Assessment of resolution and dynamic range for digital cinema

C. P. Fenimore and A. I. Nikolaev*
National Institute of Standards and Technology
Gaithersburg, Maryland 20899-8951

ABSTRACT

The proponents of digital cinema seek picture quality exceeding that of the best film-based presentation. Quantifying the performance of systems for the presentation of high quality imagery presents several challenges. For example, the dynamic range and the resolution may not be simply related to the nominal characteristics of bit-depth and pixel counts. We review some of the measurement methods that have been applied to determining these characteristics. One of the presumed advantages of high bit depth systems is to reduce the visibility of image banding. Non-uniformity of the display can be compensated in test pattern design to enable the measurement of banding contrast. The subjective assessment of banding is compared to a contrast-weighted model of noticeable image differences. Applied to a class of image banding test patterns, the metric relates dynamic range to contouring. The model produces an estimate of the visibility threshold for image contouring in a 10-bit system, superior to a simple Weber model. These measurement issues will continue to be challenges as d-cinema systems improve.

Keywords: resolution, contrast, banding, display measurement, subjective assessment, just noticeable difference, digital cinema.

1. INTRODUCTION

Digital or electronic cinema commonly is said to require picture quality exceeding that of the best film-based presentation. One aspect of the effort to realize this goal is the variety of proposed formats for digital cinema. These include a range of bit-depths from 10 to 16 bits with either linear, power law, or logarithmic scaling of code values to luminance and color, while pixel addressability has been promoted far beyond 1920 x 1080 [2]. The presumed advantages over lower formats is that these enable a wider dynamic range and higher resolution, and ultimately a greater range of grays and colors with improved picture sharpness.

How are we to judge whether or not the "enabled" features implied by high formats, are realized in digital cinema presentation systems? Specifically, how does one assess the resolution and dynamic range of a d-cinema system? The answer potentially involves two classes of measurements. The first class involves physical measurements of luminance contrast from which we determine dynamic range and resolution. While these measurements are essential, they do not provide a direct measure of the visual characteristics of the system. To do so, we consider methods for the subjective appraisal of dynamic range as it related to image banding.

Within the last decade Kelley, Boynton, and others have exposed the display community to an understanding of the effects of veiling glare and the methods for reducing them. This has led to more reliable engineering measurements of contrast and other critical characteristics of information displays [3]. We highlight some of the issues associated with the measurement of local contrast. The contrast modulation function is appealing for its robustness. The CMF is based on measurements of grille patterns for which small area measurements relate directly to the theory.

One of the presumed advantages of high bit depth systems is the elimination of visible banding, such as that seen over large scale regions with smooth gradients. Relating visible banding to the measured luminance at various points on the screen is complicated by the spatial nonuniformities of the display. We describe a family of test images and associated measurements which allow us to estimate the dynamic range and to determine banding visibility with confidence that the contrast is properly measured. The candidate patterns are designed to facilitate such measurements by orienting the low contrast patterns to avoid confounding the controlled variation in the pattern with unknown local variation in the display

* Convergent Information Systems Division, NIST, Technology Administration, U. S Department of Commerce.

luminance. Using a human vision model for picture comparison, the NIST Just Noticeable Difference metric (DCT-JND) [1], the measured banding is compared to the computed analysis of the JND-DCT metric. Using informal viewer assessment the JND-DCT performs well compared to a simple Weber threshold model for banding detection.

2. NIST IMAGE QUALITY LAB AND SIGNAL FORMATS

Most of the measurements reported here, were made in the NIST Image Quality Lab. The Lab has a video projection system, which maintains 10-bits from the stored signal through the projector. For serving digital video, we use an Accom* uncompressed high definition disk array, which stores 10-bit YUV (Abekas) images and outputs digital signals as either standard definition or high definition serial digital interconnect (SDI or HD-SDI) associated with the family of HD television standards related to SMPTE Recommendations 292M and 259M. For our projector we use a Christie Vista S3 with serial digital interfaces and internal processing. The projector is a deformable mirror device (DMD) having 3 chips [4]. The light measurements reported in this paper are made with a Minolta L-100 Luminance meter.

