1998

THE QUEEN'S AWARDS
FOR ENTERPRISE
2001

BBC

R&D White Paper

WHP 034

July 2002

Colorimetric and Resolution requirements of cameras

A. Roberts

Research & Development
BRITISH BROADCASTING CORPORATION

Colorimetric and Resolution requirements of cameras

Alan Roberts

Abstract

This document is intended to provide some of the background colour-science justification for the "BBC setup" conditions in some television cameras. Although aimed primarily at HDTV cameras and their use in attempting to mimic and thereby replace film usage, the colour-science applies equally to all forms of video cameras.

Two appendices are included, giving details of scanning and compression standards. Details of measurements and preferred setup conditions for specific cameras are available for BBC use. Manufacturers may request access to information about their own cameras, and facilities houses about cameras used on BBC projects.

Key words: camera, colorimetry, gamma, knee, aperture correction, detail enhancement, film look, shuttering.

Colorimetric and Resolution requirements of cameras

Alan Roberts

There are several schools of thought about how television cameras should be set up. Most have a direct connection with the historic limitations of the electronics in early cameras, but some are inherited from other sources. Experiments with some recent HDTV cameras have produced set-up conditions that perform well under normal usage. Although this work concentrates on HDTV cameras, the principles involved are common to all television systems, and so similar set-up conditions could be derived for any camera that had sufficiently wide-ranging and flexible control.

The main objective in deriving setup conditions for cameras has been to achieve a certain “look” that is acceptable to programme-makers. Much of the effort been spent in attempting to mimic the performance of film cameras since HDTV is currently perceived as an alternative to film rather than as an alternative to standard-definition television. That does not preclude HDTV cameras used as replacements for SDTV cameras, but the ideal setup may be slightly different in that case.

The main body of this document investigates the theoretical requirements and historic limitations of camera design, in order to throw some light onto the reasons for current practices. To ensure that this investigation is not limited to specific cameras, it is important to list the relevant properties of cameras and explain their significance.

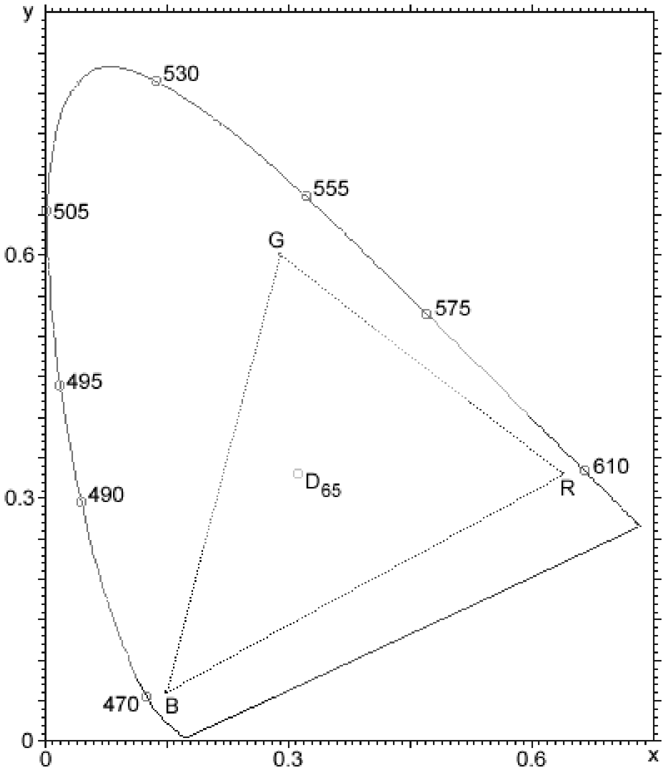
1 Colorimetric properties of cameras

The prime function of a video camera is to produce, at its output, signals which, when connected to a reference display, will cause it to reproduce the original scene to an acceptable accuracy. Thus the design and performance of the camera is defined by the display. For each television standard there is an internationally defined and agreed reference display. For the purpose of this document, only those for the PAL/SECAM/625-line[1] systems and for HDTV[2] (both 720- and 1080-line versions) will be considered.

1.1 Primary analysis

The traditional colour display comprises three electron guns in a cathode ray tube, each generating light in the phosphor screen as separate Red Green and Blue images. When equal drive signals are applied, the White Balance condition applies and the display makes a white colour, the “illuminant”. The image on-screen is seen as though lit by that illuminant. The chromaticity coordinates of those colours (primaries and white balance) are defined in the CIE1931 (xy) colour space. Conventionally, the triplet of primaries contains the whole colour gamut that the camera can analyse, any colour inside the triangle will produce a triplet of *RGB* signals, all positive or zero, which can drive an appropriate display to reproduce that colour accurately.

	PAL/SECAM 625-line		HDTV 720/1080-line	
	x	y	x	y
R	0.64	0.33	0.64	0.33
G	0.29	0.60	0.30	0.60
B	0.15	0.06	0.15	0.06
White (D65)	0.3127	0.3291	0.3127	0.3291



The primaries were chosen as a compromise between the need to display most real-world colours, acceptable brightness, and economy. The two sets of primaries are very similar, only the green is different. So a camera generating pictures for one set of primaries will not produce large colour errors if shown on a display using the other set.

The chromaticity plot above shows the EBU primaries, together with the locus of spectral colours, within which lie all visible colours. A standard colorimetric analysis[3] of these primaries derives the colour analysis matrices (simultaneous equation triplets), which connect the scene colour expressed in CIE 1931 tristimuli (*XYZ*) to the *RGB* drive signals.

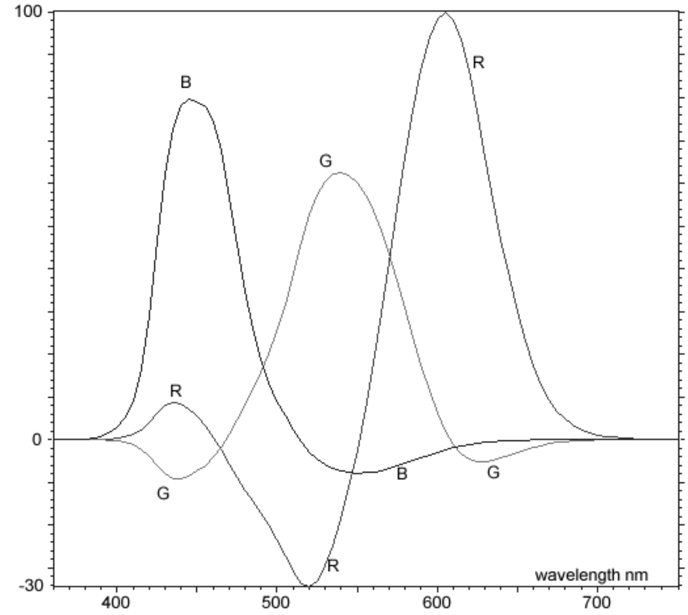
PAL/SECAM/625

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 3.2408 & -1.5373 & -0.4986 \\ -0.9692 & 1.8760 & 0.0416 \\ 0.0556 & 09.2040 & 1.0571 \end{bmatrix} . \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

HDTV (720/1080)

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 3.0632 & -1.3933 & -0.4758 \\ -0.9692 & 1.8760 & 0.0416 \\ 0.0679 & -0.2288 & 1.0693 \end{bmatrix} . \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

Passing the CIE responsivity curves (the response of the average, colour-normal, human eye)[4] for the "2°standard observer" through these equations yields the spectral taking-characteristics (colour-matching functions) that the camera must have in order to produce accurate colour analysis, shown here.

