevant since the video is not watched live. In surveillance and security using live monitoring, especially when PTZ and dome cameras are being used, low latency is essential.

6. An overview of compression formats

6.1 JPEG

The JPEG standard, ISO/IEC 10918, is the single most widespread picture compression format of today. It offers the flexibility to either select high picture quality with fairly high compression ratio or to get a very high compression ratio at the expense of a reasonable lower picture quality. Systems, such as cameras and viewers, can be made inexpensive due to the low complexity of the technique.

The artifacts show the "blockiness" as seen in Figure 1. The blockiness appears when the compression ratio is pushed too high. In normal use, a JPEG compressed picture shows no visual difference to the original uncompressed picture.

JPEG image compression contains a series of advanced techniques. The main one that does the real image compression is the Discrete Cosine Transform (DCT) followed by a quantization that removes the redundant information (the "invisible" parts).

Figure 1.   *Original image (left) and JPEG compressed picture (right).*

6.2   Motion JPEG

A digital video sequence can be represented as a series of JPEG pictures. The advantages are the same as with single still JPEG pictures – flexibility both in terms of quality and compression ratio.

The main disadvantage of Motion JPEG (a.k.a. MJPEG) is that since it uses only a series of still pictures it makes no use of video compression techniques. The result is a lower compression ratio for video sequences compared to "real" video compression techniques like MPEG. The benefit is its robustness with no dependency between the frames, which means that, for example, even if one frame is dropped during transfer, the rest of the video will be un-affected.

6.3   JPEG 2000

JPEG 2000 was created as the follow-up to the successful JPEG compression, with better compression ratios. The basis was to incorporate new advances in picture compression research into an international standard. Instead of the DCT transformation, JPEG 2000, ISO/IEC 15444, uses the Wavelet transformation.

The advantage of JPEG 2000 is that the blockiness of JPEG is removed, but replaced with a more overall fuzzy picture, as can be seen in Figure 2.

Figure 2.   *Original image (left) and JPEG 2000 compressed picture (right).*

Whether this fuzziness of JPEG 2000 is preferred compared to the "blockiness" of JPEG is a matter of personal preference. Regardless, JPEG 2000 never took off for surveillance applications and is still not widely supported in web browsers either.

6.4   Motion JPEG 2000

As with JPEG and Motion JPEG, JPEG 2000 can also be used to represent a video sequence. The advantages are equal to JPEG 2000, i.e., a slightly better compression ratio compared to JPEG but at the price of complexity.

The disadvantage reassembles that of Motion JPEG. Since it is a still picture compression technique it does not take any advantage of the video sequence compression. This results in a lower compression ration compared to real video compression techniques. The viewing experience if a video stream in Motion JPEG 2000 is generally considered not as good as a Motion JPEG stream, and Motion JPEG 2000 has never been any success as a video compression technique.

6.5 H.261/H.263

The H.261 and H.263 are not International Standards but only Recommendations of the ITU. They are both based on the same technique as the MPEG standards and can be seen as simplified versions of MPEG video compression.

They were originally designed for video-conferencing over telephone lines, i.e. low bandwidth. However, it is a bit contradictory that they lack some of the more advanced MPEG techniques to really provide efficient bandwidth use.

The conclusion is therefore that H.261 and H.263 are not suitable for usage in general digital video coding.

6.6 MPEG-1

The first public standard of the MPEG committee was the MPEG-1, ISO/IEC 11172, which first parts were released in 1993. MPEG-1 video compression is based upon the same technique that is used in JPEG. In addition to that it also includes techniques for efficient coding of a video sequence.
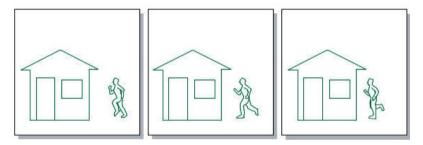
Figure 3. *A three-picture JPEG video sequence.*

Consider the video sequence displayed in Figure 3. The picture to the left is the first picture in the sequence followed by the picture in the middle and then the picture to the right. When displayed, the video sequence shows a man running from right to left with a house that stands still.

