A Comparison of Various Video Compression Methods for Use in Instrumentation Systems

James L. Rieger, PE/PTBW, and Sherri L. Gattis
Naval Air Warfare Center, Weapons Division, China Lake, CA

May 29, 1993

Abstract

Various forms of 'lossy' and 'lossless' encoding have been implemented or proposed to allow compression of still and motion picture images, with varying degrees of success which in turn depend on how *success* is defined. Proponents of various systems claim 'compression ratios' which by their nature defy comparison of one system with another and suggest there actually may be something like a free lunch. This paper compares various compression methods as well as the implications involved in using them and what happens when different systems encounter the problems associated with the uses of the restored picture.

Key Words: video, compression, instrumentation, JPEG/MPEG

1 'Real' Pictures

A picture is a picture *of something* only when objects in the picture can be recognized by the 'ultimate receiver', which may be a human or a machine, or both. As a consequence, a picture consisting only of a series of dots is meaningless unless at least some of the dots were formed into larger blobs of shape, color, and texture. If motion is involved, at least two pictures are needed with the object(s) in motion in different positions, and enough detail to show the change is required. That a picture, or series of pictures contains some order rather than just random data implies that we can reduce the amount of data we transmit or store and recover all, or at least the important parts, of the picture, without treating the data as if it were completely random .

1.1 Dimensions

A street address has two d*imensions* to it, one in the North-South direction and one in the East-West direction. That these dirctions are *orthogonal* is clear when we consider that if we're a block off on one part of the address, we cannot correct for it in any way by changing the value of the other dimension. A still or motion picture has attributes which are orthogonal dimensions:

Vertical resolution Vertical resolution is often taken to be the number of horizontal lines in a single picture. The lines need not be scanned in any particular order, but they are often scanned from top to bottom and may be interlaced. In the NTSC '525 line' television system, there are 480 to 483 of them, scanned alternately. Vertical resolution can never be greater than the number of lines in a picture, but can be—and often is—far lower.

Horizontal resolution In an analog television system, changes along the horizontal scanning line are limited by bandwidth and slew rate of the electronics, and (like the vertical resolution) by the size and shape of the scanning mechanism. Horizontal resolution can be expressed in a number of ways, with higher numbers used by the manufacturers, giving the user a false sense of security. Digitized scanning systems can have adjacent picture elements—'pixels'—completely independent, something an analog system can't do, and horizontal resolution is then due entirely to the number of pixels in a line and the shape (and 'fill factor') of the dots. In all systems, resolution is limited by the optical system which focuses the photographed scene onto the scanning mechanism.

Grayscale resolution A scene is limited by the the ratio between lightest and darkest element in it, which may be beyond the range of a human observer. Between the darkest and lightest points in a picture there may be a continuous range of gray, or many shades which appear continuous, or no intermediate shades at all, as in the printing process. The brightness range can be considered to be a linear function, but if considered as a logarithmic *density* function instead can be rendered with fewer discrete shades and still appear to be continuous.

Temporal resolution A picture which changes as a function of time has some form of temporal resolution, although the entire picture may be taken at a single instant and then displayed in a single instant or as a swept image. Interlace of the vertical (and possibly horizontal) elements in presentation, repetition of previous information, and phosphor persistence (a function of brightness and/or color) complicate the definition still further.

Color resolution In color, the reproduced scene is often altered to compensate for the imperfect illumination of the original scene, a subjective adjustment made so that the rendered colors 'look' right, even though they are not. The difference between a color and black-and-white involves two more components, the nature of which can be *hue* (the difference between red, green, blue, and orange) and *saturation* (the difference between white, pink, and red) or some other combination; the two color components plus the monochrome representation solve three equations in three unknowns. A color system using three independent color channels to represent the sprectral continuum is know as the *tristimulus model* of color. In instrumentation, there can be more colors than three or fewer, and 'colors' may be outside the range of human vision. A human eye cannot resolve a red-blue checkerboard at the same distance that a black-white board can be resolved, nor easily discern between two shades when saturation or illumination is low, hence entertainment television diminishes color signal resolutions accordingly, a simplification which may be inappropriate for instrumentation television.

For a picture file to contain the maximum amount of information, each of the above dimensions would be totally random. No 'real' picture is that chaotic, even a television receiver tuned between channels, yet only a file consisting of random information in each dimension makes full use of its communication channel.