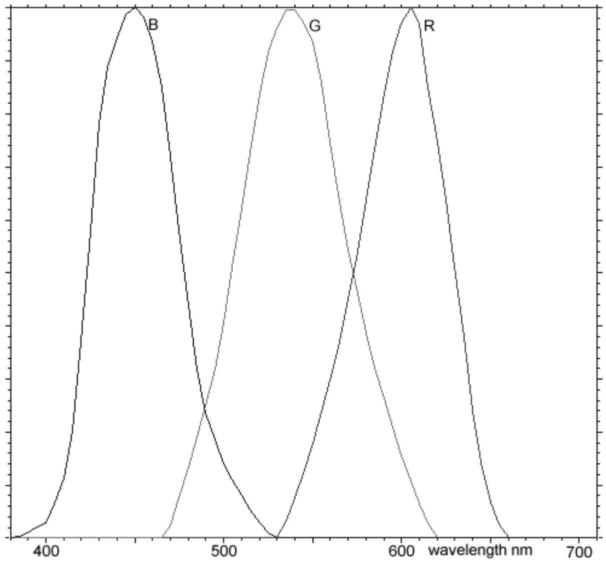


If the camera produces these amplitude responses to an equal energy illuminant (i.e. equal energy per unit of wavelength), then the same camera will accurately produce *RGB* signals for any scene colour to drive the specified display. Note that, at some wavelengths, the responses are negative. Clearly it is not possible to get such performance from any known light sensor, so some form of compromise is required and the camera no longer conforms to the "ideal", theoretical, specification.

1.1.1 Camera matrix

A "positive-lobes only" set of responsivities, such as the illustrated curves originally specified by the BBC, can be used effectively with a 3x3 linear matrix M in the camera to approximate to the ideal characteristics shown above, i.e. including negative lobes. The overlapping nature of these curves is essential for this purpose. In this way, the camera channel signals $R_cG_cB_c$ are derived from the camera sensor signals $R_sG_sB_s$:

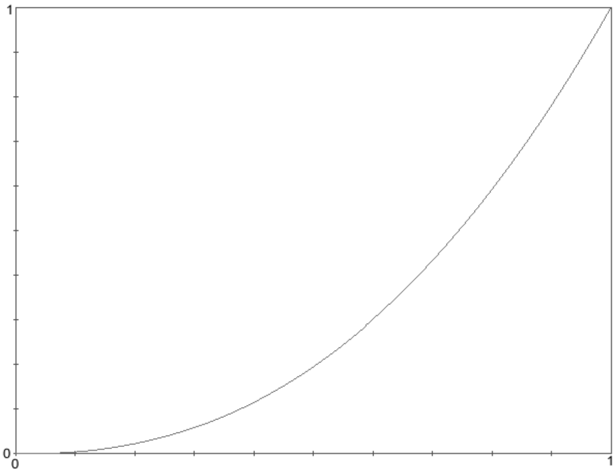
$$\begin{bmatrix} R_c \\ G_c \\ B_c \end{bmatrix} = [M] . \begin{bmatrix} R_s \\ G_s \\ B_s \end{bmatrix}$$



The coefficients of the matrix *M* are derived by an iterative mathematical process, which optimises the overall colour performance of the camera and an ideal display, according to parameters specified at the time of camera design. The result is a compromise between colour performance, sensitivity and noise. Note that this matrix should operate on the linear signals from the CCDs, and not on the gamma-corrected signals (see below). While it is possible to optimise a matrix that operates on the non-linear signals, it is generally much less successful, even though it is electrically easier to do so.

Different matrices can be used in the camera, to optimise the performance for use with any set of display primaries. The signals generated for the two sets of primaries given in Section 1.1 differ by only 4% at maximum, it is rarely worth transcoding between these sets of primaries.

1.2 Transfer characteristic (Gamma)



This is the relationship between incoming light and outgoing signal. If the display device were linear (i.e. $Light \propto Volts$) then this relationship would also need to be linear. However, the cathode ray tube (CRT) has a non-linear relationship (illustrated right), usually described as a power law:

$$Light \propto Volts^{\gamma}$$

where γ (gamma) has a theoretical value of 2.5. Practical limitations on the design of drive amplifiers tend to lower this figure to about 2.35. Accordingly, for the camera to reproduce the scene accurately, the camera should correct for this by applying a power law of 1/γ, about 0.45, to the camera signals. Unfortunately, this has never been feasible because a power law of less than unity has a slope of infinity at zero input, i.e. the gain of the “gamma-corrector” must be infinite at black. Such amplifiers cannot be built, and would produce excessive noise even if the were built. Therefore a compromise equation is used, with an offset in one or other quantity to limit the maximum slope to something achievable. Two common forms of such equations are the BBC law[5] illustrated right:

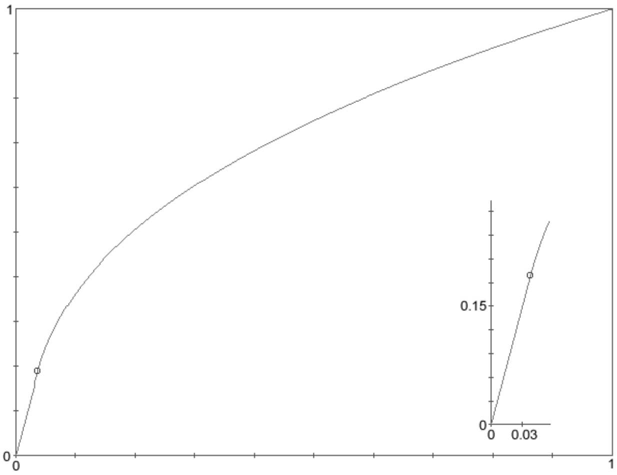
$$\text{for} \quad V_{in} > 0.037703 \quad V_{out} = \left(\frac{V_{in} - 0.02262}{1 - 0.02262}\right)^{0.4} \quad \text{else}$$

$$V_{out} = 5V_{in}$$

and the ITU HDTV law[2]:

$$\text{for} \quad V_{in} > 0.018 \quad V_{out} = \left(1.099V_{in}\right)^{0.45} - 0.099 \quad \text{else}$$

