

**VISIBILITY OF SPATIO-TEMPORAL DITHER NOISE:
EFFECTS OF DISPLAY LUMINANCE, PITCH, AND FRAME RATE**

Dr. Charles J. Lloyd – Systems Engineering Manager
Mark A. Carter – Senior Staff Engineer
FlightSafety Visual Simulation
FlightSafety International

ABSTRACT

This paper describes a quantitative model of the effects of three practical display design variables on the visibility of random dither noise used to improve grayscale performance. The magnitude of noise, and thus the size of the grey scale step that can be tolerated, increases as expected with decreasing luminance and pitch and with increasing frame rate. The data are summarized in a multiple regression model that explains 95% of the variance in the data from seven observers. When this simple technique is employed in video images, gray scale steps as large as 3% can be used under any practical conditions. For typical desktop displays with pixel pitch in the range of 1.5 to 2 arcmin, the maximum grayscale step size increases to 5-7%. As luminance decreases below 1 fL the grayscale step size that can be tolerated increases rapidly such that below 0.01 fL a binary display performs as well as a display with 8-bit grayscale. For typical simulation training display systems installed today (2.5 arcmin pitch, 60 Hz, <20 fL) 60 gray scale steps would be sufficient if they could be optimally allocated. The 255 grayscale steps available in the simplest of modern digital interfaces provide ample head room for tolerating the sub-optimal distribution of grayscale steps provided by the standard “gamma” corrections used in IGs and projectors.

INTRODUCTION

Robert’s Method

A common method used to decrease the visibility of gray scale banding that may occur in imaging systems employing grayscale sampling involves the addition of random noise to the signal, prior to quantizing. This technique serves to greatly reduce or eliminate the false contours that can occur on objects with smoothly varying grayscale. While few papers address this technique with 8-bit video, it is believed the technique is in widespread use in digital video systems. In the mid 1980s the method was applied by FlightSafety in image generators for flight simulation. In 1990 software engineers at Honeywell used the technique for displaying video images on AMLCD cockpit displays. LCD projectors produced by several vendors have used this technique for many years. Up close examination of most any large flat-

panel LCD television display reveals that some form of spatio-temporal dithering is used, ostensibly to improve the control of grayscale.

The earliest paper we have found describing this technique is by Roberts in 1962, cited, well-described, and illustrated in Schreiber (1986). In Section 4.7.2, “Randomization of the Quantization Noise,” Schreiber provides a detailed description of how and why this technique works and provides examples of static images processed using the method. Schreiber summarizes the method on pp. 101 as follows:

“This method works so well, is so easy to implement, and has so few disadvantages, that it is hard to see why it is not universally used. All that is required is to add to the signal, before quantizing, a random noise of uniform amplitude probability distribution and peak-to-peak amplitude equal to one quantization step.”

A literature search performed today brings up hundreds of papers, entire conferences in fact, on dithering, digital halftoning, ordered dithering, error diffusion, blue noise, green noise, and related topics. Recent reviews of much of this work are provided by Mese and Vaidyanathan 2002, and Ulichney, R. 2000. The great bulk of this literature applies to hardcopy (or otherwise very low bit depth) display devices or to image compression techniques. To date we have found few papers describing or evaluating Robert’s method as it applies to video systems with 8 or more bits of grayscale. Perhaps the reason for the relative dearth of research in this area is that Robert’s method is simple, effective, inexpensive, and well-described some 45 years ago.

The point of this work was NOT to evaluate methods of dithering or to promote Robert’s method above others. Quite the contrary. For this evaluation we selected the earliest, simplest, and least computationally expensive dithering technique we could find. We fully acknowledge that better dithering algorithms exist. By selecting a simple and inexpensive algorithm, the consumer of the model presented in this paper can be more confident that their

IMAGE 2009 Conference

favorite dithering algorithm will produce a reduction in their bit depth requirement that is *at least as good* as predicted here. Had we used the best performing algorithm we could find we would expect our model to have less general utility.

