PROCEEDINGS REPRINT

SPIE—The International Society for Optical Engineering

Reprinted from

Human Vision and Electronic Imaging II

10–13 February 1997
San Jose, California

SPIE
P
PROCEEDINGS SERIES

Volume 3016

Test patterns and quality metrics for digital video compression

Charles Fenimore, Bruce Field, and Craig Van Degrift

National Institute of Standards and Technology*
Gaithersburg, Maryland, 20899

ABSTRACT

Lossy video compression systems such as MPEG2 introduce picture impairments such as image blocking, color distortion and persistent color fragments, "mosquito noise", and blurring in their outputs. While there are video test clips which exhibit one or more of these distortions upon coding, there is need of a set of well-characterized test patterns and video quality metrics. Digital test patterns can deliver calibrated stresses to specific features of the encoder, much as the test patterns for analog video stress critical characteristics of that system. Metrics quantify the error effects of compression by a computation.

NIST is developing such test patterns and metrics for compression rates that typically introduce perceptually negligible artifacts, i.e. for high quality video. The test patterns are designed for subjective and objective evaluation. The test patterns include a family of computer-generated spinning wheels to stress luminance-based macro-block motion estimation algorithms and images with strongly directional high-frequency content to stress quantization algorithms. In this paper we discuss the spinning wheel test pattern. It has been encoded at a variety of bit rates near the threshold for the perception of impairments. We have observed that impairment perceptibility depends on the local contrast. For the spinning wheel we report the contrast at the threshold for perception of impairments as a function of the bit rate. To quantify perceptual image blocking we have developed a metric which detects "flats": image blocks of constant (or near constant) luminance. The effectiveness of this metric is appraised.

Keywords: digital video compression, quality metrics, test patterns, image blocking, flats

1. INTRODUCTION

The MPEG2 video compression standard[1] is being adopted widely and is finding use as the video coding method for several existing or planned commercial systems. The MPEG2 standard specifies the characteristics of the coded bitstreams and of decoders. Encoders are not specified and there is considerable freedom in encoder design and implementation. This freedom and the general complexity of MPEG systems means that testing and characterization of performance for any specific encoder is similarly complex.†

In testing proposed MPEG2 algorithms, the developers of the standard employed a set of video clips. Most of these clips originated as cinema or video, although some synthetic test patterns were used at various stages of development. The clips contain a variety of source materials and stress the encoder in a variety of ways. The material includes many levels of motion, the uncovering of hidden objects from one frame to the next, and various levels of image detail, transparency, and saturation of colors.

The principal advantage of synthetic test patterns is they can be tailored to test specific features of video systems. While synthetic patterns generally have limited visual appeal for the casual viewer-evaluator, they offer other advantages over camera-originated clips for testing and measurement. Evidence of their utility is seen in their 50 years of use in the technical evaluation of analog systems in the studio. In addition, when a synthetic test clip can be generated computationally at a test site, there is no concern about loss of fidelity in the distribution process. This eliminates the need to maintain a digital pedigree, that is a

*Electricity Division, NIST, Technology Administration, U. S. Department of Commerce.

†In this paper MPEG without an index refers to MPEG2 or MPEG1. The differences between the standards are described on the World Wide Web[2].

record of the "ancestors" of the particular test material. Such records are essential in assuring the quality of video test materials which are subject to copying.

Recently, test patterns have been used to establish MPEG2 bit stream and decoder compliance[3] and for other specific MPEG diagnostics. One example of the latter is the test pattern reported in Ladebusch[4] which is designed to permit visual determination of the quantization parameter which adaptively changes the resolution (and thus the bit rate) of transformed image data in MPEG and JPEG[5] encoders. Our test patterns have been developed for the evaluation of MPEG2 encoder performance. Here we are concerned with image blocking impairments, the origins of which in MPEG2 are identified in the Impairments Section. The details of the spinning wheel and the blocking impairments which it induces are presented in the Test Patterns Section.

