Visibility Thresholds for Compression-Induced Image Blocking:
Measurement and Models

John M. Libert and Charles P. Fenimore
National Institute of Standards and Technology, 100 Bureau Drive, Stop 8114
Gaithersburg, MD 20899-8114

ABSTRACT

The present investigation compares performance of two objective video quality metrics in predicting the visual threshold for the detection of blocking impairments associated with MPEG-2 compression. The visibility thresholds for both saturated color and gray-scale targets are measured. The test material consists of image sequences in which either saturated color or gray-scale targets exhibiting blocking are varied in luminance contrast from -44 dB to -5 dB against a constant gray background. Stimulus presentation is by the "method of limits" under International Telecommunications Union (ITU) Rec. 500 conditions. Results find the detection of blocking impairments at Michelson contrast levels between -28 dB and -33 dB. This result is consistent with values reported by other investigators for luminance contrast detection thresholds. A small, but statistically significant difference is found between the detection threshold of saturated color patterns versus luma-only images. The results suggest, however, that blocking impairment detection is controlled mainly by display luminance. Two objective metrics are applied to gray-scale (luminance-only) image sequences, yielding measures of perceptible image blocking for each frame. A relatively simple blocking detector and a more elaborate discrete cosine transform (DCT) error metric correlate well over the contrast range examined. Also, the two measures correlate highly with measured image contrast. Both objective metrics agree closely with visual threshold measurements, yielding threshold predictions of approximately -29 dB.

Keywords: blocking, DCT error, flats, quality metrics, video compression, visibility threshold, test patterns

1. INTRODUCTION

Current video compression systems such as MPEG-2 are recognized to induce various picture distortions or impairments such as image blocking, color shifts and fragmentation, "mosquito noise", and blurring in their decoded outputs. Even as compression methods are improved, persistent demands upon available bandwidth will require that video compression always push against the limits of acceptable quality. The need to monitor and adjust the quality of compressed video close to the limits of viewer acceptability has driven the development of new methods for the evaluation of picture quality.

The types and magnitudes of picture impairments in a video sequence derive from the complex interaction of the image encoding method and the spatio-temporal characteristics of the input image sequence. Because of this complexity, it often is difficult to predict precisely how a particular encoder/decoder (codec) will perform in response to an uncharacterized video sequence. Hoping to better understand the origin of various video impairments and to better control video quality measurements, NIST researchers have developed test materials to induce impairments and metrics by which to quantify them.

For most applications, the measure of acceptable picture impairment is the degree to which the degradation is noticeable to human viewers. However, it is impractical to employ human viewers to continuously monitor the quality of, for example, a digital television broadcast. Accordingly, many proposed video fidelity metrics attempt to model the human visual system to at least some degree. A question of interest to the present investigators is the degree of sophistication required of a visual model in order to quantify narrowly defined image defects.

In the present investigation, measurements are made of the visual threshold for the detection of blocking impairments by a small group of subjects. The viewers are presented with both color and grayscale (luminance only) fixed images exhibiting...
blocking impairments, characteristic of DCT-based compression, at varying levels of contrast. Then, two objective computational metrics having different levels of complexity are applied to the same video test materials, and the objective threshold predictions are compared to the subjective measurements.

2. BLOCKING IMPAIRMENT VISIBILITY THRESHOLDS

2.1 Test Sequence
A sequence of twenty digital video frames of the NIST spinning wheel test pattern [2] was compressed using the public domain software video codec, Test Model 5 (TM5) [3]. Compression was carried out at the main level, main profile, with a group of pictures (GOP) structure IPPBB. The size of the MPEG-2 data file indicated an effective compression bit rate of 1.8 Mb/s. Figures 1 and 2 show a source image and the corresponding MPEG-2 processed frame as described. A single frame, exhibiting a relatively high level of blocking impairment, was selected as the basis for the generation of a test sequence in which the contrast of the image was varied gradually.

Figure 1. Test wheel compressed at 1.8 Mb/s.