Scope

The conclusions drawn in this paper generally apply to the real-time generation, transmission, and display of video images. Robert's method can be applied effectively within real time image generators just before or as part of the process of quantizing the grayscale. We warn the reader to avoid generalizing these findings to other applications where they may not apply. For example, we are not promoting this method for use in de-contouring images after the gray scale sampling has been completed. Similarly, we think it would be inappropriate to use the method in video systems where MPEG or other efficient video compression techniques are needed as the addition pixel-level random noise would significantly reduce coding efficiency and/or effectiveness.

High Bit Depth Video

In recent years various parties have lobbied for the use of bit depths greater than 8 bits/pix for video display systems. In 2006 engineers developing a very high contrast (e.g., CR > 200,000) projector recommended the use of a high bit-depth image transmission scheme explaining that at least 10 bits/pix would be needed to effectively use the high contrast range of their product. Similarly, in 2007 engineers at an AFRL laboratory suggested that 16 to 20 bits of grayscale are required to generate and transmit video images that span the visible light and the near infra-red levels required for stimulating night vision goggles. Reinhard et. al., (2006) and other researchers in the high dynamic range imaging arena have argued that a bit depth higher than 8 bits is needed to transmit images that cover a grayscale range comparable with the real world. A common argument is that published psychophysical data indicate humans can see gray scale steps as small as 0.5 to 1% (see lower curve in Figure 7), thus 12 or more bits of gray scale are needed for very high contrast displays.

The general assertion that bit depths greater than 8 bits are needed to transmit high dynamic range video is questioned as this conclusion is strongly mediated by four important dimensions of human visual performance:

1. Threshold gray step size increases as luminance is reduced for both dithered and non-dithered images.
2. Spatial and temporal dithering significantly reduce gray step visibility.

3. Spatial dithering effectiveness increases with spatial resolution.
4. Temporal dithering effectiveness increases with frame rate.

Purpose

As explained by Roberts and others, the fundamental reason dithering techniques work is that they break up the gray scale bands and false contours that may have occurred due to limited bit depth and distribute these errors randomly across space and time. Adding the noise does not increase the total amount of noise in the image, rather, it effectively shifts the spatial and temporal frequency of the noise to levels too high to be detected by human observers.

This evaluation was designed to quantify the size of the threshold gray scale step that can be tolerated in the presence of spatial and temporal dithering. The evaluation was repeated for 72 combinations of luminance, spatial resolution (display pitch), and temporal resolution (frame rate) so that the effects of and interactions among these variables can be quantified.

Method

Equipment and Software

Images were displayed using a single chip DLP projector (InFocus, Model X3) illuminating a white screen. The zoom lens on the projector was set to the smallest image size and the projector was positioned at the nearest distance at which the lens would focus which was 1.6m (63 in) from the screen.

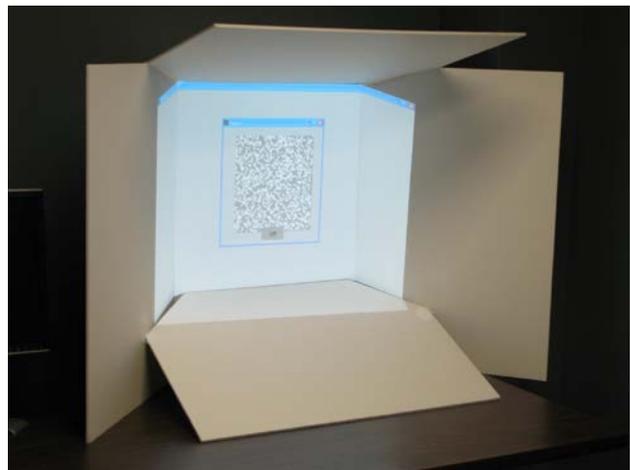


Figure 1. Screen assembly used in the evaluation. An extended surround was used to stabilize the adaptation level of the observer and to avoid high contrast edges near the visual stimulus.