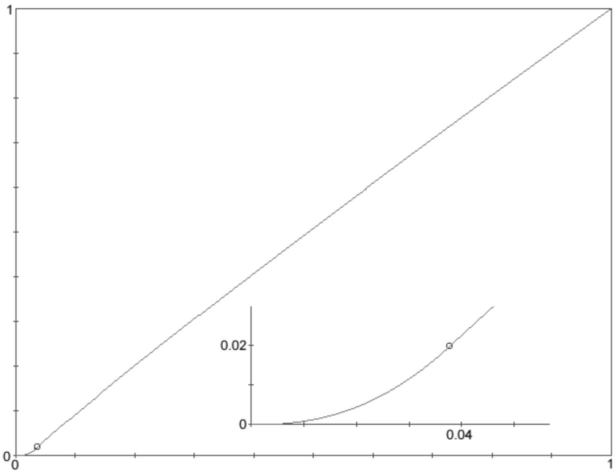
$$V_{out} = 4.5V_{in}$$



Clearly, for perfect colour rendering, a pure power law of $1/\gamma = 1/2.35$ is required, since then the overall transfer characteristic from light to light would be:

$$L_{out} = L_{in}^{\left(1/2.35\right)2.35} = L_{in}^{\ 1} = L_{in}$$

But by using gamma-correction with finite slope near black, the overall transfer curve is flattened somewhat, particularly in the near-black region.



This has the effect of crushing detail near black (e.g. shadows) and reducing the colour saturation of dark colours, while keeping noise at a reasonably low level. Unfortunately it also increases the saturation of bright colours because it acts on the individual *RGB* signals, one or more of which may be very small and thus crushed. At both ends of the contrast scale, colours are shifted in hue towards their nearest primary or secondary colour (Red, Green, Blue, Cyan, Magenta, Yellow) by an amount dependent on saturation. This leads to a poster-type colour rendering of saturated colours. To a very large extent, this is the “video look”.

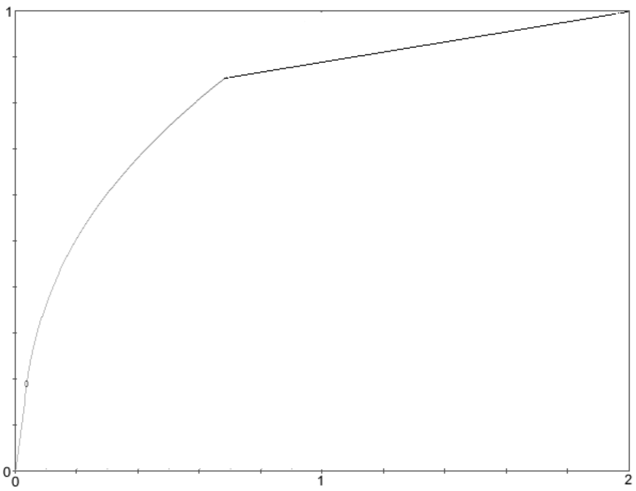
It is a strange and fortunate coincidence that the transfer characteristic of the CRT almost exactly matches the inverse of the characteristic of the human eye. The eye has a logarithmic response rather than the power law of the display, but the correspondence is good over several decades of contrast. Thus, we have a system in which the transmitted signals are well matched to human perception. Any disturbance of the signal is equally visible, at whatever signal level it may occur, and so channel noise is rendered least visible. Had the CRT not had this power law response, we would have had to correct for it anyway, in the interests of coding efficiency, using some form of non-linearity or gamma-correction.

The overall transfer characteristic, that is gamma-correction and knee, are probably the most important parameter in a television camera. More than anything else, this defines the “look” of the resulting pictures.

1.2.1 Knee

When the camera is exposed to a scene with high contrast, and the main subject is under-lit, then it is tempting to adjust the camera exposure such that the main subject looks acceptable but the highlights are over exposed and clipped. This is common where there is an externally lit window, or specular reflections, or where a light source is “in shot”. In many cameras, the input light level that causes 100% output signal is much less than the actual dynamic range of the sensors. It is quite common for only 25% or less of the actual dynamic range to be used. This means that the overloading signal may still be accessible, albeit in the range of signal levels above 100% output, which are normally clipped. There is a manufacturing advantage to doing this (it makes the camera seem to be more sensitive), and a disadvantage (it increases the level of video noise).

The knee function attempts to overcome the clipping problem by attenuating or compressing highlights, the overloading signals. The transfer characteristic follows the prescribed curve up to a “knee break-point”, above which the incremental gain is much reduced. In this way, it may be possible to cope with two or more stops (400% or more) of overexposure without the highlight being clipped. In this highlight region, colour and detail rendering will be unreal, and special precautions may be taken in the camera to protect detail and colouring. A knee curve function may operate before or after gamma-correction, and so may be curved or straight, and may have several curved or straight segments.



Typically, the break-point may be at about 90% of the output, and the slope such that 2 stops of over-exposure is unclipped. In this example, the break point is set to 85% and the slope will accommodate one stop of overload (i.e. 100% output occurs at 200% input). Some specialised cameras have the ability to set the knee much lower, even down to zero, in an attempt to get a “film-look”. This runs the risk of not being able to handle further overloads, since the entire sensor dynamic range is used for the normal range of exposure, and the camera sensitivity or photographic “speed” is reduced.

Ideally, the curve should be continuous and not have any sharp break points since they are known to cause colour contouring in saturated colours. In extreme cases, the use of a knee can simulate the “lazy S” characteristic of film, in which blacks are mildly compressed and several stops of overexposure are adequately captured.

1.3 Coding matrix

To save on transmission and storage bandwidth, the signals are coded and filtered. The filtering will be dealt with in a later section. Traditionally, we process the gamma-corrected *RGB* signals (*R’G’B’*, where the prime indicates gamma-correction) to produce a “luma” signal, by linear addition. The proportions of *R’G’B’* are somewhat arbitrary, but are defined in the PAL/SECAM/625 systems using the proportions relevant for the first American colour system (NTSC) using wide-gamut primaries:

$$Y' = 0.299R' + 0.587G' + 0.114B'$$

In colorimetric terms, the symbol *Y* represents luminance, the human experience of lightness or brightness, but in this instance the symbol is bound together with the prime and called luma since it is not truly luminance but may resemble it. Strictly, the equation for luminance is the centre row of the inverse of the analysis matrix given in Section 1.1:

$$Y = 0.222R + 0.7067G + 0.0713B$$

but it is impracticable to use this formula since the decoding display would have to undo the gamma-correction in order to extract the *RGB* signals and then re-gamma-correct them to drive the display. That would not be a problem in a linear display such as plasma, but is impracticable in CRT displays. Such a system would be said to obey "The Constant Luminance Principle" in which luminance and colour information are truly separated for transmission.

