JPEG2000, MJPEG, MPEG and H.264 in the security environment.

(When 25FPS is only 2FPS).

By Christopher Berry.

Introduction.

Digital recording of CCTV images and audio is gaining popularity amongst installers and users of large scale CCTV systems. This paper seeks to address and give advice on the various standards for compression, including JPG, MJPEG, MPEG4(H264) and Wavelet (JPEG2000).

The advice given should be taken into consideration when defining and specifying CCTV recording technology, transmission and viewing methods and when there is a potential use in Closed Circuit TV in a court of law.

Thanks is given to the Motion Picture Experts Group (MPEG) standards committee, JPEG2000 group and the US DOJ who have provided valuable assistance in the preparation of the document.

Technical introduction.

For some time CCTV has proven to be valuable in the general fight against crime, courts have long accepted video evidence as way of proving or disproving events as they occurred. Many existing CCTV systems are analogue and they are gradually being converted to digital, but without careful thought this process will actually cause a reduction in quality and consequently usability.

Traditional analogue CCTV, like any analogue system, uses considerable resource, another name for resource is cost. Therefore, do not be fooled, the conversion to digital is not an improvement in quality[1] it is a drive towards better usability and cost savings.

In the traditional analogue world image and audio is encoded into an electronic waveform of fixed length, as captured by a camera and then displayed on a “television” monitor – The Cathode Ray Tube (CRT).

The CRT (LCD[2]) monitor is made up of a number of horizontal display lines and a scanning device has the beam of electrons write the lines one at a time to create a picture. When the beam has sprayed all of the lines, the viewer sees one still frame of a video picture. The illusion of motion is then created by repeating this process a number times each second, that figure varies whether you are using either PAL or NTSC format.

1 An image is digitised and it retains quality consistently.
2 LCD Liquid Crystal Display, does not work in quite the same way as a CRT but it is mentioned simply because they exist.

Each of the frames is a still image, but each shows a progressively different stage of the motion.

It's really like watching a slide show in fast motion. You're seeing potentially 30 stills every second, but they blur together in your mind's eye to produce the illusion of motion. This little trick is called "persistence of vision." Without it, neither motion pictures nor video would exist.

The scanning electron beam starts at the top left of the picture tube and writes one horizontal line. Then when it reaches the right hand side of the picture or raster area, the beam drops down and writes the next line from left to right. In the early television systems, this process of writing lines for each frame created noticeable flickering. To minimize the flicker, engineers developed a system of "interlaced scanning", that is each second line is written and when the bottom is reached the beam restarts at the top.

The waveform requires considerable "resource".

Analogue CCTV recordings require high volumes of resource and this translates into having to use co- axial cable, fibre optic links, and CCTV VCR recording tapes. For example, a VCR tape is a typical high volume resource. A single tape can only record a few hours (minutes) of CCTV evidence yet is expensive to manufacture - plastic, metal etc. Plus VCR tape is easily damaged and has a finite life (rapidly degrades).

Volume is expensive.... the more you need the more you must be prepared to pay.

Digitisation of the CCTV image is therefore becoming increasingly popular. This process takes the analogue waveform created by the image capture device (camera) and converts it to a stream comprising a digital representation of the image. This stream of numbers when reassembled gives an accurate representation of the image waveform.

This digitisation is done at a very high sampling rate and is considered a lossless process[3].

We still have a problem with volume, we still need large amounts of resource to transmit and store that digitised image waveform.

In order to reduce the amount of Bandwidth used by video (and audio) we can take the digitised waveform and pass it through a numeric software data compressor. This compression software is known as a codec and it is used in every day life. Codecs allow digital data to be compressed and decompressed with the intention that the bit for bit result is as near perfect as possible.

Every telephone call you make has been through a codec. The television you watch (provided its digital TV) has been through a codec. You may already be familiar with the everyday names of various forms of codec- JPG is the most common one; this one is used when storing images on a digital camera for instance its name being a shortened version of the group that created it JPEG (Joint Picture Experts Group).

The use of codec and the various versions of them becomes critical in CCTV, to save money there is a temptation to use less recording media (hard disk space) and also maximise network resource where that resource may not have been planned or installed correctly.

3 Digitisation is never lossless, because no matter how high the sampling period some data is always discarded, however for the purpose of this document an ADC (analogue to digital converter) is assumed to create a bitmap.