2 The 'Ultimate Receiver'

The ultimate receiver of entertainment television is the human eye, and the intent of the system is to provide the most pleasing representation of whatever the program's producer wants the viewer to see, whether the resulting scene accurately represents what actually occurred or not. The viewer cannot examine the picture frame by frame, blow up parts of the picture, or vary the brightness and color values of parts of the scenes, although to do so would often be quite revealing. The ultimate receiver of an instrumentation television picture may or may not be a human eye, since electronic devices exist which can measure times, distances, speeds, count objects, etc., from a television signal. Even when a human is present in the loop, it is often possible to examine the resulting scene one picture at a time, and examine parts of the picture by changing the grayscale or color values and ranges. The goal of an instrumentation television system is to allow measurement of something, and the system must be optimized to allow whatever measurements to be made. Even if the picture were read and displayed 'perfectly' (whatever that means), the electronics between the pickup and display can only degrade system performance, never improve it.

3 The Nature of Compression

While *compression* has meaning in the analog domain (systems such as NTSC color television are analog compressors), the interest in instrumentation is in compression of files that result from an original digital picture or a conversion (assumed perfect) of an analog signal. Compression of such a file depends on removal of *redundancies* from the signal. The nature of redundancies are semi-intuitive—adjacent film frames look much like one another except where the scenes cut; a single dot in a photo is surrounded by many more dots of nearly the same color; the average brightness of a television picture—black-and-white or color—is close to gray. If we use shorter words to describe the common occurrences than those less common, the overall result will be a smaller file. Removal of picture attributes that don't matter to the user is another possibility, but the losses created by dropping information cannot be recovered, and degrade the reconstructed picture in a subtle or not-so-subtle way.

3.0.1 Compression Ratio

When the size of a compressed file is compared to the size of the uncompressed original, the result can be called a *compression ratio*, although the term is misleading if the compression isn't lossless. A way to inflate any system's apparent compression ratio is to start with a file whose resolution is far higher than the data it represents. Consider a standard television picture (or a VGA screen) consisting of 640 pixels in the horizontal direction, 480 pixels in the vertical direction, and three colors each of which has 8-bit resolution. Since the representation has just over 300,000 pixels, with each pixel expressed to 24-bit accuracy, we thus need nearly one megabyte of picture information to describe the picture completely. However, while we have 2^{24}, or approximately 16 million possible colors, we can have only 300,000 possible colors, and only if each pixel were a different color from all others. We could instead tabulate all the colors in the picture, and assign a unique code to each one, reducing the number of bits per pixel to no more than nineteen—with no degradation to the picture whatsoever. We have achieved a 25% <u>lossless</u> compression. There are far less than 2^{18} colors in an 'average' picture, so the 'compression' can be far more dramatic, but the actual amount will vary with picture content.

With any lossless compression system, the amount of compression achieved varies with picture content, and with the match between the real data and the data expected. Any lossless compression method—and most lossy types—decrease file size on the average, but do not produce the same size File for every picture.

3.l File Size

The reason that we wish to compress a picture file in the first place is that the uncompressed file resulting from even a single picture is huge; the statement that a picture is worth a thousand words is an understatement . A digital representation of an uncompressed single frame of television is the equivalent of several hundred pages of typewritten text. A single television picture displaces around 600 voice circuits in the telephone network under the best of circumstances. A telemetry channel, be it through a radio link or in a recording medium, is more limited in its capability than needed to handle uncompressed pictures. Since the file size produced varies with picture content and the transmission rate is fixed, a method must be used to limit the file size each picture produces to some maximum, trading resolution for consistent file size. Since the resulting compressed file is of unknown size until after compression, the system must adjust resolution to keep the transmission rate constant.

3.2 Overflow/Underflow

The size of the encode buffer should be such that it cannot overflow when fed maximum surprise at its input while the compression process adjusts resolution to servo the buffer back to normal condition. Similarly, when the picture becomes less complex and the buffer tends toward empty, the resolution should increase until the buffer status returns to normal. If the buffer is emptying even at the highest compressor resolution, the channel must send padding to prevent the buffer being emptied entirely, which will result in a malfunction.