Signal format and measured system gamma
The high definition test patterns used in this work are generated as ITU Rec. 709 digital data. We designate the three coordinates by Y', C_b', and C_r', where the prime (') distinguishes code values from luminance and color-related measured quantities. In Rec. 709, the range of luminance code values, Y', is [64, 940]. The present work does not involve color measurement and so
$C_b' = C_r' = 512$.

All measurements are made with the projector setting of contrast, C, and the brightness, Br, close to the nominal set points recommended by the manufacturer. These values are C = 39.4 and Br = 57.9. They were chosen to maximize the dynamic range of the system, as determined by viewing low contrast patterns.

3. LUMINANCE-BASED OBJECTIVE MEASUREMENTS

Measured system gamma and dynamic range
The system gamma (γ) relates code values to luminance levels. As described in [3], the model,

$$Y = a\,(Y')^{g} + Y_b\,, \qquad (1)$$

is fitted to the measured luminance for 15 full-screen level gray images similar to that in Figure 1. The gray-scale code values are spaced evenly over their full range. For our system, $\gamma = 2.49$.

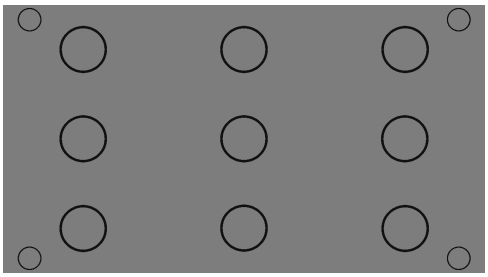


Figure 1: The measurement pattern for determining screen luminance. In determining system γ, the measurement is made at the center point. The same pattern is used in determining screen uniformity and finding local gradients in the screen luminance.

The dynamic range, *R*, is the number of code values with distinct luminance produced by the image presentation system; that is, ***R* = #(*Y* levels)**. To avoid making an exhaustive set of measurements, one may estimate *R* by measuring *Y* for a sample of code values. .

We measure the center-screen luminance levels for several consecutive code values, initially in 3 ranges: low, intermediate, and high. In an ideal system, the sampled values would follow the system γ-curve, within the system measurement error. As will be seen below, it may happen that the measured luminance fails to change between successive code values. This may result fromimproper system setup or may be an unavoidable characteristic of the system. The sampled number of distinct luminance levels provides the estimate of dynamic range.

In the present discussion, we find this number of distinct levels varies from one sampling interval to the next. In this event we add another sampling interval between the two for which a change is observed. Repeated applications of such subdivisions, yields a sequence of sample intervals. Within some intervals there is local change in the dynamic response. Other intervals exhibit locally constant dynamic response.

For the NIST system the 3 initial ranges are found in Table 1 with Code values of 64 – 68, 502-506, and 936 – 940, corresponding to low-, middle-, and high-level in the ITU-R BT Rec. 709 code value range. In the high and middle intervals the 5 distinct code values yield 5 distinct measured values. However, for the low interval, [64 – 68], the 5 code values (64, 65, 66, 67, 68) correspond to only 2 distinct luminance values with a single large change on the interval [67, 68]. Table 1 exhibits the added sampling intervals which refine the sampled data. The estimated dynamic range is 833 levels, compared to a nominal dynamic range of 876.

CODE VALUE +	OFFSET 0	1	2	3	4	Standard Error	Width of steps	Number of levels
64	0.563	0.564	0.566	0.565	0.592	0.001	4	3
76	0.644	0.645	0.649	0.669	0.668	0.0012	3	7
98	0.819	0.821	0.844	0.847	0.869	0.002	2	8
114	1.031	1.032	1.053	1.073	1.076	0.002	1.5	21
146	1.657	1.660	1.710	1.738	1.764	0.002	1	356
502	37.20	37.44	37.65	37.76	38.0	0.02	1	438
936	153.4	153.6	153.9	154.5	154.7	0.3	1	833/876

Table 1: Estimation of dynamic range from luminance measurements based on sampled luminance values. The luminance is in units of candela/m^2. The standard error is dominated by the instrumental error.

