Title
Perceptual Criteria and Design Alternatives for Low Bit Rate Video Coding

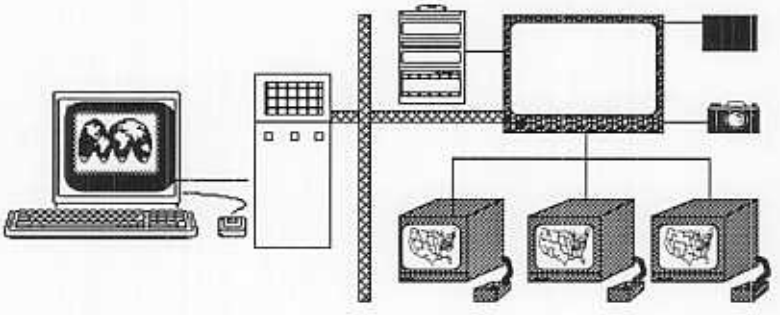
Authors
Algazi, Ralph
Hiwasa, N.

Publication Date
1993

Peer reviewed

Center for Image Processing and Integrated Computing

a CIPIC Report



Perceptual Criteria and Design Alternatives for Low Bit Rate Video Coding

by

V. Ralph Algazi and Norimichi Hiwasa

Preprint
Proceedings of the 27th Annual Asilomar Conference on
Signals, Systems and Computers
Pacific Grove, California

November 1-3, 1993

CIPIC #93-14

University of California, Davis

Perceptual Criteria and Design Alternatives for Low Bit Rate Video Coding

V. Ralph Algazi, CIPIC, University of California, Davis, and
Norimichi Hiwasa, CIPIC-HQITL and Mitsubishi Electric Corp., Ofuna, Japan

Abstract

The design of very low bit rate coders, below 64 kilobits per second, presents a number of new challenges. Such low bit rate coders are targeted to small size images, say 176 X 144, or below and are limited to head and shoulders scenes.

In this paper, we compare and rate the performance of several still image encoding methods such as DCT, Subband and Wavelets, using a new quality scale based on properties of human vision. These still image coders are embedded into video coders such as the H261 coder. The relative importance of intraframe and interframe quality and contributions to the total bit rate provide an overall design framework for these low bit rate video coders.

We discuss and illustrate, for the encoding of low bit rate video, the still image and motion impairments that are perceptually most important.

1. Physical basis and design parameters in video coding.

The coding of video sequences is based on both the properties of human perception and on properties of the data that make it possible. Without repeating here an exposition of these basic properties, we will discuss immediately the typical structure and some of the options in video coding, and discuss the underlying principles and properties as needed.

Video coders have several components. An intraframe coder which embodies one of several methods for the representation and coding of still images. We will discuss and compare some of them later in this paper. Frame to frame redundancy is quite high in video. Thus motion compensation (MC) is used to estimate one of the frames in the sequence from one or several adjacent frames. The interframe coder is the embodiment of the efficient representation and coding of the motion compensation residuals, that will be generally quite small.

2. Design alternatives.

Consider, as a reference, the raw digital data rate of a standard television image sequence. For typical image size of 640 X 480 pixels, or 525 lines video, the number of pixels is about 0.25 Megabytes per frame. At 30 frames/second, the raw data rate is 60 Mbps, and somewhat more to account for chrominance. Means and options available for the drastic decrease of the bit rate, which is a primary interest in this paper are shown in Table 1. We consider in our discussion a subset of the options listed, with emphasis on the effect on performance and image quality.

3. Some proposed standards for low bit rate video.

Some images sizes have been adopted as tentative standards in a hierarchical format framework known as H.261, which consider video encoding at rates p* 64 kbps, where p is an integer [1].

Two of these video standards, known as Common Intermediate Format (CIF) and Quarter CIF (QCIF) are shown in Table 2.