In Motion JPEG/Motion JPEG 2000 each picture in the sequence is coded as a separate unique picture resulting in the same sequence as the original one.

In MPEG video only the new parts of the video sequence is included together with information of the moving parts. The video sequence of Figure 3 will then appear as in Figure 4. But this is only true during the transmission of the video sequence to limit the bandwidth consumption. When displayed it appears as the original video sequence again.

Figure 4. *A three-picture MPEG video sequence.*

MPEG-1 is focused on bit-streams of about 1.5 Mbps and originally for storage of digital video on CDs. The focus is on compression ratio rather than picture quality. It can be considered as traditional VCR quality but digital instead.

It is important to note that the MPEG-1 standard, as well as MPEG-2, MPEG-4 and H.264 that are described below, defines the syntax of an encoded video stream together with the method of decoding this bitstream. Thus, only the decoder is actually standardized. An MPEG encoder can be implemented in different way and a vendor may choose to implement only a subset of the syntax, providing it provides a bitstream that is compliant with the standard. This allows for optimization of the technology and for reducing complexity in implementations. However, it also means that there are no guarantees for quality – different vendors implement MPEG encoders that produce video streams that differ in quality.

6.7   MPEG-2

The MPEG-2 project focused on extending the compression technique of MPEG-1 to cover larger pictures and higher quality at the expense of a higher bandwidth usage.

MPEG-2, ISO/IEC 13818, also provides more advanced techniques to enhance the video quality at the same bit-rate. The expense is the need for far more complex equipment.

As a note, DVD movies are compressed using the techniques of MPEG-2.

6.8   MPEG-3

The next version of the MPEG standard, MPEG-3 was designed to handle HDTV, however, it was discovered that the MPEG-2 standard could be slightly modified and then achieve the same results as the planned MPEG-3 standard. Consequently, the work on MPEG-3 was discontinued.

6.9   MPEG-4

The next generation of MPEG, MPEG-4, is based upon the same technique as MPEG-1 and MPEG-2. Once again, the new standard focused on new applications.

The most important new features of MPEG-4, ISO/IEC 14496, concerning video compression are the support of even lower bandwidth consuming applications, e.g. mobile devices like cell phones, and on the other hand applications with extremely high quality and almost unlimited bandwidth. In general the MPEG-4 standard is a lot wider than the previous standards. It also allows for any frame rate, while MPEG-2 was locked to 25 frames per second in PAL and 30 frames per second in NTSC.

When "MPEG-4," is mentioned in surveillance applications today it is usually MPEG-4 part 2 that is referred to. This is the "classic" MPEG-4 video streaming standard, a.k.a. MPEG-4 Visual.

Some network video streaming systems specify support for "MPEG-4 short header," which is 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, which gives a lower quality level than both MPEG-2 and MPEG-4 at a given bit-rate.

6.10 H.264

H.264 is the latest generation standard for video encoding. This initiative has many goals. It should provide good video quality at substantially lower bit rates than previous standards and with better error robustness – or better video quality at an unchanged but rate. The standard is further designed to give lower latency as well as better quality for higher latency. In addition, all these improvements compared to previous standards were to come without increasing the complexity of design so much that it would be impractical or expensive to build applications and systems.

An additional goal was to provide enough flexibility to allow the standard to be applied to a wide variety of applications: for both low and high bit rates, for low and high resolution video, and with high and low demands on latency. Indeed, a number of applications with different requirements have been identified for H.264:

> Entertainment video including broadcast, satellite, cable, DVD, etc (1-10 Mbps, high latency)
> Telecom services (<1Mbps, low latency)
> Streaming services (low bit-rate, high latency)
> And others

As a note, DVD players for high-definition DVD formats such as HD-DVD and Blu-ray support movies encoded with H.264.

