Why digital? by Vlado Damjanovski

For a non-technical person involved in CCTV, the question "Why digital?" is a pretty fundamental and important one. Although for many this would be a boring subject, explained many times and in many places by now, it would be unfair not to explain the basic concept of digital video here. This would be a good starting point for our issue devoted to DVRs because CCTV has a slightly different use than the digital video in broadcast TV.

One of the most important differences between an analogue and a digital signal, apart from the form itself, is the immunity to noise. A digital signal, having an electronic form, is also affected by noise, as is the analogue. Digital signal however, can only have two values: zeros and ones. Noise also affects digital signals, but the difference is it will only be detected when the noise values reach levels that may interfere with the digital circuit margins deciding whether a signal is zero or one. This means, digital signals can tolerate noise to an extent unimaginable with analogue video signals. As a result, this means longer distances, high immunity to external EMIs and no signal degradation, i.e., better picture quality.

The other important advantage of a digital video signal is the possibility for computer processing and storage. Once the video signal is converted into digital format it can be processed in many different ways, depending on the algorithms built in the processor or software. Some typical digital image processing include digital zooming, sharpening, video movement detection, edge detection, noise reduction, equalisation, time/date and camera ID stamping, etc. There is no difference in image quality between the copies and the original. Most importantly for CCTV applications, digital images can easily be encrypted and protected against tampering.

Let's first clarify where can we expect analogue and where digital video signals in CCTV.

Majority of CCTV cameras are still analogue. Many do have internal Digital Signal Processing circuitry (DSP), but usually, at the BNC output, we always have a composite analogue video signal (often called CVBS, or CVS). This is the popular "1Vpp" ("one volt peak to peak" - meaning one volt from the bottom of the synchronisation pulses to the maximum white level clipping).

It is fair to say though, that in the last year or so, some manufacturers have started offering cameras with digital, serial outputs. Such cameras are usually TCP/IP addressable (using the same transfer protocol as when you connect your PC to your ISP). These are still early days for such cameras, and video performance is usually lower to what an analogue CCTV camera will give you, either with lower resolution, or lower frame rate. However, they do have an advantage of having the signal converted into digital, bypassing the need for a frame grabber in your PC, a piece of equipment used to convert the analogue signal to digital (A/D).

There are alternatives to the latter, where an existing analogue camera can be connected to a stand-alone box that converts the analogue signal to digital, and this then becomes the TCP/IP addressable box.

Monitors and other display devices used in CCTV are analogue devices. Computer monitors, using S-VGA connectors, are also analogue, although if USB or FireWire connection is used, they can be called semi-digital. The display however, by nature, is always analogue, composed of continuous luminance and colour change (not zeros and ones).

Recording and storage devices, at this stage, are mostly analogue (usually using VHS or S-VHS recording formats), but, as we know, this is radically changing in the last two years. Digital Video Recorders (DVRs) are getting more and more popular and this issue of CCTV focus is dedicated to explaining some basic digital recording concepts, as well as test-comparing of some DVRs offered for our evaluation.

Last, but not least, advantage of digital video is that once the analogue video is converted into digital stream of data and entered into a computer environment it can easily be transferred, shared and used over any Local Area Network (LAN) or Wide Area Network (WAN). This re-defines the concept of CCTV installation, as we know it today. Although using the "good-old" coaxial cable will still be required, there will be more and more designs and CCTV system solutions based on the Information Technology (IT) concepts. Certainly, even with

LANs and WANs the distances and network bandwidth will need to be considered carefully. There will always be limitations and costs associated with larger and complex systems. It is very likely that many digitally designed CCTV systems will require a LAN on their own, not only because of the bandwidth, but also because of the security of closed LAN rather than an open one.

Why compression?

In order to understand the reason and the concept of various image compression techniques, let's work out the size of a typical high resolution digitised video.