Note that the raw bit rates at 30 frames per second are slightly more than 24 Mbps for CIF and 6 Mbps for QCIF. For a final bit rate at or below 64 kbps, we require a compression ratio slightly larger than 100:1 for QCIF video, at 10 kbps, the compression required would be more than 640:1. The targeted compression mandates strong restrictions on the type of image sequences that can be considered, following the framework of options of Table 1. The image size has to be a QCIF or smaller, limited camera motion is allowed, the scene is a simple "head and shoulders" video conference scene, and motion is limited.

4. Compression Goals and Means.

Let's consider briefly reasonable ranges of compression for the intraframe and interframe portions of the coder. Intraframe coding will provide substantial compression for the restricted type of scenes of interest. Clearly, we are not targeting here a very high quality image and a compression of 10 for a still image that is not too busy is reasonable. Below .3 bit/pixel, most of the known methods will result in images that are not generally acceptable. We target, for discussion, a compression of 20 for the intraframe portion of the coder, so that factors of 5 to 32 additional compressionare needed to lower the range of output rates to 64-10 kbps. An improvement factor of 5 can be achieved by interframe coding, but not a factor of 32.

Therefore, low bit rates will require decreases of both the frame rate and the image size, as well as interframe coding. Note however that lowering the frame rate will decrease the correlation between frames and, thus the effectiveness of interframe coding by motion compensation.

5. Image compression techniques and image quality for intraframe coding.

As discussed, we are targeting a compression ratio of 20 for the intraframe portion of the video coder, and therefore, we compress a QCIF image to about .4 bit/pel. Let us consider the elements of image quality at such a low bit rate taking as a prototypical example the H.261 coder.

A. Quality Factors:

This subjective quality evaluation is based on a number of impairments that can be observed in the encoded image. We identify the types of impairments and define corresponding distortion factors that can be objectively quantified.

Distortion factors are perceptually weighted measures of image impairments. These distortion factors are suggested by experience in observing artifacts due to coding and by knowledge of properties of the human visual system. A global subjective measure, the Mean Opinion Score (MOS), is a subjective assessment or ranking of a composite image quality.

We assume that MOS is a linear combination of observed disturbances D_i and that MOS is approximated by an objective picture quality scale (PQS) which is of linear combination of measurable distortion factors F_i.

The distortion factors F_i are functions of the difference between the original and reconstructed encoded image, so that PQS will be a measure of the degradation from an original. Properties of the visual system with respect to perception of luminance, contrast transfer function, and anisotropy of vision are used to weight the error e(m,n) between the original image and its encoded version [2]. We denote such perceptually weighted errors by $e_w(m,n)$.

a) Random Errors: Some random errors are visible in encoded images. We use two quality factors for these errors, obtained by weighted mean square errors measures. F1, makes use of the standard error weight used in television. F2 makes use of a more complete, two-dimensional and anisotropic perceptual weight [2].

We have observed that random errors are seldom the dominant factors in coded still frames. This is because the human visual system is more sensitive to patterns and structured misalignment errors in image than to random disturbances.

Distortion factors F1 and F2 are of the form

$$F2 = \frac{||e_w||^2}{||I||^2}$$

where $||I||^2$ is the mean square value of the image and $||e_w||^2$ is the weighted M.S. value of the error.

b) End of block effect: This effect is commonly observed in block transform coding. Define horizontal end of block discontinuity as

$$\Delta e_w(m,N) \triangleq [e_w(m,N) - e_w(m,N+1)]^2$$

where N is the block size. With a similar definition for the vertical discontinuity, we evaluate the end of block distortion factor as

$$F3 \triangleq [||\Delta e_w(M,n)||^2 + ||\Delta e_w(m,N)||^2]^{1/2}$$

c) Structured errors in any part of the image. We compute the average local correlation of weighted errors, Ry(M,N) in 8 X 8 neighborhoods to determine local error structure, and we define the distortion factor for these structure errors as

$$F4 = [||R_y||^2 + ||R_x||^2]^{1/2}$$

Thus, if errors are uncorrelated, F4 is zero.

d) Structure errors in the vicinity of high contrast edges. In the vicinity of high contrast transitions, visual impairments will be masked by the activity of the image. However, the largest errors due to the DCT also occur in the same portions of images. We define a distortion factor F5 that includes a masking function as well as a measure of image activity.