Measurement of resolution by contrast modulation

Resolution can be derived from the modulation of the luminance signal found in sinusoidal or square wave luminance (and color) patterns [3]. The family of grille patterns has a square wave profile with alternating black and white bars (Figure 2 [5]) in both vertical and horizontal orientation. The grille patterns are two-level images with a fixed code value for each stripe. For luminance measurements, the spot size of the luminance meter may critically affect the results [3].

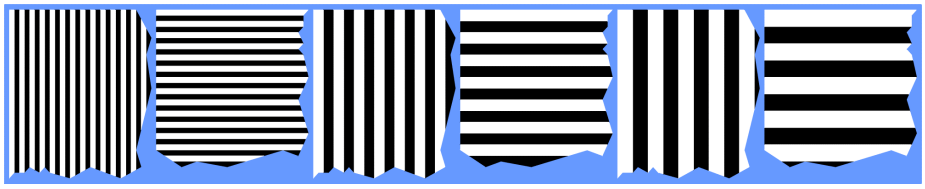


Figure 2: The grille pattern for contrast modulation measurement consists of alternating black and white stripes. The results indicated are taken from the **N-ON/N-OFF** horizontal and vertical grill patterns (1×1, 2×2, 3×3, 4×4, 5×5) used for contrast modulation and resolution measurements. Six of the ten grill patterns are shown in schematic form. Each demonstration patch represents a full screen test pattern.

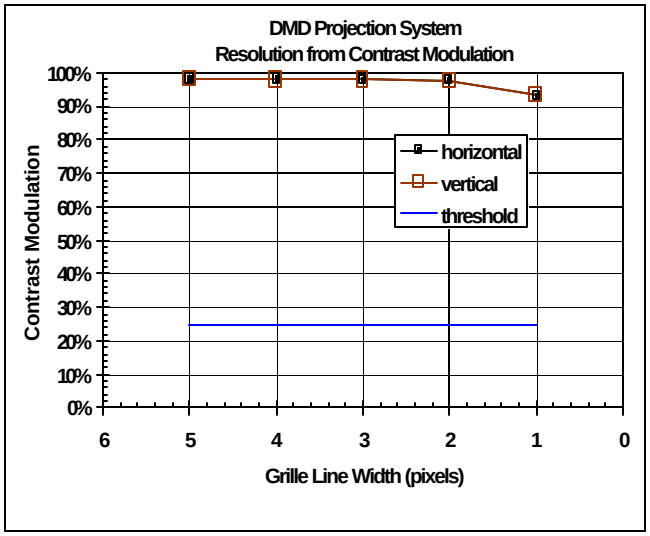


Figure 3: In a set of measurements to characterize a DMD projection display [5], the nominal resolution is identical with the resolution from Contrast Modulation. Resolution is determined by the smallest grille width for which the contrast modulation lies above the indicated threshold (22.5%).

4. PATTERNS FOR VISIBLE CONTOURING

Image banding and contouring

Banding can be seen in natural and synthetic imagery, particularly in low bit depth imagery. In Figure 4A, the painting, *Fruit Packers* [2], the gray scale image is captured at eight bits per sample, while Figure 4B has three bits per sample. The effects of banding are evident in the latter image, due to the reduction of dynamic range associated with the reduction of code values from 256 to 8. The banding shown here is intended to be illustrative, rather than characteristic, of digital cinema.

Figure 4A: *The Fruit Packers* [5] at 8 bits per sample. The original oil painting is captured with a color digital camera and reduced to grey scale.