Figure 2. Source image.
Using an initial fixed image, as in Fig. 2, with luma and chroma components \((Y', C_b, C_r)\), frames of the test sequence were constructed by interpolating between the impaired image and a flat gray image having luma \(Y'_0\) and chroma \(C_{b0}, C_{r0}\),

where \(Y'_0\) = average luma of the frame, and \(C_{b0} = C_{r0} = 128\). The interpolated frames, thus generated, have luma and chroma values given by the linear equation

\[
(Y', C_b', C_r')_t = t \cdot (Y', C_b, C_r) + (1 - t) \cdot (Y'_0, C_{b0}, C_{r0}).
\]

(1)

The non-dimensional contrast parameter, \(t\), ranges on the interval \([0, 1]\). At \(t = 1\), one would have the initial image, while for \(t = 0\), one would have a flat gray image. (Please note that in this paper, we adopt the convention of Poynton [1] in referring to non-linear encoded video luma as \(Y'\) and the luminance, i.e., the luminous flux per unit area, as \(Y\).)

In the present experiment, a sixty frame sequence was generated with display luminance contrast levels ranging from -45 dB to -5 dB. The contrast of the images was computed as

\[
C_{(dB)} = 20 \log_{10} \frac{Y_{\text{max}} - Y_{\text{min}}}{Y_{\text{max}} + Y_{\text{min}}} \text{ (dB)}
\]

(2)

where \(Y_{\text{max}}\) and \(Y_{\text{min}}\) are the maximum and minimum luminance values of the displayed image obtained by mapping video luma, \(Y'\), to display luminance values.

Both color and gray-level sequences were generated using interpolation. For the gray level images, the value 128 was substituted for the chroma components while the luma values were retained such that the displayed RGB image would exhibit gray values proportional to the luma only.

2.2 Visual Threshold Measurements

The video sequences were presented to subjects on a 19" (48 cm) Sony PVM-1944Q** professional monitor at a distance of four screen heights. Viewing conditions were as specified in recommendation ITU-R BT.500-8 [4]. Image sequences were stored on an Accom WDS XL Work Station Disk** controlled by a Silicon Graphics Indigo 2** workstation. The control software enabled the examiner to advance to specific video frames, control the rate of frame advance, and advance the sequence in either forward or reverse directions.

Subjects included five males between the ages of 25 and 55 years having normal visual acuity1 and color perception. One of the subjects was experienced in examining video and still imagery, and two others were accustomed to critical measurement of video display devices. Each subject was tested individually.

At the beginning of the session, each subject was told that he was invited to participate in an experiment designed to study the visibility of certain types of video impairments resulting from video compression. The subject was shown an example video sequence similar to the test sequence, but with coarser gradation. The examiner advanced frame-by-frame through the example sequence to illustrate to the subject the blocking impairments, the gradual reduction in contrast, and the eventual fading of the blocking structure into the gray background. The characteristics of blocking were pointed out to the subject on the display monitor, and were described verbally as the "vertical and/or horizontal structures seen in the images." The subject was told that the image frames would be advanced slowly through the sequence and that he should indicate when the blocking pattern first became visible. Again, the frame number was recorded, and the sequence was advanced further in the same

---

1 20/20 vision, either natural or corrected with glasses or contact lenses.
direction until the subject confirmed that the blocking pattern was clearly visible. The pair of ascending and descending passes constituted a single trial.

The above procedure was repeated for at least two and up to a maximum of five trials. The test for a particular stimulus was ended short of the maximum number of trials if two successive trials yielded threshold values within 1dB.

The threshold measurement procedure was performed separately for each of the three saturated color regions of the color test pattern shown in Fig. 2 and for each of the corresponding regions of the gray-scale luma channel version of the test pattern. The order of the tests was varied over the six subjects. The total of six tests took between thirty and forty-five minutes for the typical subject.

2.3 Display Contrast Measurement and Calibration

To control for peculiarities of the display system and for quantization errors in conversion of digital values to displayed video, contrast values used in the study were calculated from direct measurements of each frame of the test patterns as displayed on the test video monitor under the standard viewing conditions. Using a Minolta, LS-100 luminance meter, measurements were made of the saturated central zone of each of the colored paddles and the adjacent gray background regions. Corresponding measurements were made in the paddle and background regions of the gray-scale test sequence. The luminance value for each image region was taken as the arithmetic average of three sequential readings of the LS-100. For each color (or gray-level region), the Michelson contrast was calculated according to the following expression:

$$C_M = \frac{Y_{\text{background}} - Y_{\text{paddle}}}{Y_{\text{background}} + Y_{\text{paddle}}}$$

The present paper refers to contrast values expressed both in units of percent, \(C(\%) = C_M \times 100\%\), and in decibel units (dB), \(C(\text{dB}) = 20 \log_{10} C_M\).