B. A composite quality metric: The Picture Quality Scale: (PQS)

The five distortion factors are highly correlated and a numerical Picture Quality Scale (PQS) is defined as

$$PQS \triangleq b_o + \sum_{j=1}^{3} b_j Z_j$$

where Z_j are the principal components of the covariance matrix of the Fi's and the b's are partial regression coefficients.

By multiple regression analysis, a correlation of 0.88 with the Mean Opinion Score (MOS) taken on a five point scale is determined. (The five point impairment scale is: 5=imperceptible; 4=perceptible but not annoying; 3=slightly annoying; 2=annoying; and 1=annoying.) This PQS provides a useful quality metric for still monochromatic images [2,3].

C. Comparison of intraframe coding techniques with PQS.

By using PQS we compare the quality versus bit rate curves of several coders at about 0.4 bit/pixel. Results are

shown in Figure 1 and indicate that all three coding methods have comparable performance, with a advantage for the DCT coder. Note that the performance of such coders depend on the quantization matrix used as well as on the error free coding strategy. We have used in each case methods reported in the literature [4,5,6]. However, the quality of the coded still image does not correlate well with observed quality of video as we construct image sequence from these still frames. For instance, we are quite sensitive to temporal changes in the average gray scale value in flat portions of images.

6. Coding techniques and quality effects in video and image sequence coding.

We now discuss briefly some techniques and quality factors in image sequence coding. Of necessity, the discussion is now more qualitative.

The intraframe coding of image sequences take advantage first of the small incremental information from frame to frame. If large changes occur from frame to frame, then spatial details cannot be perceived. If image motion is small and smooth, then the incremental information in a new frame can be predicted from previously encoded adjacent frames. This motion compensation (MC) is an estimation process. Because of the accumulation of distortions in interframe coding and because of the disastrous effect of transmission errors in the quality of images that have been coded differentially, some intraframe coded image are included periodically in the overall encoding scheme.

Quality factors for sequences:

Flicker. A very objectionable degradation of quality occurs when flicker is present. Although the flicker effect depends of somewhat of viewing conditions, keeping the flat field rate at 60 frames per second will reduce the flicker effect to tolerable limits. For a low frame rate video, the flat field rate is maintained above this limit by repeating the frames at the display.

Reproduction of motion. It is generally accepted that smooth motion can be approximated in an perceptually acceptable fashion if a sequence of 24 motion frames per second is sustained. Thus, if interlaced scan is not required, a frame rate of 24 frames per second is as high as needed for a smooth motion rendition. For low bit rates, such a frame rate cannot be achieved, and jerky motion will result.

Perceptual Sensitivity to Temporal Changes. We are interested in the combined spatial and temporal sensitivity of human visual perception. Combined spatial-temporal sinusoids can be used to study this 3 D contrast sensitivity function. A major result is that the perception of moving detail is higher if the moving detail is being tracked visually. The major effects used in coding temporal changes are to control rapid temporal changes of low spatial impairments, and to allow errors at higher spatial frequencies for rapid temporal changes.

The importance of these results to our discussion are in the change in perceptual effects when forming an image sequence from coded still images, which exhibit the types of visual impairments discussed in a previous section. To address the specifics of image quality, we consider the H.261 coder, designed for low bit rate video.

7. The H 261 Coder.

In the H.261 coder, 8 X 8 blocks are transformed with the DCT. Intraframe coding follows closely the JPEG still image standard [7]. A 2 X 2 group of blocks or macroblocks, representing an array of 16 X 16 pixels is used to determine motion compensation (MC). For each macroblock, the decision is made to encode it either by interframe or intraframe methods. The decision is also made, based on the motion compensated macroblock residual energy, whether to use motion compensation or not. The H.261 may be used as a fixed rate coder, so that the variable rate digital bit stream resulting of the encoding process feeds a buffer that maintains a fixed output rate. Buffer overflow results in drastic decisions in order to maintain the fixed bit rate. These range from adaptation of the quantizer step size to discarding whole macroblock data.