Figure 4B: *The Fruit Packers* at 3 bits per sample. Banding is visible in those image areas with gradual change in gray scale, for example in the clothing.

For many display technologies, the 10-bit code values of RGB TIFF images or of Rec. 709 appear to approach the threshold for visible banding suggested by Weber's Law, which asserts the visibility of image contrast is determined by the relative change in luminance,

$$W = \Delta (Y) / Y \qquad (4)$$

The value of the visibility threshold depends on the apparent size of the banding elements and on viewing conditions. In the present case our first concern is the challenge of measuring the local contrast for projection systems. The test patterns used for measuring banding are designed to compensate for display non-uniformities. We have conducted an informal evaluation of visibility for the test patterns and consider the performance of the Weber model and another NIST-developed model for predicting visibility thresholds.

Screen non-uniformity.

The measurement of image banding is complicated by display spatial non-uniformity. The percent non-uniformity of a display is defined [3, p115] by

$$\text{Nonuniformity} = 100\% * (Y_{max} - Y_{min}) / Y_{max} \qquad (5)$$

where the luminance, Y, is measured at the 9 center points indicated in Figure 5 by the large circles. Y_{max} and Y_{min} are the maximum and minimum luminance values at the 9 points. The circles are for alignment and are not shown when measurements are made. For the NIST projector, the screen non-uniformity was measured at 3 luminance values (low, middle, and high). The nonuniformity ranges from 20 – 26 %, increasing with display brightness.

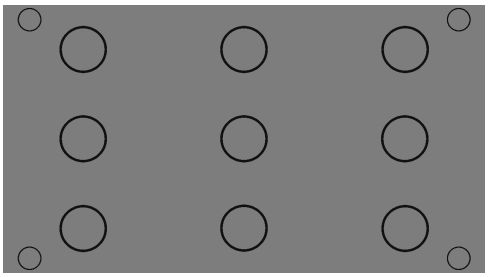


Figure 5: The measurement pattern for determining screen luminance non-uniformity, and the mean luminance over the screen. The non-uniformity of the projection system in the NIST Image Quality Lab, under the conditions described in Section 3, ranges from 20% - 26%. It is identical with Figure 1.

For low contrast patterns of horizontal or vertical stripes, loosely patterned on Figure 6, we find the apparent change in luminance between 2 stripes agrees with the change in code value. Thus, stripes with increasing code values appear brighter. The visible change in brightness is determined by the local change across edges. However, the spatial non-uniformity of the luminance makes reversals of the measured luminance rather common for low contrast edges, thus a stripe edge with increasing code value may have measured luminance at the center of the stripe that is lower.

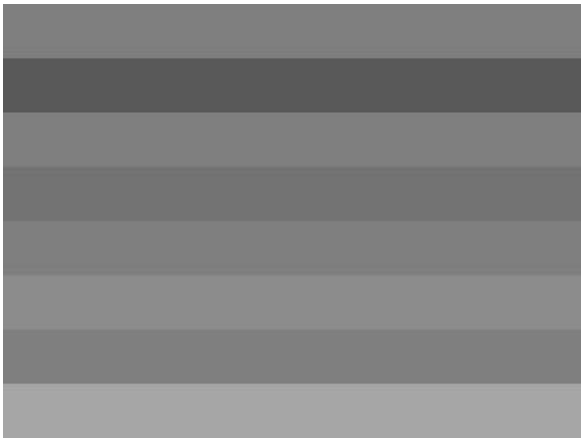


Figure 6: Screen nonuniformities can confound the measurement of contrast between 2 stripes. For low contrast patterns, the measured luminance may change in a sense opposite that of the apparent brightness between two stripes.

Test patterns for the measurement of image banding.