3.3 Delay or Latency

If the servos that keep the buffer in normal condition operate properly, the user need not to know that a buffer is operating at all, except that the delivered picture is delayed. However, if the encoder stops transmitting (the missile hits the target and explodes, for example), all data in the transmit buffer is lost. Teleconferencing systems have delays leading to the conclusion that the person on the other end is a bit slow-witted; delays in systems where the picture is being used for remote navigation are more serious. The amount and location of the delay that is tolerable depends on the use.

3.4 Error Recovery

Most systems assume that the file is received intact. Computers get upset and stop or lock up when they encounter a single bit error when reading a file. A video decoder must read through errors, display something or indicate that it can't, and recover thereafter. An instrumentation system starts receiving a signal after transmission has started, so the decoder must operate with what it gets from the time it starts receiving and onward. Errors due to gaussian noise are randomly spaced, but even if the error rate is one per hundred bits, the chances of two bits in a row being received incorrectly is slight; error-correction systems for this type of noise are fairly easily built. Dealing with errors that occur in bursts, however, generally involves transmitting data in a different order than it was created, and reshuffling back into order at the decoder, increasing delay and buffer size, so error correction is seldom used.

3.5 Limiting Conditions

While the 'average' compression, file size, picture degradation, etc., are of some interest, the item of greatest concern might be the performance of the system under the conditions where compression is at its maximum, or when the system is encountering the picture least suited for the system. The portion of the file that holds the most interesting stuff is likely not an 'average' picture.

4 Decimation

The easiest way to 'compress' anything is to leave something out in some arbitrary fashion. Transmitting every other horizontal pixel and every other line reduces the raw file size by a factor of four; the missing lines and dots can either be regenerated by making the remaining dots bigger or by *interpolating* the missing ones from the remaining ones nearby. It's also possible to reduce a picture rate by transmitting fewer pictures per second and repeating them at a faster rate—movie theaters get 72 Hz flicker from film with 24 frames per second. It's also possible to reduce the grayscale resolution significantly without much difference in average picture quality, as shown in Figure 1. With color pictures, reduction of color resolution in space, time, and intensity can be quite extreme without being obvious.

5 Delta Coding

Delta coding sends the difference between the pixel under consideration and the level predicted by a mixture of some combination of pixels already known, under the assumption that any pixel's brightness is somehow related to those around it.[1] Since the predicted value is often close to the actual value much of the time, the difference can be encoded in such a way that no difference with the predicted value is transmitted with a short code and progressively larger differences are transmitted with longer codes—a so-called *entropy coding* technique. Another method is to transmit only a single bit per pixel, its polarity being determined by whether the prediction was lower or higher than the actual value. The transmitted signal will then toggle between a **one** and a **zero** if the predictor is actually correct, so the step size appears on the display as a granularity. Since abrupt changes in brightness cause longer codes if

---

[1]The difference between any predicted and actual value is of use only if the starting value is known, so the starting value must either be transmitted occasionally or agreed on prior to transmission.

Figure 1: Dithered B&W Image @ One Bit/Pixel

variable-length codes are used, the resulting file size will vary with predictor efficiency; one-bit codes produce the same file size for any picture but cause blur and delay for rapid transitions. *Adaptive* one-bit coding, which varies step size dynamically, can decrease granularity in areas where brightness doesn't change and sharpens the effects at transitions, but reaction time may cause problems. Delta coding can be used on an analog signal without digitizing it first, which simplifies implementation. Best results are obtained when the sampling rate is an integer or an integer-and-a-half multiple of the horizontal sweep rate.

5.1 Vertical and Horizontal Delta Coding

A predictor based on the previous pixel only is the simplest possible *kernel*, and almost universally the kernel used for one-bit coding. Delta codes which use variable-length coding to transmit several possible differences with the predictor may use two previous pixels to change the ordering of the lengths of entropy codes to increase coding efficiency. The HORACE system described in RCC/TCG-210 uses two pixels in the horizontal direction with variable-length multilevel coding; as a consequence of multilevel coding the file size varies with picture content. To prevent overflow, the coding resolution is adjusted on a line-by-line basis, which the decoder follows without operator intervention.

The previous line in a picture can be used as part of the predictor, and even future[1] lines can be used to predict a single point, but the effect is similar to predicting a new line based on the lines around it. Complexity and delay increase with the selection of a more elaborate kernel, and performance may or may not improve as a result. A screen full of lettering, for example, is more accurately predicted with the previous one or two pixels in the horizontal direction than with a more-elaborate kernel.