Figure 1 shows the projected image on the screen assembly. The center portion of the screen measured 50 x 50 cm (20 x 20 in) and was positioned 3.9 m (154 in) from the observer and thus subtended 7.4 deg. One reason a long viewing distance was used is that visual acuity is generally maximized at viewing distances greater than a few meters (Luckiesh and Moss, 1941). The outer most extent of the screen measured 117 x 81 cm (46 x 32 in) and was set at a distance of 3.5 m (138 in) from the observer. The outer portion of the screen subtended 19 deg horizontally by 13 deg vertically. The noise pattern within the inner-most window measured 19.6 cm (7.7 in) which subtended 2.9 deg from the observer point of view. At this width the noise pattern was several times larger than the high-acuity foveal vision (about 1 deg wide) used by the observers for detecting the pixel level “salt and pepper” noise being produced.

The luminance of the projected image surrounding the noise image was always equal to the luminance of the noise image. The evaluation was conducted in a room with dark walls, thus, the contrast between the projected image and the walls was very high. The white panels surrounding the projected image were used to reduce this contrast and to stabilize the adaptive state of the observer at the image level.

The projector was set to “video” mode which produced a peak white that was less than half of the peak white of the “presentation” mode. It is presumed that the color wheel used in the illumination optics of this projector has a white segment that is activated in the presentation mode and deactivated in the video mode. The electro-optical response (i.e., “gamma curve”) of the projector was measured using a Minolta LS-100 luminance meter. The measured curve is plotted in Figure 2.

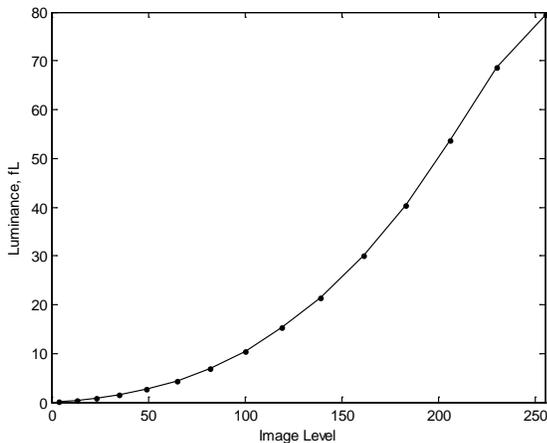


Figure 2. Electro-optical response for white of the InFocus projector when operated in “video” mode.

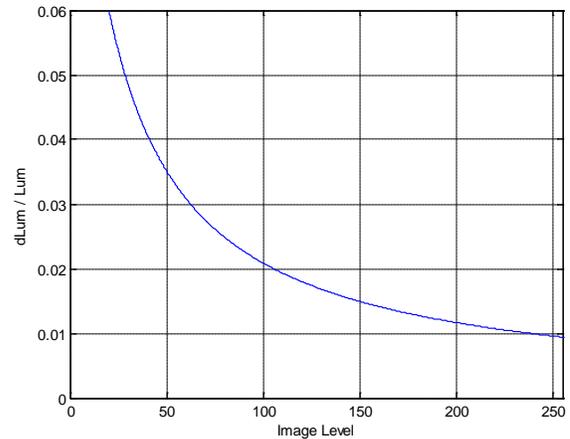


Figure 3. First derivative of the electro-optical response function showing the proportional change in luminance resulting from a unit change in image level. Note that the smallest luminance ratios (1%) occur at the highest image levels.

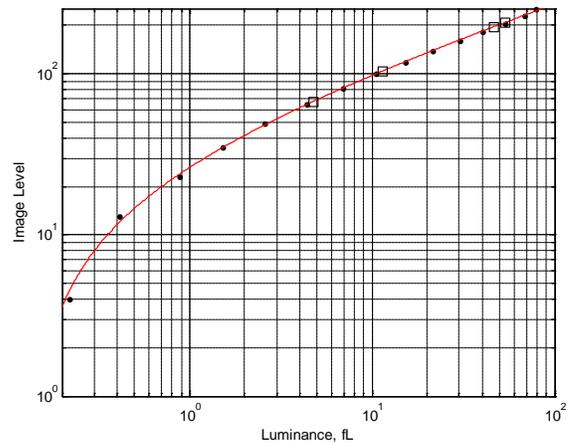


Figure 4. Inverse electro-optical response function fit to the data shown in Figure 2. Function was used to determine the levels used for the high and low portions of the noise images keeping the mean luminance constant across noise levels. The lowest two squares indicate the levels used in the darkest condition while the upper two squares indicate the levels of the brightest condition.