The pattern seen in Figure 7 permits the reliable measurement of banding. The utility of the pattern depends on the placement and orientation of the bar. The bar is centered in a region where the gradient of the luminance is smallest and oriented in the direction of the gradient, thus the luminance difference is measured in a direction, $\boldsymbol{n}$, normal to the gradient. Thus, when the luminance of a frame varies, despite having a constant code value, one determines its derivatives:

$$\tilde{\mathrm{N}}\, \boldsymbol{Y} = (\partial\, \boldsymbol{Y} / \partial\, \boldsymbol{x}\,, \partial\, \boldsymbol{Y} / \partial\, \boldsymbol{y})$$

So long as the luminance intensity for a flat-luminance frame is described by a differentiable function of position, the change due to the nonuniformity is zero in the direction of the normal because:

$$\tilde{N}\, Y_{\text{constant } Y'} \cdot \; n = 0$$

To analyze the effectiveness of this pattern, consider the sources of variation in the luminance of the banding pattern. Decompose the luminance into 2 components:

$$Y = Y_N + Y_C$$

Y_N is the variation due to the nonuniformity, while Y_C is the spatial pattern due to changing code value. As noted above, the non-uniformity is zero in the direction normal to the gradient. For the pattern in Figure 7, which is aligned with the gradient, the luminance change across the interior edge is entirely due to the change in code value, and has no component from the nonuniformity.

$$\tilde{N}\, Y \cdot \; n = \tilde{N}\, Y_N \cdot \; n + \tilde{N}\, Y_C \cdot \; n = \tilde{N}\, Y_C \cdot \; n.$$

Figure 7: The Banding Visibility Test Pattern is designed to eliminate the confounding effect of display nonuniformity on local contrast measurements.

The gradient is estimated from measurements on the grids of Figure 5 and Figure 8. The nine principal measurements in Figure 5 define 4 quadrants. Determine a "most nearly level" quadrant with the smallest gradient. Within that quadrant use the refined grid of Figure 8 to determine the gradient at the center of a "most level" region. For the NIST projector we locate the stripe pattern as shown in Figure 7.

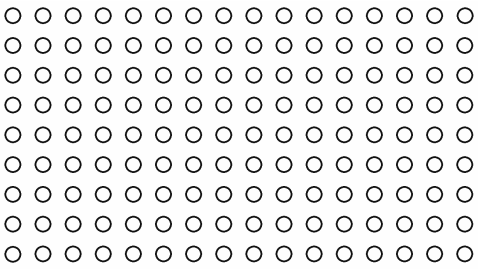
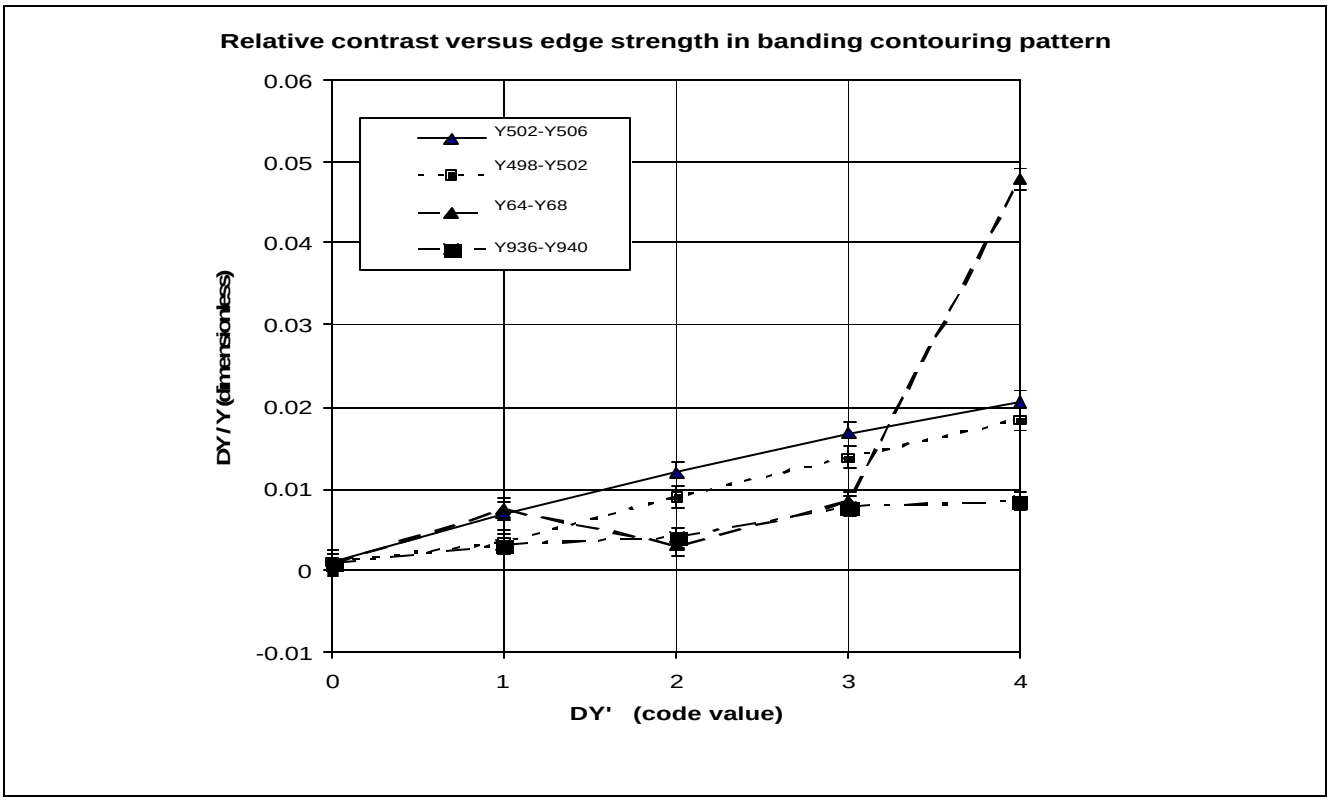