For the objective measurements (described in a later section), it was necessary that the computations be performed on displayed image luminance values rather than on the encoded luma values. Regression analysis was performed according to the method described in [6] to define the log-linear function parameters by which to map encoded luma values 0 – 255 to display luminance for each of the RGB display color guns and for white. Thus, the image luma values, \(Y'\), were converted to display luminance values, \(Y\), prior to processing by the objective metrics described below.

2.4 Blocking Sensitivity Threshold Results

The threshold value for each trial was taken as the average stimulus contrast of the two video frames selected in ascending and descending passes. The threshold for each subject was computed as the average of his trials for each stimulus. Table 1 summarizes the threshold measurements for color and gray level image sequences. The Michelson contrast for each subject is shown for both color and luma-only gray-scale sequences. Values are provided in units of percent contrast and dB. Figure 3 exhibits the threshold measurements distributed according to contrast value on the ordinate and color (or corresponding luma-only pattern) categories indicated on the abscissa. Mean values with standard error bars are shown.

Threshold differences between color and gray-scale versions of the test patterns are examined via Student's t-tests [7]. Separate comparisons are made for the three colors and are summarized in Table 2. The difference in thresholds for color versus gray-scale patterns is computed for each subject, and the mean of the differences is tested against an expected mean difference of zero. These results indicate a significant difference in thresholds for the red test pattern at the \(\alpha = 0.05\) level.

Though not quite meeting the \(\alpha = 0.05\) criterion, the result for the green test pattern suggests a difference as well.

It should be noted that while measurements are reported for each of the three color segments of the wheel test pattern, a comparison of the thresholds among the three colors is not supported by the experiment. The blocking pattern induced by

2 Although the expanded uncertainty with a coverage factor of \(k=2\) for the measurement instrument employed is 4\%, the non-linearity of the device over the range of measurements is well below 0.2\%. The principal uncertainty in these contrast measurements arises from the non-uniformity of the CRT display, both in position and in time, and can be as high as 5\%. This provides an estimate of contrast measurement expanded uncertainty with a coverage factor of \(k=2\) of 7\%.
MPEG-2 compression was similar for red and blue segments, but not necessarily identical. A rather different pattern entirely was generated in the green segment. The investigators suspect that the pattern ambiguity in the green segment, and particularly in its gray-scale version, was responsible for the large variance among thresholds for this segment.

Table 1. Blocking impairment detection thresholds.

<table>
<thead>
<tr>
<th>Pattern Segment</th>
<th>Subject</th>
<th>Contrast Threshold Color (%)</th>
<th>Contrast Threshold Gray (%)</th>
<th>Contrast Threshold Color (dB)</th>
<th>Contrast Threshold Gray (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>5</td>
<td>3.641</td>
<td>3.000</td>
<td>-28.77</td>
<td>-30.46</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>3.076</td>
<td>3.088</td>
<td>-30.24</td>
<td>-30.21</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>3.641</td>
<td>3.000</td>
<td>-28.77</td>
<td>-30.46</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4.112</td>
<td>3.220</td>
<td>-27.72</td>
<td>-29.84</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td></td>
<td>0.5565</td>
<td>0.1104</td>
<td>3.79</td>
<td>4.06</td>
</tr>
<tr>
<td>Green</td>
<td>5</td>
<td>7.515</td>
<td>1.591</td>
<td>-22.48</td>
<td>-35.97</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>3.215</td>
<td>1.801</td>
<td>-29.86</td>
<td>-34.89</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2.462</td>
<td>0.362</td>
<td>-32.17</td>
<td>-48.81</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.081</td>
<td>0.795</td>
<td>-39.32</td>
<td>-41.99</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3.632</td>
<td>1.801</td>
<td>-28.80</td>
<td>-34.89</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>3.581</td>
<td>1.1514</td>
<td>-29.41</td>
<td>-40.50</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td></td>
<td>2.4039</td>
<td>0.8003</td>
<td>6.10</td>
<td>8.57</td>
</tr>
<tr>
<td>Blue</td>
<td>5</td>
<td>1.782</td>
<td>3.300</td>
<td>-34.98</td>
<td>-29.63</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>2.493</td>
<td>2.440</td>
<td>-32.07</td>
<td>-32.25</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2.451</td>
<td>1.533</td>
<td>-32.21</td>
<td>-36.29</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.477</td>
<td>3.072</td>
<td>-32.12</td>
<td>-30.25</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.438</td>
<td>2.193</td>
<td>-32.26</td>
<td>-33.18</td>
</tr>
<tr>
<td>Mean</td>
<td></td>
<td>2.3282</td>
<td>2.5076</td>
<td>-31.04</td>
<td>-31.04</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td></td>
<td>0.3061</td>
<td>0.7071</td>
<td>4.29</td>
<td>4.00</td>
</tr>
<tr>
<td>Overall Mean</td>
<td></td>
<td>3.238</td>
<td>2.294</td>
<td>-30.574</td>
<td>-33.932</td>
</tr>
<tr>
<td>Overall Std. Dev.</td>
<td></td>
<td>1.490</td>
<td>0.946</td>
<td>3.841</td>
<td>5.380</td>
</tr>
</tbody>
</table>