Colour information is coded onto the "chroma" channels as a pair of bandwidth limited colour-difference signals:

$$P_r \propto R'-Y' \quad \text{and} \quad P_b \propto B'-Y'$$

In the compromise system widely used, some perceived luminance information travels via these colouring channels, which have limited bandwidth, and so is lost at higher frequencies. The channels are called "chroma" because they are not true chrominance channels, which would convey only genuine colour information. The issues of bandwidth limitation are discussed in a Section 2.2.

HDTV coding, as defined in ITU R.BT-709, uses a different luma coding equation:

$$Y' = 0.2126R' + 0.7152G' + 0.0722B'$$

and this is based on the coefficients of the analysis matrix for the newer primaries of Rec.709. Note that it is still a luma signal and not luminance, because it is generated from gamma-corrected values. Signals coded with this equation (and the relevant chroma equations) should be decoded with the appropriate decoding equations. Using the Rec.601 decoding equations produces large and unacceptable colouring errors.

2 Resolution properties

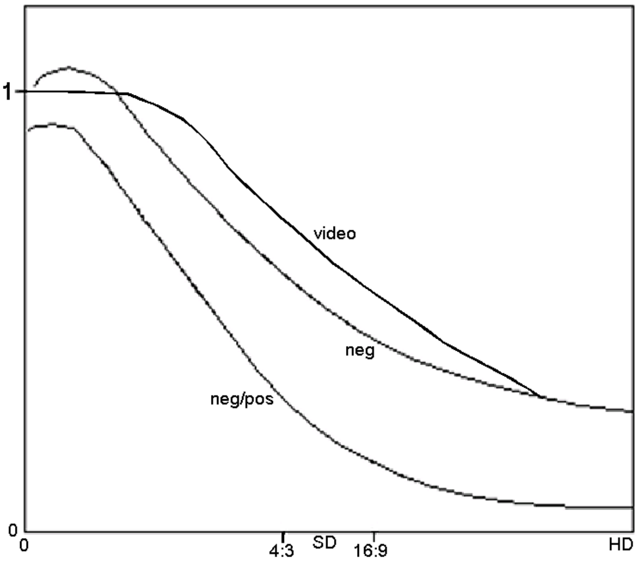
Resolution is defined in two parts, that of the camera and that of the coding system.

2.1 Properties of the camera

2.1.1 Spatial resolution

In the days of electron-beam cameras, the situation was simple, the resolution was limited vertically by the number of scanned lines in the image, and the horizontally by the size of the scanning spot. But with solid-state sensors (charge coupled devices, CCDs) the story is not so simple.

Normal practice is to allocate one row of pixels for each line of the raster (i.e. 576 rows of pixels, one for each of the active lines in the 625-line system). Interlace is achieved by reading the entire contents of the CCD at field rate (50Hz), and adding together adjacent rows of pixels to form the lines of the output signal. Thus the vertical resolution is half that which might be expected. This applies to both fields, but the row-pairing is phase shifted to achieve spatial interlace. The subjective vertical resolution is limited to about 70% of the Nyquist limit (itself about 70%) for the number of lines, at the expense of interline twitter. In pro-scan cameras, this row-averaging does not take place since interlaced fields are not needed, and so the subjective vertical frequency response is limited only by the Nyquist limit for the number of lines.



Horizontally, a different process is used. Although the digital standards for television define the number of horizontal pixels (720 for standard definition, 1280 for 720-line HDTV and 1920 for 1080-line HDTV), the camera CCDs often have many more pixels. This permits the use of good electronic filtering to define the resolution. Also, it is standard practice to offset the Red and Blue CCDs jointly from Green by exactly half a pixel spacing. This has the subjective effect of raising the horizontal resolution by about 30% when luma coding is applied (see below).

For these reasons, it is normal for a camera to have greater (and better controlled) horizontal resolution than vertical, even in a pro-scan system with square pixels.

Film resolution, by comparison, is defined at the time of manufacture. It is worth noting that, for the black and white 35mm film stock illustrated here, the contrast exceeds 100% at low frequencies due to a quirk in the film chemistry. But the resolution rapidly falls in the video bandwidth although it continues at low-level long above the limit of both standard- and high-definition systems. Any optical or printing process inevitably reduces the resolution and so the frequency response of the print combination is much lower than for the negative alone. By comparison, video cameras maintain their frequency response well in the middle-frequency range, even without aperture correction (in this instance, a high-definition camera). It is this difference, the middle-frequency image content, that dominates the perception of image judder in film, and is the reason why video-sourced images are often considered to judder more than does film, whether recorded from a pro-scan camera, or from an interlaced camera given the "film-look" temporal treatment.

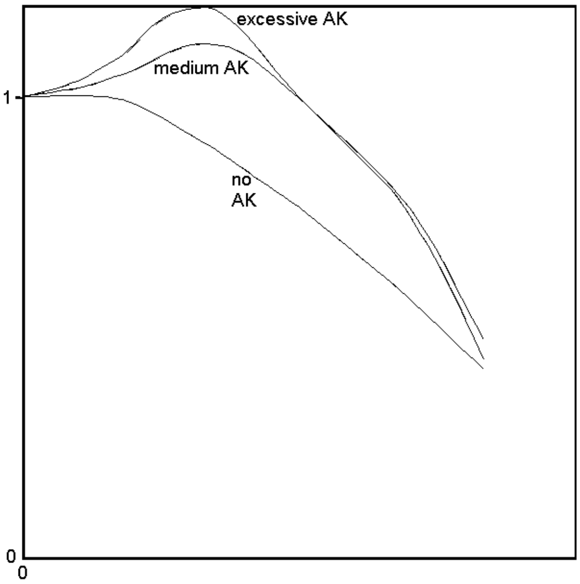
2.1.1.1 Aperture Correction and Detail Enhancement

Aperture Correction was originally a means of compensating for the erratic and irregular frequency response of electron-beam tube cameras. Since the shape of the beam-spot varied with the amount of electron charge it was collecting, so the frequency response varied as well. Also, lens performance may have been less than ideal, so there had to be some method for compensating for these deficiencies. Aperture correction (AK) was incorporated to boost the middle to high frequency content of the signal, both horizontally and vertically. Typically, this involves a transversal filter with access to samples both before and after the target sample, horizontally and vertically. The amount of boost is controlled by a detection circuit that may look at any combination of the *RGB* signals. Only image frequencies that have significant amplitude in the unenhanced image are boosted.

Detail Enhancement is a later addition to the range of image processing operations, and is generally highly adjustable, since it is not incorporated for the specific correction of an error. It is often used simply to add some "crispness" to an otherwise acceptable picture.