Having identified the impairments to be measured, we next describe our "flats" blocking detector. The definition of the detector and its optimization are presented in the Metrics Section and the detector is evaluated. We find that it is markedly less effective when it is applied to video frames which have been compressed using motion estimation than when applied to frames which have not. There are alternative methods to detecting image blocking. One approach with a statistical character is found in Melcher and Wolf[6]. Another, which is not restricted to the identification of blocking impairments, is found in Lubin's work on just noticeable differences[7]. The Conclusion discusses various approaches to the detection problem and outlines possible future work.

2. IMPAIRMENTS

Analysis of the MPEG2 standard and experience with compressed video suggests that for MPEG2 compression systems, the quantization and motion estimation stages are two main contributors to impairments. The most serious impairments appear as image blocking. Blocking arises from quantizing too coarsely the coefficients of the discrete cosine transform (DCT) and from failure of motion search to find good motion estimates.

2.1 I-frames and DCT compression blocks in MPEG2

The DCT and quantization stages of MPEG2 compression can introduce blocking impairments into video frames on the scale of the 8 x 8-pixel blocks into which each frame is decomposed. MPEG2 groups 4 DCT blocks into a single macroblock. Each macroblock is handled in one of two modes: intraframe (I-frame) compression mode in which the 4 DCT blocks making up the macroblock are encoded without reference to other frames in the video sequence and interframe mode (discussed below) in which motion estimation is used. For I-frames all of the macroblocks are DCT-encoded without motion estimation. Picture information is lost in quantizing the transform coefficients.

For an image with pixel values $s(p, q)$ on an N x N block, the 2-dimensional DCT, $S(j, k)$, is defined by:

$$S(j, k) = \frac{2}{N} \cdot C(j) \cdot C(k) \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} s(p, q) \cos\left(\frac{\pi(2p+1)j}{2N}\right) \cos\left(\frac{\pi(2q+1)k}{2N}\right), \text{ where} \quad (1)$$

$$C(k) = \begin{cases} 1/\sqrt{2} & \text{for } k = 0 \\ 1 & \text{otherwise} \end{cases}$$

and both j and $k = 0 \ldots N-1$. In MPEG2, N is taken to be 8.

The quantization of the coefficients, $S(j, k)$, occurs through integer division by the factor, $M_{QUANT} \cdot Q(j, k)$. The matrix Q is fixed while the parameter M_{QUANT} is set in a feedback loop to provide control of the bit rate. The quantization is coarser, that is $Q(j, k)$ is larger, for higher values of j and k (i.e. for higher frequencies). Similarly, as M_{QUANT} increases, the coefficients are represented with less resolution. This loss of resolution can produce visible image blocks. The blocking impairment detector which we have developed attempts to exploit the appearance of a large number of zero coefficients in quantized video frames.

2.2 Motion estimation and blocking

In interframe mode the DCT is not applied to the original frame but to the residual image formed as the difference between the original frame and a motion-estimated frame. Doing so introduces a new class of blocking impairments. Although there are two types of interframe macroblocks, predicted (P-frames) and bidirectional (B-frames), in each case motion estimation is used to find an estimate of each 16 x 16 pixel macroblock. A macroblock in the encoded (or target) frame is compared with linear translates of equal-sized blocks in encoded frames which precede and/or follow it. The block which most closely approximates the target macroblock is used as an initial estimate of the target block. The associated translation gives the value of a motion vector.

DCT encoding is applied to the residual macroblock. For rotational motion, like that in the spinning wheel, which is not tracked well by the motion estimator, portions of the residual image can be more difficult to compress than the corresponding parts of the original frame. The implementor of an MPEG encoder is free to design a search method for the motion vectors, however, it is common for motion estimation to be based on the luminance portion of the video source. This is the case for the Test Model 5 encoder[8] which we examine in this report. Because the DCT is applied to a residual which itself may contain color and luminance blocks inherited from the translated matching blocks, the structure of the blocking may be more complex for interframes than for I-frames.