Cost will therefore be a factor in deciding which codec to use, but, without careful consideration of the design and implementation irreparable damage to the CCTV waveform can be made.

One of the considered critical factors in CCTV is the **Frame Rate (FPS)**. In the USA, composite video is at approximately 30 frames per second, whilst in Europe composite video is 25 frames a second.

This has become the standard set by many consultants when specifying a CCTV system.

Most humans only require to process 16 frames per second in order to "see" motion video. The so called "persistence of motion" is effective at between 12 and 15 frames per second[4]. Composite frame rates (30/25) are set not by the necessity to fool the human eye into believing it is viewing real life, but simply because electronics gives us that rate at the frequency of electronic equipment and is traditional/historic.

Frames are often quoted when a recording system is specified without any mention of the codec to be used. In some circumstances *frame rate* is specified but *image size* is not.

Both are as important as each other, but even more so is compression. Every consultant should take care to specify the following, in this order.....compression (codec), image size and frame rate. Many ill equipped CCTV consultants miss one or more of these.

These, therefore, are the critical points that we need to discuss.

Let's first look at the codecs.

4 Moorfields Eye Hospital.

Codec.

A JPEG(Peg) codec is one that takes the digitised image frame by frame and compresses it. The compression is controllable based on various factors, but typically JPEG can compress a digitised image by a factor of 10:1 without any significant loss of quality. A JPEG image is considered nearly lossless if settings for clarity are maintained in preference to size (compression).

The newer JPEG2000 codec also improves security by enabling a complex set of markers or watermarks and can compress to an even higher level without significant loss.

An MPEG codec is one that takes the digitised image and compresses it based on a sequence of framing cycles. For example the first digitised image is compressed in a similar way to that used by JPEG but subsequent images are compressed in such a way as to avoid the necessity to send the complete frame, we simply send a representation of the image with groups of pixels to change.

This is achieved by taking the first frame (key frame) and compressing it. Then a complex algorithm is passed over the following frames which results in a significant reduction in the amount of data required to regenerate the frame. Every so often a new key frame is generated, this is based on information found in the particular invocation of the codec and the amount of change in the image (the length of time to process the frame).

MPEG2 was the first commercially used codec from the Moving Pictures Expert Group. MPEG4 is an improvement on MPEG2 giving better compression and more clarity.

H264 is the current version of MPEG4 and it uses more advanced calculations on the image stream prior to compression. Including analysis of frames both before and after the frame of interest.

MPEG/H264 codecs are considered a lossy compression method. You will have no doubt seen the effects of MPEG compression in everyday life when watching normal digital television, for example sports coverage or live event coverage when the bandwidth available is restricted the image sometimes pixelates into blocks and then gradually recovers as a new key frame is transmitted.

But, what are the effects of this on Recording and Viewing?

Recording.

Significantly, the author's view is the that main issue is one of clarity of image yet we are generally preoccupied with frame rate. An image that is a fleeting display on a monitor is compromised because we want to see it move smoothly. We forget that after that millisecond of activity we might actually need to see in clear detail frame by frame an event: a second can be a lifetime.
Because of the varied nature of the codecs in use in MPEG and H264 there is a huge difference in the results. Processor type, bandwidth, implementation of the standard are all key points.

The MPEG group had this to say about the use of these codecs in Security.

> "Generally speaking it is an unsafe policy to use these codecs for highly sensitive applications. For example, when taking a recording of an event and requiring a high resolution single frame (as might be the requirement in many circumstances) the possibility of achieving this with MPEG/H264 is unlikely – street scenes, shopping mall and anywhere single frame close identification is required should NOT be recorded using MPEG/H264 codecs. It should be noted that the MPEG committee is mainly concerned with video and audio for entertainment".[5]

In 2002 a test showing the effects of various factors governing the use of MPEG and its granularity in the production of still images demonstrated the significant differences between a still frame captured from a VCR vs those from JPG and MPEG streams. This results in justification of the JPG2000 standard- this is used in medical applications and in those that require clarity of image.

The JPEG2000 group has this to say about its codec.