5.2 Temporal Delta Coding

A delta coding that transmits the difference between an entire picture and previous (and possible future) pictures can be built; the result is in the category of *inter*frame coding as opposed to *intra*frame coding. Interframe coding works best when the scene doesn't change at all, and file size increases rapidly if something actually moves. At least one previous picture must be stored in its entirety, requiring a large memory on both ends of the link, and because the item of interest in most instrumentation television is the item that moved, interframe coding is seldom used for instrumentation systems. However, teleconferencing systems, which usually involve 'talking heads' and limited movement which can be allowed to blur while the motion continues, might be quite acceptable.

5.3 Color Delta Coding

Delta coding deals with only one dimension—brightness—over whatever other dimensions coding runs. A color image has several (often three) dimensions to it, requiring that the channel carry as many streams of data as the number of color signals, although not necessarily at the same resolution in the other dimensions. In the standard HORACE color, a color separation is sent after each line, and the user can select the grayscale and horizontal resolution to suit the intended use. The color separation sent can be alternated from one line to the next or several sent each time, and there's no restriction on the size of the separation component data, allowing colors which don't resemble the visible brightness image much (such as a radar, infrared, X-ray, etc.), an anaglyphic

[1]The concept of what the future is is negotiable in digital systems, because delays can be introduced so that the future is available as 'now' is being worked on in the past. The delay introduced depends on how far into the 'future' we need to look.

3-D pair, or two or three entirely different pictures to be sent, as long as they are in synchronization with one another.

6 Bit Plane Coding

A technique called *run-length* coding is useful when the value of brightness does not change rapidly in the coding direction, since the code is transmitted as a number which represents how many of what level is often shorter than repeating the slowly- or non-changing level over and over. However, in an analog picture the run lengths encountered often have small changes from pixel to pixel, which makes the system inefficient. On the other hand, run-length coding of the individual bits often is efficient, especially in the most-significant bits. Using a zig-zag coding direction rather than a simple horizontal or vertical motion often increases the sizes of the groups encoded. An encoding system based on run-length coding of the individual bit planes can be truncated to a maximum file length resulting in grayscale resolution being the variable dimension, and signals coded in this way can be displayed while being decoded, allowing the user to step between frames quickly until the picture of interest is located. A few of the bit planes of Figure 1 are shown in Figure 2.

7 Transform Coding

A television signal is periodic, or nearly so, in the horizontal, vertical, and temporal dimensions, and when transmitted in serial fashion produces a signal with energy clustered about the horizontal and vertical sweep rates. Hence it's possible to transform the picture into a set of orthonormal functions and then transmit the magnitudes (and possibly phases) of those functions instead of the picture. Systems have been built using Fourier, Walsh-Hadamard, Haar, etc., sets. The coefficients change slowly for the low frequencies and more rapidly for the higher frequencies, but the accuracy at which higher frequencies need to be transmitted is lower than that needed for the lower frequencies. The coefficients can be transmitted differentially. Complexity of any transformation is greater than with other processes, and involves memory at both the encoding and decoding ends of the link. Quality ranges from essentially perfect down to terrible; objects in motion and diagonal features generally fare the worst, and tend to flicker as they move.

Figure 2: Top Bitplane for Figure 1

8 Vector Coding

Since the vertical and horizontal directions in a picture are dimensions, a combination of pixels taken in a group is a *vector* in the mathematical sense, since a change in the value in any element cannot be corrected by any change in other elements. Visually, these vectors appear as mosaic tiles, and are often referred as *tiles* rather than *vectors* for that reason. A typical vector block used might be a 4x4, 8x8, or 16x16 group of pixels, with each pixel represented by a single luminance value or a vector of color values. Just as it is highly unlikely that each pixel in a large picture is of a different color from all others, so is it unlikely that all the tiles comprising a large picture are different from each other—or, even if not identical, a far smaller selection of tiles might be produced which can be used to rebuild the picture to an acceptable approximation. The tiles themselves are more likely to have pixels all of the same color, or nearly so, than to resemble color confetti, just as while all letters of the alphabet occur in normal text, blocks of three containing **the**, **ing**, or **ses** are more likely than **uum**, and **qqx** never occurs at all. A variable-length code can efficiently send vectors, whose statistics are more widely distributed than the values of each pixel.