3. TEST PATTERNS

3.1 Spinning wheel test pattern

In light of the preceding analysis of the MPEG2 encoder, we have created a test pattern with rotating motion and saturated color. Figure 1 shows one frame from the *Spinning Color Wheel* test clip. This clip is one from a family of computed test patterns. Each clip is defined by a number of user selectable attributes, the values of which for the subject clip are noted parenthetically:

- the number of frames in the test clip (20);
- the scanning mode of the frames, either 30 progressive frames per second (F/s) or 60 interlaced fields per second (f/s). In interlaced mode the algorithm generates intermediate images at 60 frames per second and pairs the lines to produce first a field of even numbered video lines and then of odd. (30 F/s, progressive);
- the background luminance (BGL), which is spatially constant across each field or frame, may either be held constant in time or may vary linearly from frame-to-frame with a range equal to the full CCIR 601[9] luminance dynamic range (BGL is constant on each frame and ranges linearly from 19 to 235 from beginning to end of the clip);
- the number of wheels in the clip (1);
- the number of paddles on each wheel (3);
- the color and initial angular size and position of each paddle (green, blue, and red paddles each of 60°, equally spaced);
- the radius of the wheels (200 pixels);
- the rate of rotation of each wheel and the velocity of translation expressed in units per frame (10° / frame counterclockwise, (10,0) pixels / frame); and
- the presence of motion blur or not (blur).

The video frames in the clip examined here have 720 x 486 pixels with 30 frames/second. When generated in interlaced mode the video frames conform to CCIR 601.

Several features of the clip are challenging to MPEG encoders. The frame-to-frame change in the luminance of the background simply adds a DC component to the motion compensated residual in regions away from the moving wheel. However, for image macroblocks at the edge of the wheel, the rotating motion assures the residual in the luminance will have a large non-DC component. An exception occurs when the luminance of the wheel closely matches that of the background, in which case there may be large chrominance components to the residual. Such a residual represents a failure of the motion estimation to reduce the bit rate. In either case there is a large residual for the DCT compression stage. The image impairments seen in P- and B-frames are qualitatively different then those in I-frames. Coding errors can show up as mispositioned color patches and as image blocking or added edges. Motion estimation can insert spurious elements into the video images which are only removed by the DCT/quantization stage at relatively high bit rates.

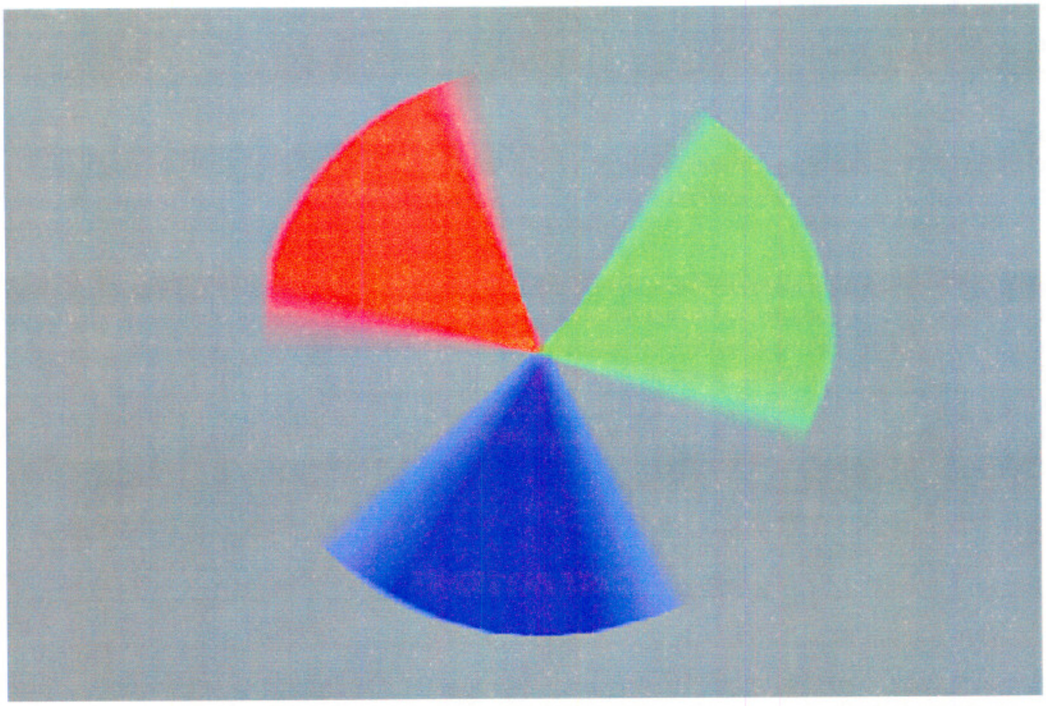

Figure 1: Frame 11 from the spinning wheel test pattern. The gray level of the background changes from frame to frame and the wheel rotates and translates with motion blur. The solid portions of the paddles have saturated RGB.