Table 2. Summary of two-tailed t-tests to compare mean differences of color versus gray-scale patterns with respect to expected mean difference of zero.

<table>
<thead>
<tr>
<th>t-Test - Color vs. Gray</th>
<th>t-value</th>
<th>t(0.025,4)</th>
<th>p[t &gt; t(crit)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>3.2048</td>
<td>2.7764</td>
<td>0.03</td>
</tr>
<tr>
<td>Green</td>
<td>2.4203</td>
<td>2.7764</td>
<td>0.07</td>
</tr>
<tr>
<td>Blue</td>
<td>-0.4349</td>
<td>2.7764</td>
<td>0.67</td>
</tr>
</tbody>
</table>
3. OBJECTIVE METRICS

3.1 Flats Detector with Local Luminance Adaptation Threshold

Fenimore, et al.,[2] describe a blocking impairment metric. This metric was designed to detect blocking structures in which 8 \times 8 image blocks contain certain homogeneities in pixel values, while contrasting with neighboring 8 \times 8 blocks by greater than some threshold value. The homogeneities involve an 8 \times 8 block having constant pixel values in either x or y directions or in both directions. Thus, a "flat" is an 8 \times 8 block of pixels having constant luma, \( Y'_0 \), inside the block cornered at image coordinate \((J,K)\), with pixels indexed by \((j,k)\), \(j - J = 0 \ldots 7\) and \(k - K = 0 \ldots 7\), for which

\[
Y'_{j,k} = Y'_0,
\]

on the block, and

\[
|Y'_{j,k} - Y'_0| > T,
\]  

at one or more of the adjacent points, \((j,k)\), where \((j - J = -1 \text{ or } 8\), and \(k - K = 0 \ldots 7\)) or \((k - K = -1 \text{ or } 8\), and \(j - J = 0 \ldots 7\)). In the study cited, \(T\) was set according to the image luma contrast at the blocking detection threshold.

In Fenimore, et al., the contrast criterion applied only to pixel contrasts across the block boundaries. The present study adds the condition that the Michelson contrast of a block classified as a flat also must exceed a visibility threshold. Thus, an 8 \times 8 image block was classified as a flat according to the original definition using a liberal threshold, \(T > 0\). But it is included in the impairment measure only if its mean luminance differs by greater than a threshold value from its surroundings. For this test, a contrast ratio is computed using the mean luminance value of the 8 \times 8 block under examination and the means of its surrounding neighbors which share a boundary according to the following expression:

\[
Y'_{j,k} = Y'_0,
\]

\[
|Y'_{j,k} - Y'_0| > T,
\]  

at one or more of the adjacent points, \((j,k)\), where \((j - J = -1 \text{ or } 8\), and \(k - K = 0 \ldots 7\)) or \((k - K = -1 \text{ or } 8\), and \(j - J = 0 \ldots 7\)). In the study cited, \(T\) was set according to the image luma contrast at the blocking detection threshold.
\[ C_{R_{j,k}} = \frac{\text{Max}(|\bar{Y}_{j,K} - \bar{Y}_{j-8,K}|, |\bar{Y}_{j,K} - \bar{Y}_{j+8,K}|, |\bar{Y}_{j,K-8} - \bar{Y}_{j,K+8}|)}{\bar{Y}_B}, \]  

where \( \bar{Y}_{j,K} \) designates average luminance of the 8 x 8 image block cornered at pixel \((j,K)\) and \( \bar{Y}_B \) designates the average luminance of the 24 x 24 enclosing block. The value, \( C_{R_{j,k}} \), was compared to a visibility threshold determined via the subjective measurements. The value 0.03 was used in the execution of this impairment metric. This threshold value is approximately equal to the average value obtained in subjective tests for the color and gray-scale test patterns.