Figure 8: The sampling pattern for orienting the banding test pattern. By staging the measurements as shown first on Figure 5, followed by Figure 8, one needs measure only in one quarter of the larger mesh.

Figure 9: The edge strength is measured by averaging 3 luminance values on either side of the edge.

5. MEASUREMENTS FOR VISIBLE CONTOURING

The Banding Test Pattern was used to produce the contrast measurements shown in Table 1. Each datum is determined by averaging three measurements. The measurement points are equally spaced on one side of the interior edge, as shown in Figure 9. In the data plot, Figure 10, the contrast is normalized by the luminance. The number of data points for mid-level code values are double those for the high and low intervals.



DCT Perceptual Error Metric

The NIST-JND metric carries out the analysis of the difference between two images using the discrete cosine transform (DCT). The DCT and the quantization of its coefficients are the basis for JPEG and MPEG compression schemes. 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, $I_{m,n}$, of an N x N block of pixels, $i_{j,k}$, are given by

$$I_{m,n} = \sum_{j=0}^{N-1} \sum_{i=0}^{N-1} i_{j,k}\ c_{j,m}\ c_{k,n}\ , \quad m,n = 0,\ldots,\ N-1, \tag{6}$$

where

$$c_{j,m} = \sqrt{\frac{1}{N}}\ , \quad m = 0,$$

$$= \sqrt{\frac{2}{N}}\ \cos\left(\frac{p\ m}{2N}\left[2j+1\right]\right), \quad m > 0\ .$$