If we take a high resolution colour camera in CCTV, the number of picture elements, or pixels, that would be "frame grabbed" using the CCIR-601 for PAL standard recommendation (NTSC is calculated similarly), we would have:

720 × 576 picture elements = 414,720 pixels

This is valid for 1 TV frame = 2 fields. Each of these pixels is represented by 8-bits of luminance quantisation, i.e., 28 = 256 levels, thus 414,720 pixels luminance are represented by 414,720 bytes of information (important to remember is that 8 bits = 1 byte). If the image is in colour, there are 3 primary colours to be encoded with 256 levels, making 16 million colour combinations (this comes from 256^3 = 16,777,216).

So the total number of bytes required to digitally describe each of the possible three primary colours for every pixel is 414,720 × 3 = 1,244,160 bytes = 1,215kB.

This is a digital representation of a full PAL TV frame with no compression!

This file size could represent an uncompressed BMP file of one TV frame. BMP is an image format where every pixel is represented with a byte without compression. To put it simple, only one frame of uncompressed video fits on one 3.5" floppy diskette. But don't forget, we have (in PAL) 25 such frames every second.

Now, if we want a real time recording and playback of such files, for PAL we would have to have 1,215kB × 25 frames per second = 30,375kBs = 30MBs.

For NTSC such calculation will bring more or less the same number since we have more frames per second, but the vertical resolution is smaller.

Please note that this is 30MBs! Capital "B" means Bytes, small "b" means bits.

The data flow in computer language is usually measured in megabits per second (Mbps). Even if we accept that a byte is equal to 8 bits (but it can also be 16 bits, depending on the application) the 30 MBs will be equivalent to 240 Mbps.

This is a very high data flow

number indeed, even for today's standards. The maximum data flow we can have inside a computer is limited, usually, by the motherboard data bandwidth, but most of all by the hard disk mechanical limitations. The fastest hard disks we know of, using the ultra-wide SCSI, have theoretical maximum transfer rate of 40~60Mbps, although realistically this is half this value.

So, from the above it is clear that video images can hardly be recorded (stored) in their raw format. Even computers have bandwidth limitations and uncompressed data of a real time video is just too much.

The only practical solution would then be to compress the video images.

The following are the most common video compression types in CCTV today:

- MPEG-1 (also known as motion predictable, data between 1.5Mbps and 3.5Mbps)
- MPEG-2 (motion predictable, data between 3.5Mbps and 20Mbps)
- H.320 series (H.261; H.263, data between 64kbps and 1.5Mbps)
- JPEG and Motion-JPEG (still image DCT transformation)
- Wavelet (still image wavelet transformation)
- JPEG-2000 (A standardised Wavelet based compression)
- Others (proprietary compressions) such as Multi-Layer JPEG (ML-JPEG), Hybrid MPEG/ Wavelet compression, etc…

MPEG-1

MPEG works with motion pictures. MPEG-1 is targeted at processing real time images at bit rates from 1.5Mbps up to 3.5Mbps. This data speed is easily achievable, and was initially designed for laser disk speeds, hard disk and the CD-ROMs.

It is possible to have MPEG-1 running at higher bit rates, but the real intention is to have images of equal or better than VHS quality and up to the resolution quality of S-VHS.

An interesting information for audio lovers, using the popular MP3 format - this is practically MPEG-1's audio layer.

MPEG-2

MPEG-2 was accepted by the ISO in 1993.

The MPEG-2 standard specifies the coding formats for multiplexing high quality digital video, audio and other data into a form suitable for transmission or storage. So, MPEG-2, like MPEG-1, does not limit its recommendations to video only. MPEG-2 encoding can produce data rates between 3.5 Mbps and 20 Mbps.

MPEG-2 works with Intra pictures (I), Predicted pictures (P) and Bi-directional pictures (B), digitised with CCIR-601 sampling parameters of 704(H) X 576(V) pixels.

There are two MPEG-2 variants: MPEG-2 Main Profile@Main Level and MPEG-2 422Profile@Main Level.