### 3.2 DCT Perceptual Error Metric

The basis for JPEG and MPEG compression schemes is the DCT and the quantization of its coefficients. Typically, images are divided into 8 x 8-pixel blocks, each of which is transformed into a set of 64 coefficients representing the weights of the DCT basis functions from which the image can be reconstructed during the decoding process. The DCT transform coefficients, \( y_{u,v} \), of an \( N \times N \) block of pixels, \( Y_{u,v} \), are given by:

\[ y_{u,v} = \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} Y_{j,k} c_{j,u} c_{k,v}, \quad u, v = 0, \ldots, N - 1, \]  

where

\[ c_{j,u} = \begin{cases} \frac{1}{\sqrt{N}}, & u = 0, \\ \frac{2}{\sqrt{N}} \cos \left( \frac{\pi u}{2N} \right), & u > 0. \end{cases} \]

Compression of the image results from quantization of the DCT coefficients. Thus, we can designate the blocked DCT of the image as \( Y_{u,v,b} \), where \( u, v \) are the indices of the DCT frequency, which range from 0 to 7, and \( b \) is the block index. Each DCT block is quantized by dividing its coefficients by the corresponding elements of a quantization matrix, \( q_{u,v} \). Therefore, the quantized DCT coefficients are given by

\[ k_{u,v,b} = \text{Round} \left( \frac{y_{u,v,b}}{q_{u,v}} \right). \]

The quantization error, then, is expressed as:

\[ e_{u,v,b} = y_{u,v,b} - k_{u,v,b} \cdot q_{u,v}. \]  

Several investigators [8, 9, 11, 12] report methods by which to model the visibility thresholds for this quantization error. Peterson et al., [8] report experimental measurements of the detection thresholds for individual DCT basis functions in each of the R, G, B dimensions of a color display. Using their own data and that reported by van Nes and Bouman [10], Ahumada and Peterson [9] determine a best-fitting function which defines the visibility threshold of the DCT quantization error, \( T_{u,v} \), as a function of spatial frequency and the attributes of the display device. The parametric dependence of the Ahumada and Peterson function is given as:

\[ T_{u,v} = f_{\text{ap}}(u, v, L, p_x, p_y) \]

where

\[ u, v = \text{frequencies of DCT}, \quad u = 0, \ldots, 7, \quad v = 0, \ldots, 7, \]

\( L \) = luminance of display, including

\[ L_0, \] the veiling luminance on the screen and

\( L_Y, \) luminance contributed by the image,
As is pointed out by Watson [12], the matrix $T_{u,v}$ incorporates only a general dependence upon the mean luminance of the display, making no correction for local image-dependent variation in mean luminance. He proposes that the threshold matrix, $T_{u,v}$, can be adjusted according to local luminance values for each image block using the power law function:

$$T_{u,v} = T_{u,v} \left( \frac{Y_{0,0,b}}{\overline{Y}_{0,0}} \right)^{a_t}$$  \hspace{1cm} (10)$$

where

$$Y_{0,0,b} = \text{the dc value of the } 8 \times 8 \text{ image block, } b$$

$$\overline{Y}_{0,0} = \text{the average dc term of the image}$$

$$a_t = \text{luminance masking exponent given by Ahumada and Peterson, taking the value 0.649.}$$

Watson proposes further that the visibility model incorporate an adjustment for contrast masking such that the perceptual error matrix for block, $b$, is:

$$m_{u,v,b} = \operatorname{Max} \left( T_{u,v,b}, \frac{|Y_{u,v,b}|^{w_{u,v}}}{T_{u,v,b}^{1-w_{u,v}}} \right)$$ \hspace{1cm} (11)$$

where $w_{u,v}$ is an exponent for each frequency, ranging from 0, no masking, to 1. We use the value 0.0 for the dc term and 0.7 for all other DCT frequencies, as reported in Watson [12]. Thus, it is possible to compute perceptually just-noticeable-differences (jnd's) of the processed image as:

$$jnd_{u,v,b} = e_{u,v,b} / m_{u,v,b}$$ \hspace{1cm} (12)$$

While more elaborate methods have been described for spatial error pooling [14, 15, 16], in the present experiment we merely formed the linear sum of the absolute values of the jnd components in each block found to be greater than or equal to one. Thus, a block was considered to display a visible error if it was found to differ from the source image by at least one jnd. In order to facilitate comparison with the flats metric, a normalized value was computed as the percentage of image blocks having visible errors.