6.11 MPEG–7

MPEG-7 is a different kind of standard as it is a multimedia content description standard, and does not deal with the actual encoding of moving pictures and audio. With MPEG-7, the content of the video (or any other multimedia) is described and associated with the content itself, for example to allow fast and efficient searching in the material.

MPEG-7 uses XML to store metadata, and it can be attached to a timecode in order to tag particular events in a stream. Although MPEG-7 is independent of the actual encoding technique of the multimedia, the representation that is defined within MPEG-4, i.e. the representation of audio-visual data in terms of objects, is very well suited to the MPEG-7 standard.

MPEG-7 is relevant for video surveillance since it could be used for example to tag the contents and events of video streams for more intelligent processing in video management software or video analytics applications.

6.12 MPEG–21

MPEG-21 is a standard that defines means of sharing digital rights, permissions, and restrictions for digital content. MPEG-21 is an XML-based standard, and is developed to counter illegitimate distribution of digital content. MPEG-21 is not particularly relevant for video surveillance situations.

7. More on MPEG compression

MPEG-4 is a fairly complex and comprehensive standard that has some characteristics that are important to understand. They are outlined below.

7.1 Frame types

The basic principle for video compression is the image-to-image prediction. The first image is called an I-frame and is self-contained, having no dependency outside of that image. The following frames may use part of the first image as a reference. An image that is predicted from one reference image is called a P-frame and an image that is bidirectionally predicted from two reference images is called a B-frame.

> I-frames: Intra predicted, self-contained
> P-frames: Predicted from last I or P reference frame
> B-frames: Bidirectional; predicted from two references one in the past and one in the future, and thus out of order decoding is needed
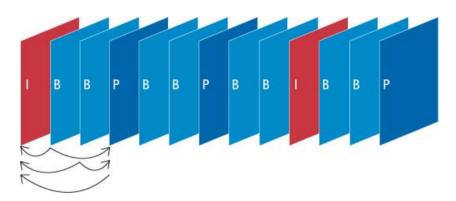

Figure 5. *The illustration above shows how a typical sequence with I-, B-, and P-frames may look. Note that a P-frame may only reference a preceding I- or P-frame, while a B-frame may reference both preceding and succeeding I- and P-frames.*

The video decoder restores the video by decoding the bit stream frame by frame. Decoding must always start with an I-frame, which can be decoded independently, while P- and B-frames must be decoded together with current reference image(s).

7.2 Group of Pictures

One parameter that can be adjusted in MPEG-4 is the Group of Pictures (GOP) length and structure, also referred to as Group of Video (GOV) in some MPEG standards. It is normally repeated in a fixed pattern, for example:

> GOV = 4, e.g. IPPP IPPP ...
> GOV = 15, e.g. IPPPPPPPPPPPPPP IPPPPPPPPPPPPPP ...
> GOV = 8, e.g. IBPBPBPB IBPBPBPB ...

The appropriate GOP depends on the application. By decreasing the frequency of I-frames, the bit rate can be reduced. By removing the B-frames, latency can be reduced.
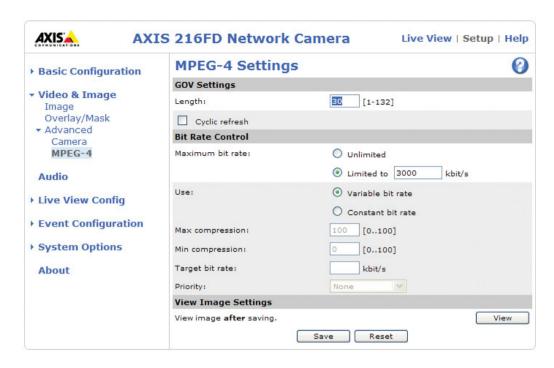
Figure 6. *An interface in a network camera where the length of the Group of Video (GOV), i.e. the number of frames between two I-frames, can be adjusted to fit the application.*

7.3 Variable and constant bit rates

Another important aspect of MPEG is the bit rate mode that is used. In most MPEG systems, it is possible to select the mode, CBR (Constant Bit Rate) or VBR (Variable Bit Rate), to be used. The optimal selection depends on the application and available network infrastructure.