The JPEG—for Joint Photographic Expert Group—coding system uses a combination of the techniques discussed above to provide an efficient coding of most pictures. An input picture is first divided into 8x8 tiles, with one tile representing the luminance value and (if the picture is in color) the next tile representing one of two color separation components corresponding more or less to the *U* and *V* components used for European color TV.[2] Each tile is then transform coded with the discrete cosine transform into 64 components, the lowest representing the DC value (average brightness) of the tile and the rest indicating spatial frequencies in the vertical, horizontal, and along diagonals. Not surprisingly, many of the diagonal terms are so near zero in amplitude that they can be safely ignored, and the picture reconstructed 'losslessly'.[3] For further compression, the amplitudes of the significant spatial components are transmitted with variable length codes and the amplitude changes are bracketed—with the bracketing level determining how crude the reconstruction will be—and the zeros are run-length coded. The coefficients are read out in zigzag fashion beginning with the DC term, and for further compression all 64 of them need not be transmitted. Differing amounts of loss can be assigned to the color separations, which can be bracketed more broadly and fewer coefficients sent. For any given compression thresholds, the File size will be variable; scenes with high contrast areas and sharp edges make larger files. The effects of 'normal' and 'extreme' JPEG coding can be shown for any given picture, but there are surprises when random pictures are used.

The decoder is told what compression thresholds are used at the start of each picture, and resolution cannot be adjusted during a picture. If a file is too large, the entire picture is not transmitted, with the last rows of tiles at greatest risk. Because coding is adjusted for each picture rather than for each line or row of tiles, the encoder buffer must be longer and the delay consequently greater than with the simpler line by line or bit by bit coding adjustments possible with delta coding alone. The benefit of the increased complexity is that a JPEG file is

---

[2]For each tile reconstructed, the luminance value tile uses its corresponding color tile and for the other color separation uses an average of the tiles on either side.

[3]Not counting roundoff and truncation errors and what we did to the color separations.

typically far smaller in size than the files created by other schemes, which are limited to no less than one bit per pixel. In color, the user is limited to the color separation ratios allowed by JPEG and to three colors assumed to all contain essentially the same scene—both limitations which are fully acceptable for some uses and wildly inappropriate for others.

9 Feature Coding

Feature coding is used to keep track of a countable number of objects in the scene, and stores them with regard to position, size, color, or whatever. The theory is that, for example, a scene containing a bouncing red ball is most easily described by keeping track of the size and position of the ball as it moves, and displaying that information at the receiver with little concern about what else is in the picture. While the complexity of such a system increases as the number of objects and with what attributes and precisions we wish to measure, the amount of data that must be exchanged is minimal. Systems which send teleconferencing data by selecting an appropriate face and keeping track of about twelve points on the face for transmission have been demonstarted, and a similar system is used to make animated cartoons. The parts of the picture that aren't coded may be distorted or missing altogether, but if what is to be measured is miss distance, feature coding might be appropriate.

10 Fractal Coding

Perhaps the strangest coding method is that called *fractal* coding, where shapes, colors, and textures are drawn by a system not unlike 'paint by numbers'. The descriptor list can be quite small, involving things as simple as 'color bars' or 'a field of flowers'. Computer screen scenes often lend themselves well to encoding of this type, in part because most scenes are created from primitive elements and in part because what the screen depicts did not have its origin in nature. In some sense, the picture thus generated is identical to the one described, although the resemblance is tenuous if the description is vague or terse. On the other hand, if a tree-bark pattern is placed where a tree trunk is blocked out, that the tree bark thus reconstructed looks as good as the original but doesn't represent it point by point may not matter. A *fractal* coding system can be efficient, but reducing something down to its fractal descriptors is the most time-consuming and computer-intensive technique of all.

11 Conclusion

While we've said often that the type of coding used for pictures, and how to set the knobs on a given system for best results depends on what is wanted, the statement still seems like a copout. Yet without knowing what attributes are essential, which are 'nice to have', and which are superfluous, there is no simple answer. The method used should be determined on the basis of the needs of the user, channel capacity, noise performance, cost, and availability. Even with those limitations, however, there's something available which can do almost anything reasonable for any user.