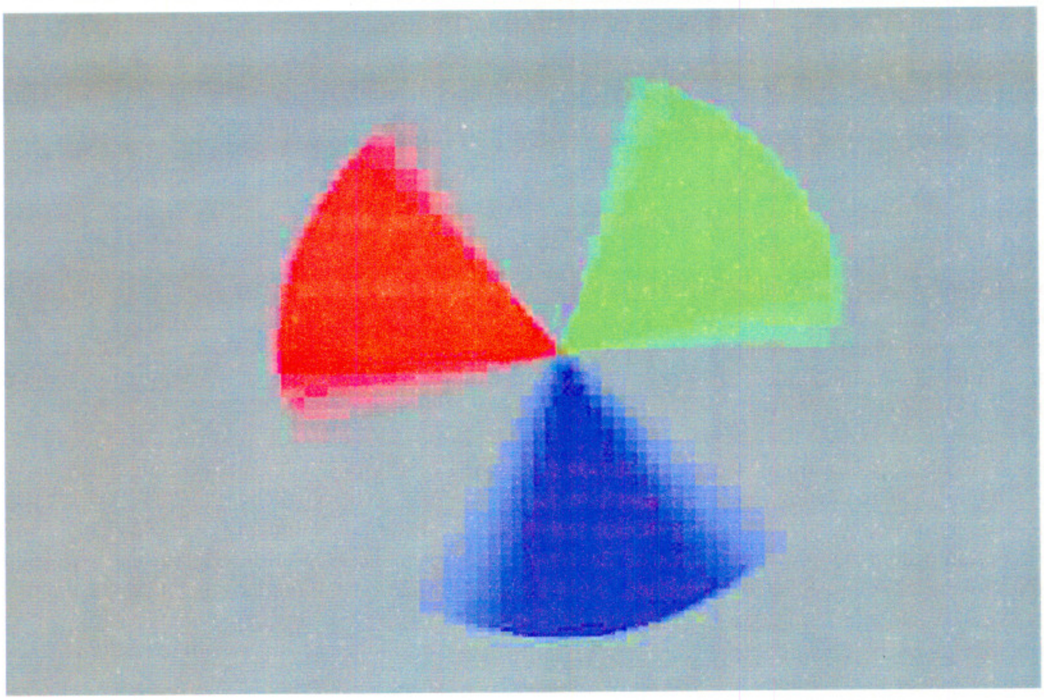

Figure 2: Frame 12 (I-frame) from the spinning wheel test clip MPEG2-compressed at 1.7 Mb/s. Image blocking is seen along the trailing and leading edges of each paddle as a regular, rectangular pattern of edges. The smallest block size is 8 x 8 pixels.

3.2 Encoding of test patterns and rating of impairments

In order to determine the threshold bit rate at which perceptible impairments appear, the test clips have been MPEG2 encoded for a variety of bit rates. With the exception noted below, the public domain Test Model 5 (TM5) encoder/decoder, which we used and is associated with the MPEG2 committee, proved to be relatively reliable and stable. The compression was carried out at the main level, main profile, with a group of pictures (GOP) structure IBBPBB. It was found that at start-up of the clip, the first two B-frames (frames 2 and 3), exhibited impairments at high bit rates. For this reason our data was generated by appending an extra GOP at the beginning of the test clip which was discarded before collection of our data. The anomaly may be associated with the initialization of buffers in the encoder.