Experiments with the H.261 Coder. To examine the combined effect of all the possible causes of image impairments at low bit rates, we encoded a standard video sequence "Miss America" using the H.261 coder, with some of the options that will be of use in illustrating the image quality issues. The H.261 coder simulator is made available by the portable Video Research Group (PVRG) at Stanford University [8]. For a target bit rate, say 64 kbps, we have the option of buffer size, initial quantization step, and frame rate. We chose a 16,000 bit buffer that does not result in buffer overflow for our experiments, and thus allows normal operation of the coder.

We now consider several frame rates for this target bit rate. For high frame rates, the rendition of motion is improved, while for lower frame rates, the distortion of the isolated frames will be decreased, at the expense of increasing the jerkiness of motion.

Intraframe and Interframe Coding. It is informative to consider the contribution to the total bit rate in the H.261 coder. At 15 frames/sec, the interframe coder achieves approximately 3.5 higher compression than the intraframe

coder. Accounting for the chrominance information, the compression factor for the intraframe coder is about 20 as anticipated. The bits needed to encode both the intraframe and interframe information increase as the frame rate decreases.

8. Discussion of Image Quality issues.

The goals of our study were two-fold:

1. To determine the relative importance of each of the distortion factors, identified for still image coding, when these images are displayed in a time sequence.
2. To determine the importance of jerky motion at reduced frame rates.

With respect to the still image distortion factors, some conclusions are easy to reach for the low bit rate we have to maintain. The random error factors, F1 and F2 are not important because other effects are much larger. The end of block factor F3 may be quite detrimental, principally when the entire macroblock is suddenly degraded because of buffer overflow. It is not yet clear to us that this effect could not be mitigated by careful quantizer selection, and with acceptable buffer size and delay. The structured error factors F4 and F5 are the dominant factors in all cases. In smooth portion of the image, such as the face of Miss America, the factor F4 measures image blotchiness. The factor F5, measures visible blocking artifacts near higher contrast transitions, such as in the neck area of Miss America.

With respect to the effects due to motion, we also reach readily some preliminary conclusions. For a "head and shoulders" scene such as Miss America, the smoothness of motion is not a critical issue. Frame rates of 10 or even 5 frames/sec are more acceptable than a poor still image quality. If the structured impairments measured by F3, F4 and F5 are significant in the frames of an image sequence, their perceptual impact is increased. Structured errors move across portions of the image in a slow, random fashion are quite annoying. Thus the trade-off seems to favor strongly higher quality still images and a low frame rate to achieve the overall goals of a very low bit rate system.

9. Discussion and Conclusions

We examined the performance of the H.261 coder on a QCIF sequence at 64 and 32 kbps and 30, 15 and 5 frames/sec. At 64 kbps, 15 frames per second appear to achieve the best compromise between motion rendition and image quality. At 32 kbps, it is not possible to achieve 30 frames/sec at all, and 5 frames/sec is the preferred choice. To go below 32 kbps, we could use the H.261 coder with a smaller image, say 128 X 128, and achieve 20 kbps. Below that rate, it does not appear that the H.261 DCT based coder is suitable, or either is the block based motion compensation. Subband and wavelet coder may provide some incremental advantages because distortions track more closely the moving object. Other methods, such analysis based motion compensation or careful use of spatio-temporal visual perception, may have more promise, without going to the complexity of an image synthesis coder [9].

One of the objectives at CIPIC, is to pursue the development of quality metrics for image sequences coding. The current study was of help in providing a rough cut through some major issues.

References

[1] Ming Liou, "Overview of the px64 Kbps Video Coding Standard." Communications of the AIM, Vol. 34, No. 4, pp. 59-63, April 1991.