> "Traditional surveillance technology has been quite slow to embrace the advantages of digital image processing. In part this has been because the sheer volumes of data have required analogue storage methods such as lapsed time video recorders, and in part because the cost of moving to a digital base have been prohibitive. In the last few years however, costs have fallen dramatically, whilst processing power and capabilities have improved equally fast. This allows many of the shortcomings of traditional surveillance applications to be addressed, whilst also considering many of society's concerns about privacy and intrusion.
>
> Movement detection, and many more sophisticated forms of image analysis can be coupled with new sensor technology to allow much more pro-active monitoring and alarms. Use of 'region of interest' enhancement allow accurate identification of suspects within single frames while excluding from analysis, and subsequent public exposure, the innocent bystander. Tight control can be exerted over the user of surveillance technology - for example showing sufficient details to allow recognition of an individual found to have passed a stolen cheque, whilst not permitting enough detail to allow a corrupt viewer of the scene to be able to view and copy a signature being made.

5Munich 2004.

> The need for stored evidence to be of a sufficient high quality however also raises concerns and a need for protection against tampering and fabrication. It is very easy within the digital environment to change either subtly or completely aspects of an image, and the meta data surrounding it. Techniques such as encryption and watermarking can be used to help protect against this risk, but there is a real need for well accepted media management techniques which can help reduce risks in this area, for example using trusted third parties and crypto-technology. In addition, it is important that evidence is not segmented, being kept in a single file to avoid the obvious risks of mis-information."

MPEG codecs are also poor when image enhancement is used, for example, electronic zoom.

Additionally, the same 2002 test demonstrated some blurring in high speed applications when images are viewed in slow motion[6] on MPEG in relation to JPEG images.

Of course this is a generality and typically the higher cost recording devices support better performing codecs than the lower cost versions. In addition a codec that allows you to define the expected bandwidth is considered superior to one that has no such feature. The bottom line is that not all codecs are equal.

Despite manufacturers' claims it is ***impossible*** to record at 30/25 frames per second using MPEG/H264 as a considerable amount of the original data frame by frame is lost. At best this can only be described as a "virtual" frame rate. Typically a new full key frame is created every 15 frames, this gives us true recording ratio of 2 frames per second on even the highest bandwidth devices - it is possible to fool the human eye but the mathematics are clear.

The frame rate specified by many consultants is often used as a minima acceptable. Little if any regard is made of clarity and evidencial usability. The authors view is that it is always preferable to record in MJPEG (moving JPEG) format or JPEG2000 when evidence has to be clear and each frame is part of the vital evidence.

However, there are some significant issues in recording in MJPG/JPEG2000 format. The biggest one is volume, the processor used in the recording device also has to be capable of processing the images quickly and there is a requirement for more storage media (more recording volume). The consequence is increased cost.

But, in danger of over repetition, should cost be part of the critical decision making process when specifying a sensitive CCTV system?

MPEG is designed specifically to mimic the firing of electrons at a CRT so that the human eye is fooled. However, in the security industry we are not viewing recorded images for entertainment. In a high risk security environment the quality has to be as near to that found in the original analogue waveform as possible.

6This blurring is independent of the shutter speed selected in the camera.

Live viewing/Transmission.

It is the exact opposite of the above when we are live viewing the images or transmitting them for immediate evaluation of an event. In this circumstance MPEG/H264 is considered a better solution.

Responders work better when the scene they are viewing is closer to the real event timing. It becomes easier to pan and tilt cameras and it is significantly easier to follow interesting targets. This is difficult to achieve with JPG compression simply because more bandwidth and even more processing more is required.

Its far easier to fool the human with motion than it is by critical examination of a still image.

Conclusion.

"Volume" is the important factor in this discussion. The MPEG/H264 codecs are designed specifically to save bandwidth and whilst improvements have been made, the loss of image quality, particularly where enhancement of single frames is required, is significant and therefore precludes its use in high security recording.

Bandwidth[7] should never be used as a reason for using either MPEG or H264 codecs for recording video in a sensitive security environment. ***Unless*** the following criteria can be met.

1. Minimum target size recorded should be no less than 25% of the recorded image.
2. Lighting can be guaranteed to be at least ten times that given by the camera manufacturer as the low light level for operation (MPEG4 codecs do not perform well in low light levels).

If cost ***is*** a critical factor then In order to save costs of recording medium It is preferable to record at a lower frame rate (15FPS) using the JPG codec than at supposed higher rates using MPEG/H264; remembering that a human can only need process 15/16 frames per second in any case for live motion to be imagined.