Figure 2 shows typical impairments in an I-frame taken from the Test Wheel. For this figure, the encoding rate was reduced to 1.70 Mb/s, where the blocking impairments are readily visible. When measuring the threshold of perceptibility of the impairments significantly higher bit rates were used. The conditions of the testing were made close to those specified in CCIR Recommendation 500[10] where possible. Thus the illumination of the room, the reflected light from the display, and the viewing distance (4 x screen height) conformed to Rec. 500. We used a 19" Sony PVM-1944Q professional monitor for viewing. It did not have SMPTE phosphors. Also, contrary to Rec. 500, we used a single viewer. This proved to be adequate for the purpose of evaluating the "flats" blocking detector.

We measured the threshold for perception of impairments on the static frames and found separate thresholds for I-, P-, and B-frames as presented in Table 1. The threshold bit rates differ significantly, by more than the indicated resolution of 0.5 Mb/s. The variation is a consequence of the bit-allocation scheme of Test Model 5 and may indicate directions for improvement of that implementation of MPEG2, as noted in the conclusions. In the current study, the data is used as a guide in the selection of test material with just noticeable impairments.

Table 1: Threshold bit rate for perceptibility of impairments in the MPEG2 compressed Spinning Test Wheel. The indicated bit rates are measured with a resolution of $\pm$ 0.5 Mb/s.

	I-frames	P-frames	B-frames
Bit rate at the threshold for perception (Mb/s)	3.0	5.0	6.0

4. METRICS

4.1 A model for perception of blocking impairments - flats

The flats detector was developed based on the observation that the DC component, $S(0, 0)$, of the DCT (1) appears to dominate many of the image blocks in our test patterns. The detector searches a video frame for flat image blocks in which the luminance, Y, is constant and differs from the surrounding pixels, or is constant in x or in y. The detector is luminance-based, that is it attempts to account for chrominance-based blocking effects by its effect of the luminance. The difference between the level inside the block and in the adjacent pixels is a threshold parameter, T, in our model.

A *flat* is an 8 x 8 block of pixels having constant luminance, Y_0, inside the block cornered at (J, K), having pixels indexed by (j, k), $j - J$ and $k - K = 0 \ldots 7$, for which

$$Y(j, k) \equiv Y_0, \text{ on the block and}$$

$$|Y(j, k) - Y_0| > T, \text{ at one or more of the 32 adjacent points, } (j,k)$$

where $(j\text{-}J = -1 \text{ or } 8, \text{ and } k\text{-}K = 0 \ldots 7)$ or $(k\text{-}K = -1 \text{ or } 8, \text{ and } j\text{-}J = 0 \ldots 7)$. In the present work, $T = 10$. The technique for setting T is discussed in the succeeding section on contrast.

Similarly, a *horizontally ruled flat* is an 8 x 8 block of pixels having constant luminance, $Y_0(k)$, along each horizontal line inside the block,

$$Y(j,k) \equiv Y_0(k), \text{ on the block and}$$

$$|Y(j,k) - Y_0(k)| > T, \text{ at one or more of the points adjacent to the row with}$$

k fixed and j-J = -1 or 8. A *vertically ruled flat* is defined similarly.

4.2 Visibility of MPEG2-induced image blocking, the effect of contrast

In order to set the threshold parameter in the flats detector, we have studied the visibility of the blocking impairments when the contrast of an impaired frame is varied, keeping the mean luminance unchanged. For a fixed image, with luminance and chrominance components (Y,U,V), one can interpolate between the image and the flat gray image with level luminance and chrominance, (Y_{av}, U_0, V_0) where $Y_{av} = Average(Y)$ and $U_0 = V_0 = 128$. The interpolated frames will have luminance and chrominance given by the linear combination

$$(Y, U, V)_t = t \cdot (Y, U, V) + (1-t) \cdot (Y_{av}, U_0, V_0).$$

The non-dimensional contrast parameter, t, ranges on the interval [0, 1]. At $t = 1$ one has the initial image, while for t=0 one has a flat gray image having the same average luminance. It is apparent that any blocking seen in the original will be extinguished as the contrast is reduced to zero

Using the I-frame (frame #12) taken from the Spinning Wheel test clip compressed at four different bit rates, we have applied the contrast reduction to find the contrast threshold for perception of blocking as a function of bit rate. Table 2 shows the value of the contrast parameter at the threshold for each of the color components (RGB)

Table 2: Threshold to which contrast must be reduced to suppress visible blocking. The contrast parameter is measured with a resolution of 0.05.