Dither Noise

The results of a preliminary evaluation indicated that dither noise visibility is maximized when the mean level of the

IMAGE 2009 Conference

image is at half way between the lower and upper states. In other words, thresholds are lowest when on average half the pixels are rounded up to the next available level and half are rounded down. Thus, the 50% point was used for this evaluation so that the resulting data represent the “worst case” viewing conditions where the observer is maximally sensitive to the noise.

Previously published evaluations of visual sensitivity show that sensitivity is maximized when the observer is adapted to the luminance level of the stimulus. Thus, the surround luminance was set equal to the mean luminance of the noise patterns.

Display Design Variables

The visibility threshold of the dithering noise was measured as a function of 72 combinations of three experimental variables: display luminance, pitch, and frame rate.

Luminance Six logarithmically spaced luminance levels were used in this evaluation. The upper level was selected to be high enough that noise visibility would asymptote. The lowest luminance was set with the goal of achieving a maximum step size of about 50%

For this evaluation the smallest gray scale step size (luminance ratio) achievable by the projector was desired, thus the projector was operated towards the high end of the luminance range where the step-to-step luminance ratios are the smallest (see Figure 3). This was accomplished by placing neutral density filters in front of the lens to reduce the luminance, rather than commanding the projector to low luminance levels. Three filter conditions were used in the evaluation, no filter, single filter, and double filter. For the white produced by this projector the luminance transmittance was 0.116 for the single and 0.0135 for the double filter conditions.

Table 1. Projector levels, ND filters, and luminance levels produced at the screen for the six luminance conditions.

Luminance, fL from Projector	Neutral Density Filter	Luminance, fL at Screen
50	None, T = 1	50
16.5	None, T = 1	16.5
47.4	Single, T = 0.116	5.5
15.5	Single, T = 0.116	1.8
37.0	Double, T = 0.0135	0.50
14.8	Double, T = 0.0135	0.20

Display Pitch In this evaluation the viewing distance was held constant so that changes in visual acuity that can occur with changes in distance did not confound the results. Display Pitch was changed at the display by changing the number of native projector pixels used to create each image pixel. For example, for the finest pitch condition, 2x2 projector pixels were used to create each image pixel. For each image pixel each of the native projector pixels were commanded to the same gray level.

At the native resolution of the projector and image magnification used in this evaluation the native pixel pitch was 0.65 mm/pix (.026 in/pix). At the 3.9 m viewing distance this produced 0.57 arcmin/pix. The number of native projector pixels used for each image pixel and the resulting image pitch are provided in Table 2 for each of the four pitch conditions.

Table 2. Number of projector pixels, image pitch, and number of pixels across the display for the four resolution conditions.

Number of native projector pixels per image pixel	Image Pitch, arcmin	Number of image pixels across 2.9 deg window
2 x 2	1.14	150
3 x 3	1.71	100
4 x 4	2.28	75
6 x 6	3.42	50

The high end of this range represents the pitches that have been sold into the flight simulation training market over the past few years. The small end of this range approaches the pitches that will be required to achieve the “eye limited” resolution that several military customers have described as their ultimate goal.

Frame Rate For all conditions in the evaluation the frame rate of the projector was fixed at 60 Hz. Three “effective” frame rates were produced by controlling the number of projector frames over which the new noise pattern was displayed. The effective frame rates were 60, 30, and 15 Hz. In the 30 Hz condition the same noise pattern was displayed for two consecutive frames while it was displayed for four consecutive frames in the 15 Hz condition.