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

With VBR, a predefined level of image quality can be maintained regardless of motion or the lack of it in a scene. This is often desirable in video surveillance applications where there is a need for high quality, particularly if there is motion in a scene. Since the bit rate in VBR may vary--even when an average target bit rate is defined--the network infrastructure (available bandwidth) for such a system needs to have a higher capacity.

8. MPEG comparison

Looking at MPEG-2 and later standards, it is important to bear in mind that they are not backwards compatible, i.e. strict MPEG-2 decoders/encoders will not work with MPEG-1. Neither will H.264 encoders/decoders work with MPEG-2 or previous versions of MPEG-4, unless specifically designed to handle multiple formats. However, there are various solutions available where streams encoded with newer standards can sometimes be packetized inside older standardization formats to work with older distribution systems.

Since both MPEG-2 and MPEG-4 covers a wide range of picture sizes, picture rates and bandwidth usage, the MPEG-2 introduced a concept called Profile@Level. This was created to make it possible to communicate compatibilities among applications. For example, the Studio profile of MPEG-4 is not suitable for a PDA and vice versa.

Note: MPEG-2, MPEG-4 and H.264 are all subject to licensing fees.

9. Conclusion – Still pictures

For single still pictures both JPEG and JPEG 2000 offers good flexibility in terms of picture quality and compression ratio. While JPEG 2000 compress slightly better then JPEG, especially at very high compression ratios, the momentum of the advantage compared to the price to pay for the extra complexity, makes it a less preferred choice of today.

Overall, the advantages of JPEG in terms of inexpensive equipment both for coding and viewing make it the preferred option for still picture compression.

10. Conclusion – Motion pictures

Since the H.261/H.263 recommendations are neither international standards nor offer any compression enhancements compared to MPEG, they are not of any real interest and is not recommended as suitable technique for video surveillance.

Due to its simplicity, the widely used Motion JPEG, a standard in many systems, is often a good choice. There is limited delay between image capturing in a camera, encoding, transferring over the network, decoding, and finally display at the viewing station. In other words, Motion JPEG provides low latency due to its simplicity (image compression and complete individual images), and is therefore also suitable for image processing, such as in video motion detection or object tracking. Any practical image resolution, from mobile phone display size (QVGA) up to full video (4CIF) image size and above (megapixel), is available in Motion JPEG.

However, Motion JPEG generates a relatively large volume of image data to be sent across the network. In comparison, all MPEG standards have the advantage of sending a lower volume of data per time unit across the network (bit-rate) compared to Motion JPEG, except at low frame rates. At low frame rates, where the MPEG compression cannot make use of similarities between neighboring frames to a high degree, and due to the overhead generated by the MPEG streaming format, the bandwidth consumption for MPEG is similar to Motion JPEG.

MPEG–1 is thus in most cases more effective than Motion JPEG. However, for just a slightly higher cost, MPEG–2 provides even more advantages and supports better image quality – comprising of frame rate and resolution. On the other hand, MPEG-2 requires more network bandwidth consumption and is a technique of greater complexity. MPEG–4 is developed to offer a compression technique for applications demanding less image quality and bandwidth. It is also able to deliver video compression similar to MPEG–1 and MPEG–2, i.e. higher image quality at higher bandwidth consumption.

If the available network bandwidth is limited, or if video is to be recorded at a high frame rate and there are storage space restraints, MPEG may be the preferred option. It provides a relatively high image quality at a lower bit-rate (bandwidth usage). Still, 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 Motion JPEG.

Looking ahead, it is not a bold prediction that H.264 will be a key technique for compression of motion pictures in many application areas, including video surveillance. As mentioned above, it has already been implemented in as diverse areas as high-definition DVD (HD-DVD and Blu-ray), for digital video broadcasting including high-definition TV, in the 3GPP standard for third generation mobile telephony and in software such as QuickTime and Apple Computer's MacOS X operating system.