In this example of a high-definition camera, it was decided that it would be best to use no correction at all when shooting drama or wildlife, which would otherwise have been shot on film. For sport or light entertainment, the "medium" setting may be suitable, but the "excessive" setting would certainly give too much edge-enhancement for any type of recording.

Both types of enhancement usually operate vertically as well as horizontally, and always produce significant edge enhancement in middle-frequencies, exactly the range of frequencies thought to be important for the visibility of motion judder in film. In some cases, there may be two separate correctors, Aperture and Detail, which may be adjusted independently. This allows more refined control over the frequency response. In exceptional circumstances, a negative Detail Enhancement may suppress middle frequency content and more closely reproduce the performance of film.



Horizontal correction is simple, just using pixel delays, but vertical correction involves line delays. In the normal case of interlaced cameras, the vertical information must be taken from lines of the interlaced field rather than from adjacent frame lines (i.e. lines in the other field of the interlaced pair). Cameras set to produce non-interlaced scans can perform aperture correction using genuine spatially adjacent lines, and so can produce a cleaner picture. More complex schemes are available, but may not yield significant benefits. The amount of correction is controllable, as is the centre frequency and spread of the boost. A threshold circuit may prohibit action until a significant level of detail is detected, to prevent undue amplification of noise.

There may be extra circuits to avoid excessive amplification of detail in skin tones (wrinkles). Typically, a matrix is used to detect skin tones and apply a lower level of correction for just that range of colours. Also there may be special provision to ensure that detail is preserved in highlight parts of the picture where detail is necessarily compressed by a "knee" function. This may involve detecting detail in the signal before compression, and applying correction to the compressed signal. The resulting picture may look "more real", but at the expense of short-term signal excursions that go significantly above peak white level.

Similarly, there may be circuits to protect the hue and saturation of colours that are compressed. Failure to do this results in highlight-compressed parts of the picture unnaturally losing all colour.

Judicious amounts of Detail Enhancement can be beneficial, in that the pictures can be given a degree of sparkle or crispness that they would not otherwise achieve. In this example, the source picture is shown (left), without any correction. The same image is then "medium" corrected (below left) and "over" corrected (below right). The medium correction has sharpened the picture a little, but the over-correction has added graininess and unpleasant edge effects. These are visible as dark and/or light bands around contrasty edges. This effect is a significant part of the "video look".

Excessive application of Detail Enhancement or Aperture Correction can generate edge signals of such high amplitude that they could be clipped, with unpleasant results. Also, this artificial detail can cause problems in compression systems, by effectively overloading the compressor.

2.1.2 Temporal resolution

The temporal resolution of the camera is set by the image capture rate. Each image may be a field containing only half the lines or a frame containing all of them, depending on the system. Thus for 625-line systems, the temporal resolution is defined by the 50Hz field rate, while for 525-line systems is it defined by the 59.94Hz field rate[1]. The emerging HDTV systems are more complicated in that they may be either interlaced or pro-scan, and the frame rate may be selectable by the user. Nevertheless, the temporal resolution is always set by the image rate (whichever is the higher, of field rate and frame rate). Generally, the higher the image rate, the smoother the rendition of motion and the sharper each image is when the scene contains motion, since the exposure time is shorter (being the inverse of the image rate).

A commonly used setting in HD cameras operates at 24 frames/second, in pro-scan mode. This, to some degree, mimics the temporal performance of film shot at 24fps. In this case, the juddering motion portrayal is preferred and actively sought by the programme-maker, to emulate film. The resultant motion judder

---

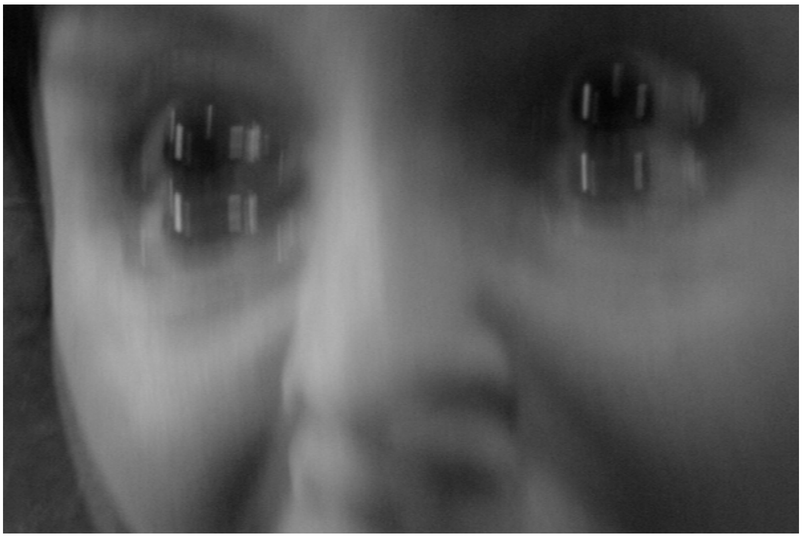
[1] The field rates given are those of the most commonly used systems. There are other, less common, systems used across the world, such as 525/60 (monochrome), 625/59.94 and 525/50.

(which results from the double-showing of each frame) is often thought to be excessive, almost certainly because the video-produced pictures are sharper than those from film. Under these conditions, a negative Detail Enhancement can ameliorate the problem. When a camera is used in pro-scan mode to emulate film, it is necessary to invoke a shutter, to shorten the exposure time, otherwise motion portrayal is too blurred. Film also conventionally uses a 50% shutter. Pro-scan cameras running at significantly higher speeds generally do not need a shutter to shorten exposure time.

2.1.2.1 Shuttering

Film cameras always incorporate a shutter, so that light can be blocked while the film is moved from one frame to the next. Video cameras do not need shutters since the charge pattern accumulated on the CCDs can be transferred into a store in the period (blanking interval) between one image (field or frame) and the next. However, there may be good reasons for wanting to shorten this exposure time. It can be used just to reduce the light input to the camera, to emulate film, to give sharp images for slow-motion usage, or to eliminate beating effects (e.g. when shooting a computer monitor screen, or when using 60Hz lighting with a 50Hz camera, etc.).

A shutter can be applied in two ways, mechanically and electrically. Film cameras use a mechanical shutter, a rotating vane or blind in the light path, usually near the focal plane. Typically, the shutter is open for 50% of the frame duration of 1/24 second, and it is usual to express this in degrees as a 180º shutter. Video cameras more often use an electronic mimic of the shutter, by allowing light-induced charge to accumulate in the CCD and then dumping it all at the appropriate time before the CCD is read, allowing the CCD then to continue accumulating charge for the wanted period. Thus the video camera exposure is always taken during a period that finishes at the end of the image interval, instead of in the middle as in film. This can lead to synchronisation errors (lip-sync) unless it is allowed for.



This photograph is from part of a video image, showing two successive pro-scan frames overlaid. Particularly, it shows the catch-lights in the eyes as sharp, moving lines with gaps where the shutter was closed for 50% of the frame period.