bit rate (Mb/s)	green blocking	red blocking	blue blocking
1.70	0.20	0.05	0.05
2.00	0.55	0.15	0.10
2.50	1.00, no blocking	0.60	0.55
3.00	1.00, no blocking	1.00, no blocking	1.00, no blocking

Using the thresholds found in the above measurements, we apply the flats detector, with various values of the parameter T, to the contrast-reduced images. For values of T near 0, the *flats* detector finds many false positives in images for which no blocking is visible. For values of T near 10, there is a great reduction in such false positives. Further increasing the value of T can lead to false negatives, that is it may cause a failure to detect blocking. For this reason we have set $T = 10$ in the blocking detector.

4.3 Evaluation of the *flats* detector

Figure 3 shows the luminance signal for the impaired *Spinning Wheel* I-frame seen in Figure 2. It has been overlaid with white markers showing the flats (small squares), and the horizontally and vertically ruled flats (horizontal and vertical lines, respectively.) The detector misses about 5% of the blocks, those seen in the trailing edge of the green paddle. The luminance values shown in Figure 3 suggest the reason for this false negative: the luminance of the green is very close to that of the gray background. The hypothesis that the luminance channel will identify all chrominance blocks fails when the changing background nearly matches the luminance of the wheel. The problem cannot be corrected by increasing the sensitivity of the detector, i.e. by decreasing the threshold parameter, T. The detector behaves similarly for smaller values of T.

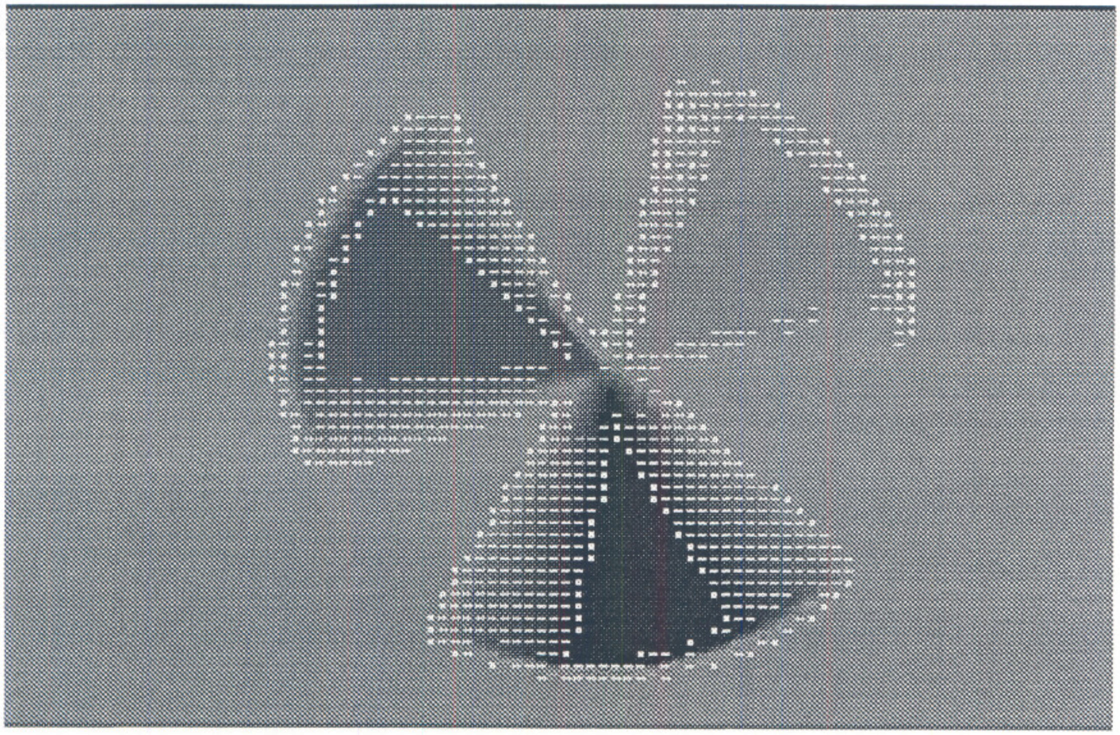

Figure 3: The *flats* blocking detector applied to the I-frame in Figure 2. The blocking seen on the trailing edge of the green paddle in Figure 2 (right center), is barely perceptible in this luminance-only image and thus is not readily detected by the *flats* detector.