H.264 is now a widely adopted standard, and represents the first time that the ITU, ISO and IEC have come together on a common, international standard for video compression. H.264 entails significant improvements in coding efficiency, latency, complexity and robustness. It provides new possibilities for creating better video encoders and decoders that provide higher quality video streams at maintained bit-rate (compared to previous standards), or, conversely, the same quality video at a lower bit-rate.

There will always be a market need for better image quality, higher frame rates and higher resolutions – with minimized bandwidth consumption. H.264 offers this, and as the H.264 format becomes more broadly available in network cameras, video encoders and video management software, system designers and integrators will need to make sure that the products and vendors they choose support this new open standard. And for the time being, network video products that support several compression formats are ideal for maximum flexibility and integration possibilities.

11. Acronyms

The following is a description of the acronyms used in this white paper.

Blu-ray – A high-density optical disc format for the storage of digital media, including high-definition video.

CCIR 601 – A standard for digital video for picture size of 720 x 485 at 60 interlaced pictures per second or 720 x 576 at 50 interlaced pictures per second.

CIF – Common Intermediate Format. Video of picture size 352 x 288 at 30 pictures per second.

DVD – Digital Versatile Disc. A standard to store digital audio and/or video on a CD-sized disc.

HD DVD – High Density DVD, or High-Definition DVD. A rival format to Blu-ray.

HDTV – High-Definition Television. A standard for television with significantly higher resolutions than traditional formats. For example, 1920 x 1080 pixels at 30 pictures per second.

IEC – International Electrotechnical Commission. International Electrotechnical Commission. An international standards and conformity assessment body for all fields of electro technology. Homepage at: http://www.iec.org.

Interlaced – A technique used in television system where the picture is divided into two half pictures containing every other line each. When displayed, first the odd lines are displayed then the even lines followed by the odd lines of the next picture and so on. This is the opposite of Progressive Scan.

ISO – International Standards Organization. A worldwide federation of national standards bodies from some 140 countries. Homepage at: http://www.iso.ch.

ITU – International Telecommunications Union. An international organization within the United Nations System where governments and the private sector coordinate global telecom networks and services. Homepage at: http://www.itu.int.

JPEG – Joint Photographic Experts Group. The committee responsible for developing the JPEG and JPEG 2000 standards. Homepage at: http://www.jpeg.org.

JVT – Joint Video Team. A partnership between ISO/IES and ITU that has developed the H.264 video codec standard.

MPEG – Motion Picture Experts Group. The committee responsible for developing the MPEG standards. Homepage at: http://www.chiariglione.org/mpeg/.

NTSC – National Television Standards Committee. This is the standard for the analog television format used in for example in Japan, and the US. NTSC specifies 525 lines at near 60 pictures per second, i.e. Interlaced video.

PAL – Phase Alternating Line. This is the standard for the analog television format used, for example, in Europe with 625 lines at 50 half-pictures per second, i.e. Interlaced video.

Progressive Scan – Each picture in the video sequence is the full picture displayed all in once. This is the opposite of Interlaced.

QCIF – Quarter CIF. Video of picture size 176 x 144 at 30 pictures per second.

VCEG – Video Coding Experts Group. A group within ITU that has developed for example the H.261 and H.263 recommendations for video-conferencing over telephone lines.

XML – Extensible Markup Language. A general-purpose markup language that supports a wide variety of applications.

AXIS
COMMUNICATIONS

About Axis Communications

Axis is an IT company offering network video solutions for professional installations. The company is the global market leader in network video, driving the ongoing shift from analog to digital video surveillance. Axis products and solutions focus on security surveillance and remote monitoring, and are based on innovative, open technology platforms.

Axis is a Swedish-based company, operating worldwide with offices in 18 countries and cooperating with partners in more than 70 countries. Founded in 1984, Axis is listed on the OMX Nordic Exchange, Large Cap and Information Technology. For more information about Axis, please visit our website at www.axis.com.

AXIS®
COMMUNICATIONS