Motion judder results, not from camera shuttering, but from the double display of images. In conventional movie film, this is a necessary evil because the projector would otherwise show at 24 frames per second, and that would cause very upsetting brightness flicker, where the whole picture is seen to consist of separate flashing exposures. The alternative would be to shoot film at 48 frames per second or higher, costing at least twice as much but showing each image only once. By interposing a mechanical shutter in the projector, each frame can be shown twice, and the 48Hz brightness flicker is not then seen as a problem at the relatively low brightness used in cinemas (e.g. around 30-50cd/m^2). The projector shutter will normally be 180º for this purpose. Television display of film achieves the same effect by transmitting image content from each frame twice, at 50Hz[2]. No display shutter is needed since the CRT produces only brief impulses of light from each pixel even though television pictures are watched at much higher brightness (e.g. around 200-400cd/m^2 in the home) and so large area flicker at 24 or 25Hz would be a major problem.

Attempts to mimic film temporal response by applying a shutter in the pro-scan camera have mixed success. It is easy to achieve a 180º shutter simply by dumping half the charge, but the resulting image appears to judder rather more than would a film image. This is because video cameras, by and large, produce more

---

[2] Although film is normally shot for the cinema at 24fps, it is displayed in 625-line television at 25Hz, 4% fast. Thus the double exposure occurs at 50Hz, the field frequency of the system.

resolution in the middle frequency range than does film. Experiments into this effect continue, but judicious use of negative Detail Enhancement has been demonstrated to give a better emulation of film.

2.2 Camera coding resolution

The standard coding system, described in Section 1.3, exploits a well-known limitation of human vision, that we are less sensitive to colour transitions than luminance ones [6]. So the chroma or colouring channels can have narrower bandwidth, and thus occupy less spectrum or bit-rate, than the luma channel. The bandwidths of the luma and chroma signals are defined in the coding system. Normally, the chroma channels have equal bandwidths, but there is some advantage to be gained by having unequal bandwidths, since the colour detection properties of the eye are most spatially sensitive to orange-to-grey colour changes [6]. Thus it could be profitable to phase-rotate these signals such that one signal lay on the orange/blue axis with the other orthogonal to it. Then optimum bandwidths could be assigned to the channels, as was originally done in the NTSC system but was later dropped for economic simplicity in coders and decoders. For PAL-coded composite signals in System I (UK and Eire) the luma limit is 5.5MHz and for chroma 1.3MHz (3dB attenuation) but it is rare for domestic decoders to deal accurately with much more than about 800kHz of chroma. For digital, component, systems the specification deals more with sampling ratios than bandwidths, expressed relative to the luma channel; the following table refers only to these ratios:

System		luma, % of system bandwidth	chroma, % of system b/w	
	ratios		horizontal	vertical
PAL (System I)	N/A	100	25	50
SECAM	N/A	100	25	50
miniDV (consumer, 625-line)	4:2:0	100	50	50
miniDV (consumer, 525-line)	4:1:1	100	25	100
DVC-PRO	4:1:1	100	25	100
Production vtrs (D1, D5 etc)	4:2:2	100	50	100
HDTV (HD D5, D6)	4:2:2	100	50	100
HDTV (HDCAM-1080)	3:1:1	75[1]	25	100
HDTV (720-line)	4:2:2	100	50	100
MPEG-2 (contribution)	4:2:2	100	50	100
MPEG-2 (distribution)	4:2:0	100	50	50

The table gives the sampling ratios, not the actual bandwidths. Filter characteristics are defined in the relevant standards. The factor 4 for luma derives from the original sampling system (pre-1982) where the sampling was precisely locked to a multiple of the colour sub-carrier frequency in composite systems (3 times for 625-lines, 4 times for 525-lines).

The filters used to define the bandwidths depend on the system. Typically, the luma filter is defined to be flat to the channel limit (5.75MHz for 625-line systems) and then falling rapidly into the stop-band (greater than 12dB attenuation at 6.75MHz, exactly half-sampling frequency). Note that this does not eliminate the possibility of aliasing resulting from excessive high frequency image content, above 6.75MHz, but it will always be at low amplitude. The chroma filters fall more smoothly (flat to 2.75MHz, greater than 6dB attenuation at 3.375MHz, exactly half the chroma sampling frequency)[1]. Thus the chroma channels do not necessarily carry the frequency content that they theoretically could.

2.3 Transmission coding resolution

Ideally, the recording or transmission system should be transparent, what goes in comes out. In practice, some compression is likely, and there may be some sub-sampling as well. The compression system may be intra-frame or intra-field (in which only information from one frame or field contributes to the transmitted data blocks), or may be more complex, involving longer sequences of frames and motion prediction. In general, the higher the compression ratio, the more artefacts will become apparent.

3 References

1 International Telecommunications Union (ITU) Recommendation ITU-R BT.470-4 “Conventional, enhanced and high-definition television systems”. Also ITU-R BT.601-5 “Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios”

2 Society of Motion Picture and Television Engineers (SMPTE) standard 274M “1920 x 1080 Scanning and Analogue and Parallel Digital Interfaces for Multiple Picture Rates”. Also SMPTE standard 295M “1920 x 1080 50Hz – Scanning and Interfaces”, and SMPTE standard 296M “1280 x 720 Progressive Image Sample Structure – Analogue and Digital Representation and Analogue Interface”. Also ITU- BT.709-2 “Parameter values for the HDTV standards for production and international programme exchange”.

3 W.N.Sproson, “Colour Science in Television and Display Systems”; Adam Hilger Ltd, Bristol, 1983 (Ch. 2). Also Dr.R.W.G.Hunt, “The Reproduction of Colour”; Voyager Press, sixth edition 2002 (earlier editions from Fountain Press are still valid).

4 International Commission on Illumination (CIE) Publication No.15

5 BBC specification TV2248 “Colour Television Cameras”

6 BBC Research Department Report T-068 (1958/20), K. Hacking, “Colour transitions”

4 List of Appendices

A	Scanning standards for television and computer displays
B	Compression systems for tape recording

APPENDIX A : Video scanning systems

This Appendix tabulates some known display standards. Information is gleaned from published standards documents, and from the manuals of specific equipment. The sources are given in the tables. These lists are not intended to be exhaustive, but are thought to be accurate at the time of writing.