One technique in the compression of an image is the quantization of the DCT coefficients. Thus, we can designate the blocked DCT of the image as $I_{u,v,b}$, where u and 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, $q_{u,v}$. The quantized DCT coefficients are given by

$$k_{u,v,b} = Round\left(I_{u,v,b} / q_{u,v,b}\right) \tag{7}$$

In the present application we compare 2 images ***i*** and ***j*** which have transformed representations $I_{u,v,b}$ and $J_{u,v,b}$. The quantization error, then, is expressed as

$$e_{u,v,b} = I_{u,v,b}\ -\ J_{u,v,b} \tag{8}$$

As described in [1] and developed by Watson et al. [7], the visibility thresholds use the visibility of DCT basis functions and adjust the contrast sensitivity function for the display characteristics. The DCT error thresholds are also adjusted to account for image-dependent effects of luminance adaptation and contrast masking. Designating the perceptual error threshold matrix as $m_{u,v,b}$, it is possible to compute perceptually just-noticeable-differences (jnd's) of the error image as

$$jnd_{u,v,b} = e_{u,v,b}\ /\ m_{u,v,b.} \tag{9}$$

In the present experiment the jnd's for each DCT block were pooled using Minkowski summation to yield a perceptual error value for each 8 x 8 image block according to

$$p_b = \left(\sum_u \sum_v \left| jnd_{u,v,b} \right|^{\ b}\right)^{\frac{1}{b}}\ , \tag{10}$$

where the value for b was set to 1 in the present investigation. The JND measure for the images is the maximum of the values of p_b.

DISCUSSION AND CONCLUSIONS

The 2 input images used to compute the DCT-JND estimate of the visibility of the edge are (1) the test pattern with the interior edge assigned various intensities and (2) the similar pattern with an edge of zero strength. Any visible differences between the two images is attributable to the visibility of the edge in the first image. This allows us to use the dual input measure of image differences to model the visibility of the edge in a single image.

The simulation of the visible edge strength has been compared to the results of an informal appraisal of visibility. The data suggest the DCT-JND is superior to the simple Weber threshold model (4). In particular, for a limited sample size values of the DCT-JND above 2 units appear to correlate well with detection of contouring in the informal evaluation.

Although our system did not achieve full dynamic range, we are within a few percent of the nominal range. Using the sampling method for detecting measurable and visible banding, it appears that for the bright end of the code value scale, banding may be visible for edges having only one unit differences. If confirmed in formal tests, it would suggest a need for the use of systems with 10 or more bits per color sample for presentation as well as post-production of digital cinema entertainment.

Another intriguing aspect for future investigation is the power of the JND-DCT to predict visible differences in very high quality systems. The JND tools were developed in an 8-bit environment. Predictive power for 10-bit or higher systems would be of great interest to the imagery quality community. As nominal resolutions and bit depths increase there will be a continuing need to validate the effective resolution and dynamic range and to validate our models.

REFERENCES

1. J. Libert and C. Fenimore, 'Visibility thresholds for compression-induced image blocking: measurements and models', *Human Vision and Electronic Imaging IV*, SPIE Vol 3644, San Jose CA, pp 197-206, 1999.

2. Robert Rast, "A Report of the SMPTE Digital Cinema Study Group", *Digital Cinema 2001 Conference Proceedings*, NISTIR 6591, Gaithersburg, MD, 2001.

3. VESA, *Flat Panel Display Measurement Standard, Version 2.0*, 2001.

4. Paul Breedlove, "The DLP Projector", *Digital Cinema 2001 Conference Proceedings*, NISTIR 6591, Gaithersburg, MD 2001.

5. P.A. Boynton and C. Fenimore, *NISTIR 6792*, "Characterization of Projection Systems for the MPEG-4 Digital Cinema Compression Scheme Evaluation," Gaithersburg MD, [September 2001].

5. Tom N. Cornsweet, *Visual Perception*, Orlando, Florida, 1970.

6. Edward Fenimore, *The Fruit Packers*, oil on canvas, 1978.

7. A. B. Watson, "Toward a perceptual distortion metric for digital color video," *Human Vision and Electronic Imaging III*, SPIE Vol 3299, San Jose CA, 1998.

---

* In this paper, specific references are made to commercial products in order to specify the conditions and methods of conducting the investigation. Such references are not an endorsement by NIST of any product, nor should they be taken to imply that the equipment is necessarily the best for the purposes of this study.