One point that is very often missed when defining a recording system is image size. This is confusing for the end user and again is very often misunderstood.

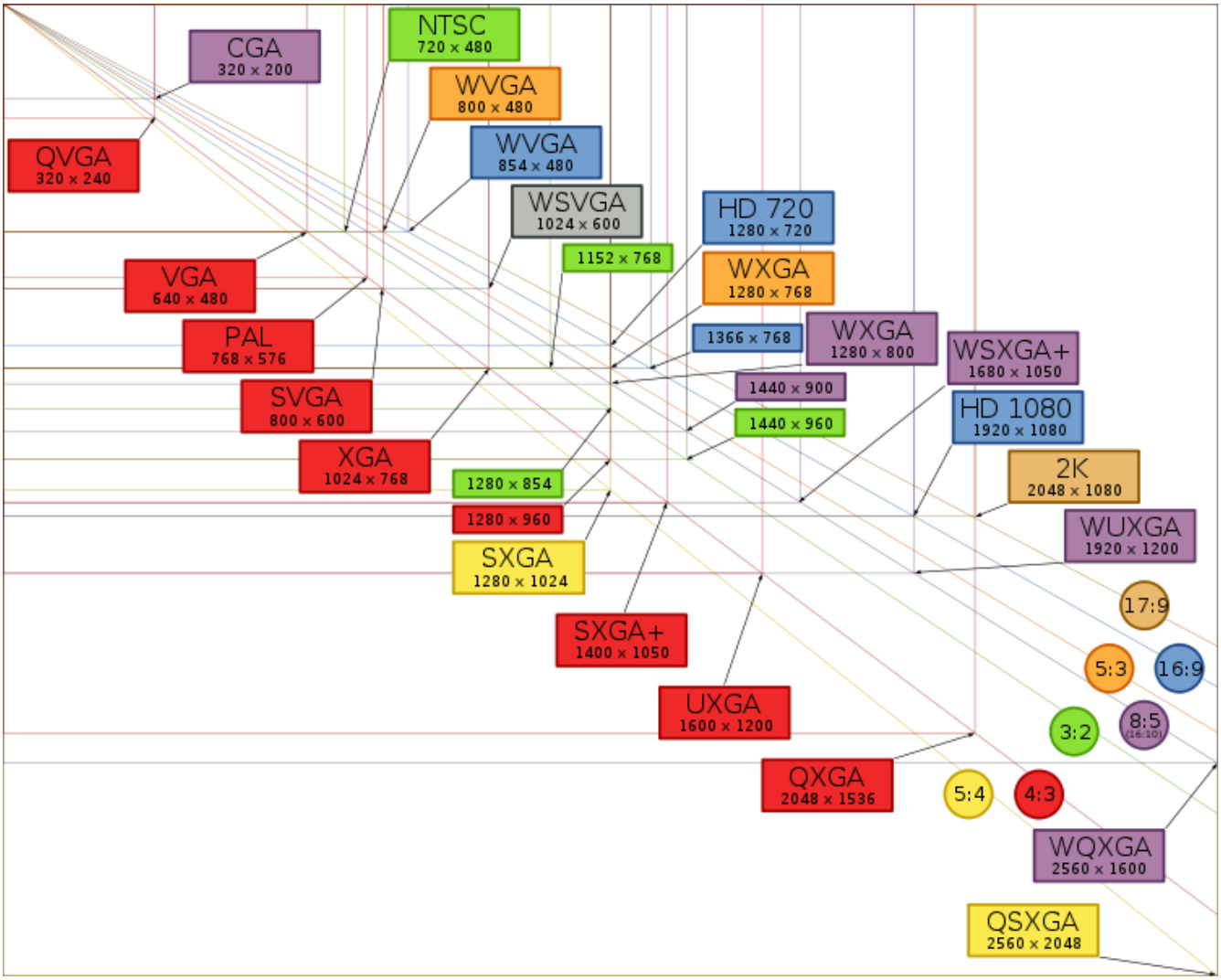
In a sensitive recording application images should be recorded at a minimum of 768*576[8] giving us the equivalent of the PAL format, whereas, in normal day to day "general" recording can be at 640*480[9] .

Of course Mega-Pixel cameras give us the opportunity to record in higher resolutions and these should be considered in many circumstances.