Television standards:

		Aspect Ratio	Pixel size		Scan rate			SDI	
System	Version		H	V	H (kHz)	V (Hz)	Sync	Mb/s	Spec
SDTV 525i	Analogue composite	4:3	704	483-487	15.73	59.94	C-neg		
	Digital	4:3/16:9	720	480	15.73	59.94	C-neg	270/360	ITU601
SDTV 625i	Analogue composite	4:3	702	575	15.63	50.00	C-neg		
	Digital	4:3/16:9	720	576	15.63	50.00	C-neg	270/360	ITU601
SDTVp	483 SMPTE 293	16:9	720	480	31.47	60.00	trisync		
	575	16:9	720	575	31.25	50.00			
HDTV 750 (1280 x 720p)	23.98 SMPTE 296	16:9	1280	720	17.98	23.98	trisync	1484	SMPTE 296
	24 SMPTE 296	16:9	1280	720	18.00	24.00	trisync	1485	SMPTE 296
	25 SMPTE 296	16:9	1280	720	18.75	25.00	trisync	1485	SMPTE 296
	29.97 SMPTE 296	16:9	1280	720	22.48	29.97	trisync	1484	SMPTE 296
	30 SMPTE 296	16:9	1280	720	22.50	30.00	trisync	1485	SMPTE 296
	50 SMPTE 296	16:9	1280	720	37.50	50.00	trisync	1485	SMPTE 296
	59.94 SMPTE 296	16:9	1280	720	44.96	59.94	trisync	1484	SMPTE 296
	60 SMPTE 296	16:9	1280	720	45.00	60.00	trisync	1485	SMPTE 296
HDTV 1125 (1035i)	59.94 SMPTE 240	16:9	1920	1035	33.75	59.94	trisync	1484	SMPTE 292
	60 SMPTE 240	16:9	1920	1035	33.82	60.00	trisync	1485	SMPTE 292
HDTV 1125 (1080i)	48	16:9	1920	1080	27.00	48.00	trisync		
	50 SMPTE 274	16:9	1920	1080	28.13	50.00	trisync	1485	SMPTE 292
	59.94 SMPTE 274	16:9	1920	1080	33.72	59.94	trisync	1484	SMPTE 292
	60 SMPTE 274	16:9	1920	1080	33.75	60.00	trisync	1485	SMPTE 292
HDTV 1125 (1080p)	23.98 SMPTE 274	16:9	1920	1080	26.97	23.98	trisync	1484	SMPTE 292
	24 SMPTE 274	16:9	1920	1080	25.00	24.00	trisync	1485	SMPTE 292
	25 SMPTE 274	16:9	1920	1080	28.13	25.00	trisync	1485	SMPTE 292
	29.97 SMPTE 274	16:9	1920	1080	33.72	29.97	trisync	1484	SMPTE 292
	30 SMPTE 274	16:9	1920	1080	33.75	30.00	trisync	1485	SMPTE 292
	50 SMPTE 274	16:9	1920	1080	56.25	50.00	trisync		
	59.94 SMPTE 274	16:9	1920	1080	67.43	59.94	trisync		
	60 SMPTE 274	16:9	1920	1080	67.50	60.00	trisync		
HDTV 1250	50i SMPTE 295	16:9	1920	1080	31.25	50.00	trisync	1485	SMPTE 292
	50p SMPTE 295	16:9	1920	1080	62.50	50.00	trisync		
	50i Eureka 95	16:9	1920	1152	31.25	50.00	trisync		

Notes: 625/50p (575) and 1125/48i (1080) references occur in some display devices such as projectors, and are not acknowledged standards for broadcasting or production.

Frame and field rates that are not integers are calculated from the next highest integer divided by 1.001, thus 59.94 is an approximation to 60/1.001. Time codes for such systems are all of the “Drop Frame” variety, which is needed to ensure that the hours/minutes/seconds on tape match real clock time.

Computer display standards

System	Version	Aspect Ratio	Pixel size H	Pixel size V	Scan rate H (kHz)	Scan rate V (Hz)	Sync
VGA (640 pixels)	VGA-1 VGA350	4:3	640	350	31.47	70.09	H-pos, V-neg
	VESA 85 (VGA350)	4:3	640	350	37.86	85.08	H-pos, V-neg
	NEC PC98	4:3	640	400	24.82	54.42	H-neg, V-neg
	VGA-2 (TEXT)/VESA 70	4:3	640	400	31.47	70.09	H-neg, V-pos
	VESA 85 (VGA400)	4:3	640	400	37.66	85.08	H-neg, V-pos
	VGA VESA 60	4:3	640	480	31.47	59.94	H-neg, V-neg
	Mac 13”	4:3	640	480	35.00	66.67	S on G
	VGA VESA 72	4:3	640	480	37.86	72.81	H-neg, V-neg
	VGA VESA 75 (IBM M3)	4:3	640	480	37.50	75.00	H-neg, V-neg
	VGA VESA 85 (IBM M4)	4:3	640	480	43.27	85.01	H-neg, V-neg
SVGA (800 x 600)	SVGA VESA 56	4:3	800	600	35.16	56.25	H-pos, V-pos
	SVGA VESA 60	4:3	800	600	37.88	60.32	H-pos, V-pos
	SVGA VESA 72	4:3	800	600	48.08	72.19	H-pos, V-pos
	SVGA VESA 75 (IBM M5)	4:3	800	600	46.88	75.00	H-pos, V-pos
	SVGA VESA 85	4:3	800	600	53.67	85.06	H-pos, V-pos
832 x 624	Mac 16”	4:3	832	624	49.72	74.55	H-neg, V-neg
XGA (1024x 768)	XGA VESA 43 (8514)	4:3	1024	768	35.52	86.96	H-pos, V-pos
	XGA VESA 60	4:3	1024	768	48.36	60.00	H-neg, V-neg
	XGA VESA 70	4:3	1024	768	56.48	69.96	H-neg, V-neg
	XGA VESA 75	4:3	1024	768	60.02	75.03	H-pos, V-pos
	XGA VESA 85	4:3	1024	768	68.68	85.00	H-pos, V-pos
	Mac19”	4:3	1024	768	60.24	74.93	H-neg, V-neg ?
1152 x 864	VESA 70	4:3	1152	864	64.00	70.02	H-pos, V-pos
	VESA 75	4:3	1152	864	67.50	75.00	H-pos, V-pos
	VESA 85	4:3	1152	864	77.49	85.06	H-pos, V-pos
1152 x 870	Mac21”	4:3	1152	870	68.68	75.06	H-neg,V-neg ?
1152 x 900	SUN LO	?	1152	900	61.80	65.96	H-neg, V-neg
	SUN HI	?	1152	900	71.71	76.05	C-neg

Wide screen versions, achieved by adding pixels horizontally or cropping vertically

Aspect Ratio	Pixel size Extra H	Pixel size Crop V
16:9	853	360
16:9	853	360
16:9	853	360
16:9	853	360
16:9	853	360
16:9	1067	450
16:9	1067	450
16:9	1067	450
16:9	1067	450
16:9	1067	450
16:9	1109	468
16:9	1365	576
16:9	1365	576
16:9	1365	576
16:9	1365	576
16:9	1365	576
16:9	1365	576
16:9	1536	648
16:9	1536	648
16:9	1536	648
16:9	1536	652