3.3 Results of Objective Metrics

Each of the objective measures was applied to the luma ($Y$) channel of frames 30 - 59 of the 60 frame sequence. Prior to processing, the luma values were converted to display luminance, as described previously, in order to facilitate comparison with visual threshold measurements.

Only a single impairment value was calculated for each frame, i.e., the percentage of $8 \times 8$ blocks exhibiting perceptible error. No attempt was made to differentiate among the three segments of the test pattern. Thus, comparison of objective threshold with the subjective values is made with respect to the grand mean and standard deviation of the six tests. Noting that the threshold results for the gray-scale version of the green pattern may have been unduly influenced by outliers, the mean value of -29dB is yielded from the other five tests. If the gray-scale version of the green pattern is included, the grand mean is -31dB. Thus, to the extent that the computational metrics are able to model the subjective blocking detection, the objective metric values would be expected to reach zero in the neighborhood of these average visual threshold values.

Figure 3 exhibits the objective metric values plotted against the maximum luminance contrast of each frame. Both metrics find the visual threshold at around -29dB, in good agreement with the subjective threshold for blocking detection. Moreover, the trends of the two metrics correlate well over the entire sequence, with Pearson $R = 0.988$. In general, the metric outputs correlate highly with the luminance contrast values for test pattern frames. Pearson $R$ correlation coefficients are summarized in Table 2.
Table 2. Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>$C_M$</th>
<th>DCT error</th>
<th>Flats</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_M$</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCT error</td>
<td>0.943</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Flats</td>
<td>0.970</td>
<td>0.988</td>
<td>1</td>
</tr>
</tbody>
</table>

4. DISCUSSION OF RESULTS

The present experiment finds a small, but statistically significant, difference in blocking sensitivity measurements for color versus gray images. The actual differences are small, however. Moreover, the small sample size used in the experiment may not support a firm conclusion as to the significance of these differences. The present values of 3% and 2% are consistent with the range of values reported by others for luminance contrast sensitivity. Levi and Klein [17], for example, give a range of 0.02 – 0.04 as typical measures of the Weber constant, $\Delta L/l$. In that the NIST spinning wheel test pattern is composed of low spatial frequency regions having relatively uniform luminance value, it is quite reasonable that block detection should be mainly a function of luminance adaptation following the Weber relationship. Moreover, the objective metrics yield a relatively good prediction of the visual threshold for blocking using only the luminance channel of the encoded video, adjusted to the luminance outputs of the display color guns.

For the detection of blocking, the flats detector appears to perform as well as the more sophisticated DCT error metric with respect to visual threshold prediction. It is computationally simple, and does not require a reference image sequence. The limitation of the flats detector, however, is that it is triggered by a relatively restricted range of local pixel value configurations. Moreover, its triggering conditions are best expressed within the original boundaries the $8 \times 8$ discrete cosine transform. Accordingly, offsets of flat features from their original positions due to other features of MPEG-2 compression, such as motion estimation, might diminish the performance of the flats detector for B and P frames of the sequence. The DCT error metric, too, operates on the $8 \times 8$ block. However, because it involves a frequency decomposition, it could potentially be used to detect a range of image impairments in addition to blocking.
In spite of greater computational complexity over that of the flats detector, the DCT error metric remains a comparatively efficient vision model. Extracting its frequency channel information from the DCT, a fundamental component of MPEG-2 compression, this image fidelity metric efficiently captures the quantization error characteristic of DCT encoding schemes. Its potential application directly to the encoded DCT may realize additional computational efficiencies for some applications. Thus, compared to vision models based on more computationally expensive image transforms, e.g., those using wavelets, DCT-based error models may prove quite economical and effective for fidelity measurement of MPEG-2 compressed video and other DCT-encoded images.

ACKNOWLEDGEMENTS

We wish to acknowledge support of this research by the Advanced Technology Program of National Institute of Standards and Technology.

REFERENCES


* Several commercial products are mentioned in this paper in order to specify the experimental setup. Such mention does not constitute an endorsement of the product, nor imply that the equipment is the best available for the purpose described.