7 Bandwidth, Storage/transmission costs.

8 Some IP cameras/encoders offer a slightly lower pixel ratio to the encoded PAL composite video, however, this is a red herring as the top two lines and bottom one of a PAL encoded video are lost giving us a lower overall pixel ratio for a composite analogue signal when compared to a true IP encoded signal.

9 PAL/NTSC discussion relating to true image size D-PAL and D-NTSC are not covered by this document.


CGA
320 × 200
NTSC
720 × 480
WVGA
800 × 480
QVGA
320 × 240
WVGA
854 × 480
WSVGA
1024 × 600
HD 720
1280 × 720
1152 × 768
WXGA
1280 × 768
VGA
640 × 480
1366 × 768
WXGA
1280 × 800
WSXGA+
1680 × 1050
PAL
768 × 576
1440 × 900
HD 1080
1920 × 1080
SVGA
800 × 600
1440 × 960
XGA
1024 × 768
2K
2048 × 1080
1280 × 854
1280 × 960
WUXGA
1920 × 1200
SXGA
1280 × 1024
17:9
SXGA+
1400 × 1050
5:3
16:9
UXGA
1600 × 1200
3:2
8:5
(16:10)
QXGA
2048 × 1536
5:4
4:3
WQXGA
2560 × 1600
QSXGA
2560 × 2048

Comment.

Generally the security industry appears to be accepting H264(AVC) as the de-facto current standard for both recording and live viewing. Comments similar to “it is what we see on HD TV so it must be acceptable in the security industry” are prevalent, however, if you actually look closely at your HD TV you will notice that TV programming generally has the image size (the viewing of characters) at above 25% of the viewable screen.

You need also to consider that the recorded DVD entertainment “film” is not dependent on bandwidth/volume or cost and cannot be used as a general comparator.

It is generally impossible to play back single frames in MPEG, simply because those individual frames are poor quality. Instead slow motion viewing is simply the playback of the key frames (2FPS) or “snatched frames” - a frame composed of the original key frame and the changes. The changes are dependent on the algorithm.

Bear in mind a human being can rotate a full 360 degrees in somewhat less than one second. So it is likely that a person's image, whose identification may be required, could be compromised because we are not seeing all of the available data. The clear images of a person or persons could be available only in one or two *real* frames. The chance of losing that evidence in an MPEG stream is simply too great for its use to be accepted.

To repeat, where a manufacturer is claiming single frame playback from an MPEG based system you need to be extremely careful that they are not simply showing key frames or making a “snapshot” of the virtual image. Indeed if they are claiming real playback it is likely that they are using increased numbers of key frames (high motion) which makes the argument for higher compression pointless.

Luckily many major IP camera manufacturers are members of the MPEG committees and they understand these factors and provide dual streams. JPG(JPG2000) for recording and MPEG(H264) for viewing (fooling the human eye).

And this is how it should be in a high security environment (you decide what constitutes high security), record in JPEG/JPEG2000 but display in MPEG4/H264 so that the operators can view in real time.

Do not make the assumption that “what you see is what you get” - that may be so when you are watching your next sports event, rather your brain tells you its seen everything, but, in our environment it is that **one frame** that potentially captures your bad guy - so it had better be a good one.

It is also doubted that a critical examination of an MPEG compressed CCTV stream could pass in a court of law, simply because the lossy nature of the codec. In fact we can say better than that.....MPEG can be contested[10].

It has long been the held belief of most western courts of law that evidence (visual or otherwise) should be an accurate representation of the original event. Anything that materially alters the actual original image (the compression technique) could be considered altering that original evidence. Both MPEG and JPEG committees freely accept this situation and is why the JPEG2000 committee specifically included strong watermarking in the specification – you have to show that your single frame has not been messed with.

10 US DOJ

I have deliberately avoided Mega-pixel conversation, but this will eventually overcome its cost /performance issues (Moores Law) and the features of MPEG4/H264 will be an advantage when viewing and manipulating early use of the image or to view intent.

Ends.

References.

PD6777, BSI's Guide to the practical implementation of JPEG 2000.
MPEG committee various published documents.
US DOJ analysis of evidence presented in court.

About the Author.

Christopher Berry has been involved in CCTV since 1982 and is the founder of Initsys Ltd.
Initsys Ltd has several recording and viewing products all of which support the codecs described here.