The MPEG-2 MP@ML was originally developed for the distribution of video signals. For this reason, it is primarily used for DVB (Digital Video Broadcasting) and DVD (Digital Versatile Disc). These applications demand the best possible image quality at the lowest data rates possible (max.15 Mbit/s).

The interesting variation for us is the MPEG-2-422P@ML, which has been designed specifically for post-production and CCTV application requirements.

The main issues in CCTV are whether material can be edited, authenticated, encrypted, reverse playback of multiplexed video streams.

MPEG-2 422P@ML, "I-frame only" meets all of these demands: the use of "I-frame-only encoding" guarantees easy, frame-accurate control, and sampling at 4:2:2 provides excellent image quality. In addition, the standard used for decoding is independent of proprietary decoder solutions, making such MPEG variant a very attractive standard for CCTV.

More importantly (at least for CCTV), such MPEG-2 encoded video sequence can be paused, played-back in reverse and do anything else required for multiple camera recording.

H.320

This group of standards define streaming audio and video, similar to MPEG standards. The original standard was designed to support video running over switched digital circuits at bandwidth ranging from 112kbps to 1.5Mbps and was designed to enable video conferencing over narrowband ISDN (N-ISDN) networks. The H.320 standard is an umbrella standard that refers to numerous other standards to specify how certain required functionality is provided.

Video compression capabilities can be supplied in accordance with either the H.261 or the H.263 standards, depending upon factors such as available bandwidth and sampling rate.

H.261

The H.261 standard was specifically developed to serve as the video compression part of the H.320 videoconferencing standard. That means that it was optimised to compress video for transmission over ISDN lines that provide a range of bandwidth from 64kbps to about 1.5Mbps.

Like the MPEG standards, H.261 specifies formats appropriate for both storage and transmission of compressed video. Moreover, since bandwidth over ISDN is available in increments of 64kbps, the H.261 standard permits the compression options to be adjusted in a manner that increases required bandwidth in 64kbps increments to get higher video quality.

The data formats originally defined for use in this standard have become important for a number of other compression These formats specify the size (image resolution) of the compressed video image and the maximum frame rates.

The first of them is referred to as the Common Interchange Format (CIF) and specifies image resolution at 352X288 pixels or roughly half the resolution of a standard analogue PAL or NTSC TV image. There is a format for even lower resolution - Quarter CIF (QCIF) 176X144 pixels.

H.263

The H.263 standard was specifically developed to enable videoconferencing and storage in low-bandwidth environments but can also support high-bandwidth.

By incorporating a more efficient video-compression algorithm, the H.263 standard offers higher video quality than H.261 at every level of bandwidth, including ISDN.

Higher resolutions were added to exploit capabilities made possible by newer transmission and compression technologies. Sub-QCIF (SQCIF) permits video as small as 128 horizontal pixels by 96 vertical pixels to be transmitted. There are also larger image resolutions, four times and sixteen times the size of the CIF image-i.e., 704 horizontal x 576 vertical pixels and 1,408 horizontal x 1,152 vertical pixels respectively.

JPEG

JPEG is an ISO standardized image compression mechanism. It refers only to still digital images. Such still images in video can be either TV fields or TV frames.

There is a JPEG derivative in CCTV that some like to refer it to as Motion-JPEG. Motion JPEG does not exist as a separate standard but rather it is a rapid flow of JPEG images that can be played back at a sufficiently high rate to produce an illusion of motion.

JPEG works by transforming blocks of 8×8 picture elements using the Discrete Cosine Transformation (DCT). The DCT is based on Fourier transformation of time signals into frequency domain signals. The Fourier transformation is a very good method for analysing signals in frequency domain, the only problems is it always works with an assumption of time domain signals being periodical and infinite. This is not the case in reality and this is why an alternative to Fourier transformation was introduced in the sixties and this is called fast Fourier transformation (FFT).

Compression factors achieved with JPEG compression are quite high (well over 10 times) and the picture quality loss appears insignificant to the human eye. JPEG is designed to exploit the known limitations of the human eye, like the fact that fine chrominance details aren't perceived as well as fine luminance details in a given picture.