Experimental Design

A full-factorial, within-observer, experimental design was used for this evaluation, meaning that all combinations (6 x 4 x 3 = 72) of each of the three experimental variables was evaluated by each observer. Each observer was presented

IMAGE 2009 Conference

with the conditions in a different random order so that any unavoidable noise or drift in observer ratings, such as those caused by practice and fatigue, would be distributed randomly throughout the data and would not bias the results.

Observers

Seven people (six male), all employees of FlightSafety International participated in this evaluation. The ages of the observers ranged between 24 and 50 years and the mean age was 31.4 years. All observers reported good distance vision. Data collection took approximately one hour each with instructions and practice trials requiring about ten additional minutes.

Task

For each experimental condition the threshold noise visibility was measured using the psychophysical “method of adjustment.” At the beginning of each experimental trial the magnitude of the noise was set above the threshold so that it was clearly visible. The noise magnitude was slowly reduced to the point where the observer could no longer see it at which point the observer pressed a “reverse” button on the keyboard which began slowly raising the noise magnitude. As soon as the observer could again see the noise they pressed the reverse button. This pattern of raising and lowering the threshold was repeated for about 40 sec or 8 to 10 transitions per experimental condition. The threshold is defined as the geometric mean of the upper and lower transition points.

Results and Discussion

Regressors

Using the stepwise multiple regression tool supplied with the MATLAB Statistics Toolbox, hundreds of candidate models were considered which contained combinations of the following regressors used to predict either threshold or $\log_{10}(\text{threshold})$:

- luminance, luminance², luminance³,
- $\log_{10}(\text{luminance})$, $(\log_{10}(\text{luminance}))^2$, $(\log_{10}(\text{luminance}))^3$
- $(1/\text{luminance})$, $(1/\text{luminance})^2$, $(1/\text{luminance})^3$,
- pitch, pitch²,
- $\log_{10}(\text{pitch})$, $(\log_{10}(\text{pitch}))^2$
- rate, rate²,
- $\log_{10}(\text{rate})$, $(\log_{10}(\text{rate}))^2$
- All two-way interactions amongst these regressors

The best fitting model was selected on the basis of maximizing the R^2 correlation (minimizing the RMSE) while requiring the fewest number of terms and using terms with the lowest powers and fewest interactions.

The model that was settled on was:

$$\begin{aligned}\log_{10}(\text{thresh}) = & b(1) + b(2)*\log_{10}(\text{lum}) \\ & + b(3)*(\log_{10}(\text{lum}))^2 + b(4)*\log_{10}(\text{pitch}) \\ & + b(5)*\log_{10}(\text{rate}) \\ & + b(6)*(\log_{10}(\text{lum}))^3*\log_{10}(\text{pitch}) \\ & + b(7)*\log_{10}(\text{lum})*\log_{10}(\text{rate})\end{aligned}$$

Where:

- thresh is the threshold of the mean observer, (proportional change in luminance)
- lum is the mean luminance, fL
- pitch is the display pitch, arcmin

The coefficients for this model are: $b =$

$[-1.054 \quad -0.2704 \quad 0.03011 \quad -0.9180 \quad 0.3084 \quad 0.04148 \quad -0.07224]$

A plot of the fitted model with the raw data showed no evidence of systematic deviations of the model from the data. Similarly, comparisons across plots of the data from each observer showed the shape of model was consistent across observers.

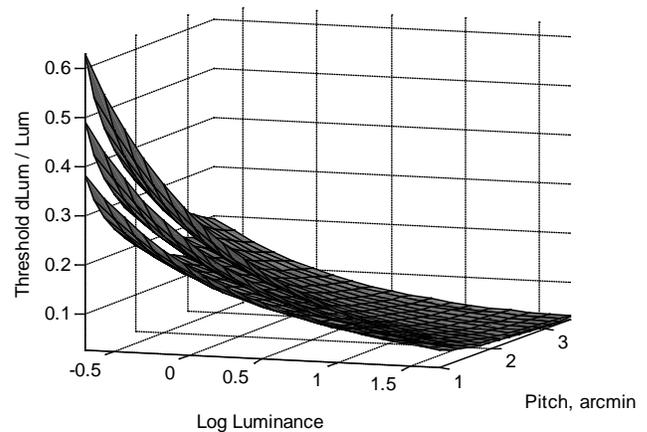


Figure 5. Model of threshold luminance step size ($dLum / Lum$) as a function of mean luminance (fL) and display pitch (arcmin). Top surface describes the 60 Hz condition while the bottom surface shows the 15 Hz condition. $p < 0.0001$, $R^2 = 0.951$, $RMSE = \text{threshold} / 5.72$