[2] M. Miyahara, K. Kotani and V. R. Algazi, "Objective Picture Quality Scale (PQS) for Image Coding," SID Symposium 92, Boston, MA, May 1992.

[3] V. R. Algazi, Y. Kato, M. Miyahara and K. Kotani, "Comparison of Image Coding Techniques with a Picture Quality Scale", SPIE Technical Program on Photonics Instrumentation Conference, Vol. 1771, San Diego, CA, July 1992.

[4] J. D. Johnston, "A Filter Family Designed for use in Quadrature Mirror Filter Banks," ICASSP, April 1980.

[5] I. Daubechies, "Orthonomal Bases of Compactly Supported Wavelets," Commun. on Pure and Applied Mathematics, Vol. XLI, 1988.

[6] J. W. Woods, Ed., "Subband Image Coding," Kluwer Academic Publishers, Norwell, MA, 1991.

[7] JPEG - W. B. Pennebaker and J. L. Mitchell, "JPEG, Still Image Data Compression Standard 1993", Van Nostrand Reinhold, New York.

[8] PVRG-P64 Codec 1.1, Andy C. Hung, Stanford University, June 1993.

[9] U. Golz and R. Schafer, "Considerations on the Possibility to Exchange Temporal Against Spatial Resolution in Image Coding," Signal Processing Image Communication 2, 1990, 39-51.

Acknowledgements

We are very grateful to Sandeep Shetty for assistance with the H.261 coder simulation.

We also acknowledge support of the UC Micro Program, Pacific Bell, Lockheed and Hewlett Packard.

A. Restrict the image and video parameters.	B. Improve the efficiency of the intraframe, or still image coder.	C. Improve motion prediction and compensation (MC).
a) The image content.	a) DCT/JPEG Standards	a) Block matching standards techniques.
b) The image size.	b) Wavelet.	b) Analysis based MC.
c) The image motion and frame rate.	c) Subband.	c) Interpolative MC.
d) The camera motion.	d) Other coders: Model based coding and texture synthesis.	d) Use image synthesis.

Table 1. Alternatives for Achieving Lower Bit Rates

Parameter	CIF		QCIF	
Screen Size	Assummed to be about 3 to 10 inches			
Screen Aspect Ratio	4:3			
Spatial Resolution	288(lines x 352(pixels) | 144(lines) x 176(pixels)			
Temporal Resolution	29.97 frames/s (progressive scanning)			
Colorimetry (lines x pixels)	Y	288x352	Y	144x176
	Cr	144x176	Cr	72x88
	Cb	144x176	Cb	72x88
Gradation	8 bits			
Transfer Speed	64 kbit/s - 1.5 Mbit/s, 2 Mbit/5			

Table 2. CIF and QCIF Standards in H.261

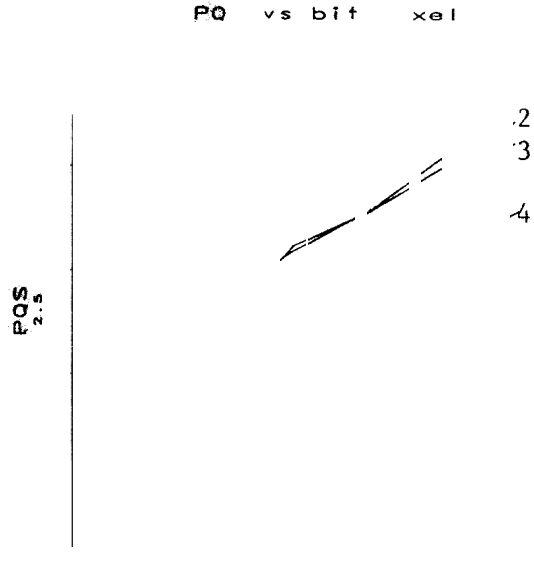


Figure 1.
1) DCT (H.261);
2) wavelet-(scalar Q + Huffman) / 16 dim LVQ)
3) subband-(scalar Q + Huffman)/16 dim LVQ)
4) subband-(scalar Q + Huffman)