When a highly compressed image is enlarged a blocky appearance is evident. Compressions of up to 100× can be achieved.

Since JPEG files are independent of each other, when used in CCTV recording, they can be easily played back in reverse direction, playback speed can be increased or reduced and copied as single files or group of files.

Wavelet (and JPEG-2000)

For many decades, scientists have wanted more appropriate functions than the sines and cosines, which comprise the bases of Fourier analysis to approximate choppy signals. By their definition, sines and cosines are non-local functions (they stretch out to infinity). This is the main reason that they do a very poor job of approximating sharp changes, such as high-resolution details in a finite, two-dimensional picture. This is the type of picture we most often have in CCTV time-lapse recording, as opposed to a continuous stream of motion images in broadcast television.

Wavelet analysis works differently. With wavelet we can use approximating functions that are contained in finite domains. Wavelets are functions that satisfy certain mathematical requirements and are used in representing data or other functions in wavelet analysis. The main difference compared to the FFT analysis is that the Wavelets analyse the signal at different frequencies with different resolutions, i.e., many small groups of waves, hence the name Wavelets.

The wavelet algorithms process data at different scales or resolutions.

The wavelet analysis tries to see details and the global picture, or as some wavelet authors have said, "see the forest and the trees" as opposed to Fourier analysis which "sees just the forest."

Wavelets are well-suited for approximating data with sharp discontinuities.

Wavelet compression transforms the entire image as opposed to 8×8 sections in JPEG and is more natural as it follows the shape of the objects in a picture.

ISO is completing the proposal for standardising JPEG-2000 compression (ISO15444), based on wavelet theory. At low bit-rates, the performance of JPEG2000 outpaces that of the current JPEG.

JPEG2000 offers features, in an integrated code stream, that is not possible with the current JPEG image compression standard. Because individual tile components can be accessed from the code stream, different spatial regions of the image can be decoded and displayed. Clever error recovery strategy can be used to conceal errors in the bit stream and this makes JPEG2000 suitable for use over error-prone wireless channels.

Recording or transmission of only a particular region of an image, without the whole high-resolution set can be done.

Another important feature of the structure of JPEG2000 compression system is the ability to decode, from a single code stream, different spatial resolutions of the image. Using the same code stream, images with different fidelity up to lossless can be decoded.

Many expect JPEG2000 to become the accepted, universal file format for digital images, whether on the Web or from digital cameras, printers, faxes, remote sensors, wireless transmissions or CCTV.

D-VHS

This is a digital variation of the VHS format, which is now already available through JVC and Sanyo.

This is a new recording format that uses existing, and cheap S-VHS tapes.

It could be an attractive solution for CCTV, especially interesting for retrofitting existing multiplexer systems that use analogue VHS or S-VHS VCRs.

D-VHS uses virtually the same head mechanism as the existing VHS, but the interesting part is that a 180 minute tape can be used to record live video (PAL) for 12 hours. There are other, time-lapse modes, of 24, 48, 72, 96, 120, 168, 240, 360, 480, 720 and 960 hours.

The compression scheme used is JPEG.

The picture quality can be adjusted from low quality 16.8kB per field to high quality 520TVL frame mode with files size of 100kB.

DV format

DV is the digital video recording system used as a basis for the Mini DV, DVCAM™ (Sony) and DVCPRO™ (Panasonic) videotape formats. It uses Y:Cr:Cb = 4:1:1 encoding (luminance Y : chrominance red : Chrominance blue) and 5:1 Discrete Cosine Transformation (DCT) inter-field motion compression.

A DV tape can store up to 270 minutes of video in standard mode, which equates to 60GB of data. A 90 minutes Mini DV tape can store up to 60 minutes of video in standard mode, or 90 minutes in long play mode. The amount of video and audio data stored is equivalent to around 20 GB of data.

Both of these formats can be used with the Sony HSR series digital recorders. :-)