The final model summarizing the mean data of the seven observers is plotted in Figure 5 as three surfaces representing the threshold levels for the 60, 30, and 15 Hz conditions. Figure 6 provides a contour plot for the 60 Hz. Viewing condition as it is easier to read the data from this type of plot.

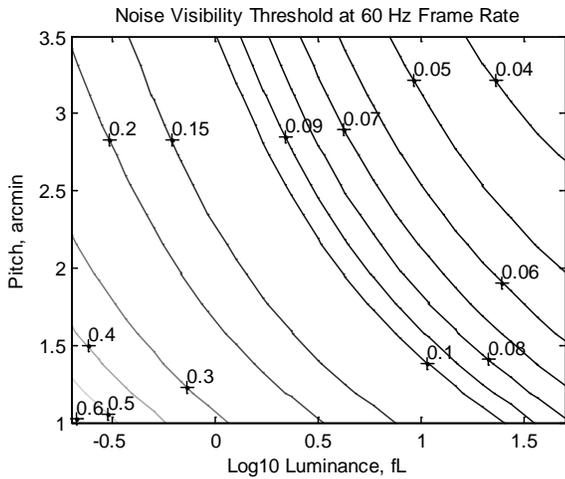


Figure 6. Contour plot of the top surface (60 Hz condition) shown in Figure 5, showing threshold luminance step size (dLum/Lum) as a function of luminance and display pitch.

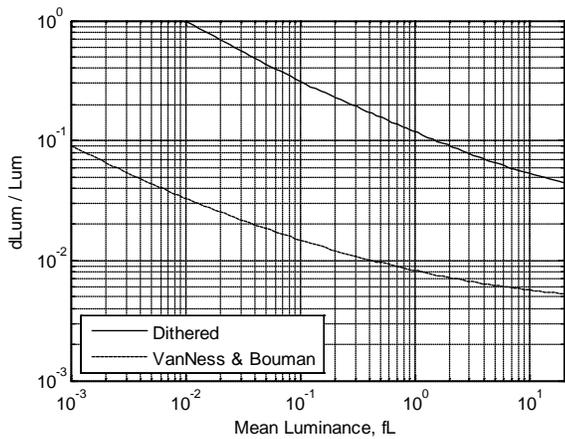


Figure 7. Lower curve: Threshold step size as a function of adapting luminance derived from the data of Van Ness and Bouman (1967). Upper curve: Size of the threshold gray scale step required when using spatial-temporal dithering at 60 Hz frame rate and a display pitch of 2.5 arcmin.

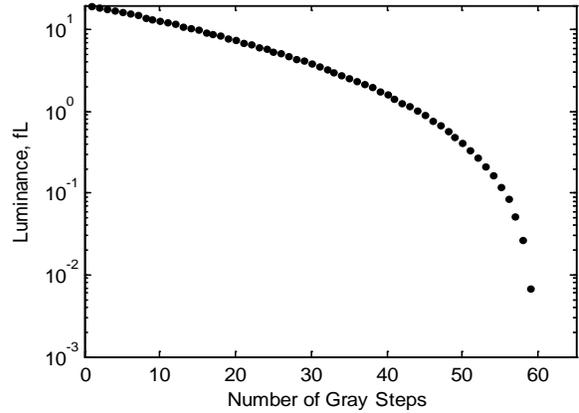


Figure 8. Optimal distribution of threshold gray scale steps spanning luminance levels less than 20 fL for the 60 Hz, 2.5 arcmin case.

Table 3. Minimum number of gray levels as a function of frame rate and pitch for a display with a maximum luminance of 20 fL.