1280 x 960	VESA 60	4:3	1280	960	60.00	60.00	H-pos, V-pos	16:9	1707	720
	VESA 75	4:3	1280	960	75.00	75.00	H-pos, V-pos	16:9	1707	720
SXGA (1280 x 1024)	SXGA VESA 43	5:4	1280	1024	46.43	86.87	H-pos, V-pos	16:9	1820	720
	SGI-5	5:4	1280	1024	53.32	60.06	H-neg, V-neg	16:9	1820	720
	SXGA VESA 60	5:4	1280	1024	63.97	60.01	S on G	16:9	1820	720
	SXGA VESA 67	5:4	1280	1024	70.78	67.00	?	16:9	1820	720
	SXGA VESA 75	5:4	1280	1024	79.98	75.03	H-pos, V-pos	16:9	1820	720
	SXGA VESA 76	5:4	1280	1024	81.20	76.18	H-pos, V-pos	16:9	1820	720
	SXGA VESA 85	5:4	1280	1024	91.15	85.02	H-pos, V-pos	16:9	1820	720
UXGA (1600 x 1200)	UXGA VESA 60	4:3	1600	1200	75.00	60.00	H-pos, V-pos	16:9	2133	900
	UXGA VESA 65	4:3	1600	1200	81.91	65.53	H-pos, V-pos	16:9	2133	900
	UXGA VESA 70	4:3	1600	1200	87.50	70.00	H-pos, V-pos	16:9	2133	900
	UXGA VESA 75	4:3	1600	1200	93.75	75.00	H-pos, V-pos	16:9	2133	900
	UXGA VESA 85	4:3	1600	1200	106.25	85.00	H-pos, V-pos	16:9	2133	900
QXGA (2048 x 1536)	QXGA VESA 60	4:3	2048	1536	95.34	59.94	?	16:9	2731	1152
	QXGA 60max	4:3	2048	1536	95.76	60.00	?	16:9	2731	1152

Notes: VESA is the registered trade mark of Video Electronics Standard Association.
Aspect ratio of non-480-line VGA is assumed to be 4:3 with non-square pixels

APPENDIX B : Tape compression systems

This Appendix lists some better know video digital compression systems. The compression ratios are only approximate, and are usually only quoted as being “about n:1”. Calculations based on the image format and these compression ratios have revealed that the approximation is often poor, since it frequently results in data rates well above the payload capability of the tape format (see below). The bit rates and compression ratios are, therefore, only indicative and not precise. The reader should also bear in mind that this table takes no account of audio signals. Even miniDV, the variety commonly found in consumer camcorders, requires a substantial bit-rate to support a stereo pair sampled at 48kHz in 16-bit uncompressed form, i.e. 192kB/s or 1.536Mb/s. Thus the available bit-rate for compressed video and meta-data is only 23.464Mb/s, so the compression ratio must be a little above 5:1 for the entire signal to fit into 25Mb/s. At the other end of the scale, all the professional compressors include at least 4 audio channels at 48kHz with at least 16-bit coding, requiring over 3Mb/s just for sound.

	Camera				Filter and sub-sample preprocessor					Compressor		SDTI
	Pixels		Frames		Pixels		Coding		Raw	Compressed		
	W	H	f/s	I/P	ratio	W	Y'CC	bps	Mb/s	N:1	Mb/s	
DV	720	480	59.94	2		720	4:1:1	8	124.292	5	24.858	25
	720	576	50	2		720	4:2:0	8	124.416	5	24.883	
DVC Pro	720	480	59.94	2		720	4:1:1	8	124.292	5	24.858	25
	720	576	50	2		720	4:1:1	8	124.416	5	24.883	
	720	480	59.94	2		720	4:2:2	8	165.722	3.3	50.219	50
	720	576	50	2		720	4:2:2	8	165.888	3.3	50.269	
Digi Beta	720	480	59.94	2		720	4:2:2	8	165.722	2	82.861	
	720	576	50	2		720	4:2:2	8	165.888	2	82.944	
DVC Pro HD	1280	720	59.94	1	3/4	960	4:2:2	8	662.889	6.7	98.939	?
	1280	720	60	1	3/4	960	4:2:2	8	663.552	6.7	99.038	
	1920	1080	50	2	3/4	1440	4:2:2	8	622.080	6.7	92.848	
	1920	1080	59.94	2	2/3	1280	4:2:2	8	662.889	6.7	98.939	
	1920	1080	60	2	2/3	1280	4:2:2	8	663.552	6.7	99.038	
HDCAM	1920	1080	50	2	3/4	1440	3:1:1	8	518.400	4.3	120.558	270
	1920	1080	59.94	2	3/4	1440	3:1:1	8	621.459	4.3	144.525	
	1920	1080	60	2	3/4	1440	3:1:1	8	622.080	4.3	144.670	
	1920	1080	23.98	1	3/4	1440	3:1:1	8	497.167	4.3	115.620	
	1920	1080	24	1	3/4	1440	3:1:1	8	497.664	4.3	115.736	
	1920	1080	25	1	3/4	1440	3:1:1	8	518.400	4.3	120.558	
	1920	1080	29.97	1	3/4	1440	3:1:1	8	621.459	4.3	144.525	
	1920	1080	30	1	3/4	1440	3:1:1	8	622.080	4.3	144.670	
HD-D5	1280	720	59.94	1		1280	4:2:2	10	1104.815	3.5	315.661	360
	1280	720	60	1		1280	4:2:2	10	1105.920	3.5	315.977	
	1920	1080	50	2		1920	4:2:2	10	1036.800	3.5	296.229	
	1920	1080	59.94	2		1920	4:2:2	10	1242.917	3.5	355.119	
	1920	1080	60	2		1920	4:2:2	10	1244.160	3.5	355.474	
	1920	1080	23.98	1		1920	4:2:2	10	994.334	3.5	284.095	
	1920	1080	24	1		1920	4:2:2	10	995.328	3.5	284.379	
	1920	1080	25	1		1920	4:2:2	10	1036.800	3.5	296.229	
	1920	1080	29.97	1		1920	4:2:2	10	1242.917	3.5	355.119	
	1920	1080	30	1		1920	4:2:2	10	1244.160	3.5	355.474	

Notes

Digibeta compression is "about half" according to manuals. 2:1 produces data rates nicely inside 88Mb/s, tape payload rate. Processing is 10-bit but only 8 bits are believed to go to the compressor.

HDCAM spec. is 8-bit 4:2:2, 3:1:1 filtered and subsampled then about 4.3:1 compressed, claimed to be equivalent to 1920 4:2:2 compressed 7:1. Bit-rate sums confirm this.

HD-D5 spec. is 10-bit 4:2:2, compressed "about 3:1", that puts bit-rates above SDTI, 3.5:1 is probably nearer the truth.