The problem with false negatives is even more severe in the case of P- and B- frames for which the detector fails to find over 25% of the visible blocks. In addition to the problem that chrominance positioning errors go undetected by the luminance-based flats, the geometric character of the blocks is more complex for P- and B-frames. For I-frames the edges of the DCT blocks are known beforehand to fall along pixels with position indexed by multiples of the 8-pixel-long sides of the block. As noted earlier, for P- and B-frames the DCT is applied to a residual that will contain any blocking impairments found in the associated motion macroblock. The motion of the macroblocks can destroy the 8 x 8-pixel coarse graining on which the flats detector depends.

5. CONCLUSION

We report here on one synthetic video test clip for stressing MPEG encoders. When compressed and decompressed with Test Model 5, the test pattern exhibits DCT blocking and motion blocking impairments, the severity of which is a decreasing function of the bit rate. Current work is focused on additional test patterns in order to exercise other aspects of the MPEG encoding process. For example, the effects of noise in the encoding process will be studied by its addition to the existing patterns.

The test wheel has also proven to be useful in revealing several weaknesses and performance limitations in Test Model 5. The startup problems with the first few B-frames noted in the third section should lead users of the current algorithm to avoid using the first few B-frames. The variations in bit rate at the threshold for impairment perception in B-, P-, and I-frames, seen in Table 1, suggests that bits are being allocated to I-frames which might better be applied to intercoded frames.

The *flats* blocking detector is a modest success for I-frames. While it fails to report some blocks, there seems to be an opportunity to upgrade the detector by using color information. In the case of B- and P-frames, the higher rate of false negatives appears to be based on more fundamental limits of the detector and alternative techniques are being studied. One alternative is found in the work of Melcher and Wolf which is moderately successful with the spinning wheel test pattern but is less suitable for test patterns with strongly directional spatial patterns, e.g. grained patterns or zone plates. Another is the just noticeable differences (JND) of Lubin. We hope to compare the performance of the flats detector to that of the JND. The development of a better blocking detector for interframe blocking is an immediate challenge for video quality metrics.

REFERENCES

1. ISO/IEC 13818-2, *Generic Coding of Moving Pictures and Associated Audio - Part 1: Video*, International Organization for Standardization, [1995].

2. *MPEG Moving Picture Expert Group FAQ* (Frequently Asked Questions), Web site located at *www.crs4.it/~luigi/MPEG/*.

3. ISO/IEC 13818-4, *Generic Coding of Moving Pictures and Associated Audio - Part 1: Compliance Testing*, International Organization for Standardization, [1995].

4. Uwe Ladebusch, *Visual Evaluation of the Encoding Parameters of JPEG and MPEG Encoders*, preprint, Universitaet Braunschweig (Braunschweig, Germany), [1996].

5. G. Wallace, "The JPEG still image compression standard", *Comm. of the ACM*, 34:4, pp 30-44 [1991].

6. D. Melcher and S. Wolf, *Objective Measures for Detecting Digital Tiling*, contribution to ANSI,ANSI-T1A1.5/95-104, [1995}.

7. J. Lubin, *A Visual Discrimination Model for Imaging System Design and Evaluation*, David Sarnoff Research and Development Report, Princeton NJ, [1995].

8. ISO/IEC JTC1/SC29/WG11/N0400, *Test Model 5 (draft)*, MPEG93/457,[1993] Software is available by FTP on the World Wide Web at ftp://ftp.crs4.it/mpeg/programs/.

9. CCIR Rec. 601-2, *Encoding Parameters of Digital Television for Studios.*

10. CCIR Rec. 500-4, *Method for the Subjective Assessment of the Quality in Television Pictures* [1990].

Disclaimer: Certain commercial products are mentioned in this paper. Such mention should not be construed as an endorsement of the product by NIST, nor is it a statement that the product is the best for the purposes of the experiments described here.