Pitch, arcmin	15 Hz	30 Hz	60 Hz
3.5	124	101	82
3	107	87	71
2.5	90	74	60
2	73	60	48
1.5	56	45	37
1	38	31	25

Table 4. Minimum number of gray level steps/decade in luminance for luminance levels greater than 20 fL.

Pitch, arcmin	15 Hz	30 Hz	60 Hz
3.5	83	71	62
3	73	63	55
2.5	63	55	47
2	53	46	40
1.5	42	36	32
1	30	26	23

IMAGE 2009 Conference

Conclusions

The results of this evaluation allow the reader to quantify the significant reductions in the bit-depth required to transmit and display video images free of gray scale sampling artifacts that is afforded through the use of the simplest of dithering algorithms. Modern displays typical of the desk top (e.g., about 50 fL, ≥ 2.5 arcmin, ≥ 60 Hz) would require only about 80 gray steps per primary if the levels were optimally spaced. Similar results are indicated for the worst case display systems typical of flight simulation training industry (e.g., 3 arcmin, 30 Hz, and ≤ 10 fL) where about 80 levels would be required. These results suggest that the use of more than 8 bits per pixel is not indicated for transmitting and displaying high dynamic range video for any practical simulation training display system of today. The rapid advances in resolution and frame rate of displays and image generators will further reduce the demand for high bit depth image encoding.

Extension to NVG

This evaluation, conducted in early 2007 focused on the visibility of noise in visible light video. A companion evaluation was conducted in Q2 2007 in which observers viewed the noise patterns through night vision goggles. The results of this evaluation indicate that the same bit depth extension method described here works for stimulated NVG applications. This companion evaluation will be described in a future paper.

About the Authors

Dr. Charles J. Lloyd has 23 years of experience in the area of display systems and applied vision research at such organizations as the Displays and Controls Lab at Virginia Tech, the Advanced Displays Group at Honeywell, Lighting Research Center at Rensselaer Polytech, Visual Performance Inc., and BARCO Projection Systems. Charles now works at FlightSafety International where he manages the development of next-generation display and alignment systems. Charles has published/presented more than 50 papers in the field.

Mark A. Carter is a Senior Staff Engineer for the Visual Simulation Systems division of FlightSafety International in St. Louis, MO. He has participated in the design of VITAL visual systems software for the past twenty years. Mr. Carter is the primary architect for the scene graph, paging, and rendering software technology underlying VITAL visual system products. He is currently working on PC-based sensor channels and on next generation out-the-window visual system designs. He received his associate in science degree from the University of the State of New York.

References

- [1] Luckiesh, M. and Moss, F. K. 1941. The variation in visual acuity with fixation distance. *J. Opt. Soc. Am.* 31, pp 594-595.
- [2] Mese, M. and Vaidyanathan, P. P. 2002. Recent advances in digital halftoning and inverse halftoning methods. *IEEE Trans. on Circuits and Systems*, 49(6). pp. 790-805.
- [3] Reinhard, E., Ward, G., Pattanaik, S., and Debevec, P. 2006. High Dynamic Range Imaging. Elsevier, New York.
- [4] Roberts, L. G. 1962. Picture coding using pseudo random noise. *IRT Transactions on Information Theory*, IT-8, pp.145-154.
- [5] Schreiber, W. F. 1986. Fundamentals of Electronic Imaging Systems: Some Aspects of Image Processing. Springer-Verlag, New York.
- [6] Ulichney, R. 2000. A review of halftoning techniques. Compaq Computer Corp, Cambridge, MA.
- [7] VanNess, F. L. and Bouman, M. A. 1967. Spatial modulation transfer in the human eye. *J. Opt. Soc. Am.* 57, pp. 401-406.

Document Location

This document is stored on the following site:
<http://www.charleslloyd.us.com>.

The address is:

http://charleslloyd.us.com/Visibility_Of_Spatio-Temporal_Dither_Noise_Effects_Of_Display_Luminance_Pitch_And_Frame_Rate.htm