

Effects of Monitor Calibration Functions on Imagery Interpretability

Jon Leachtenauer
J/M Leachtenauer Associates Inc., Annandale, Virginia

Andrew S. Biache and Geoff Garney
Autometric Inc., Springfield, Virginia

Abstract

The inherent relationship between digital command level and output luminance of a CRT is defined by physics and follows a power law function. The contrast sensitivity of the human eye varies as a function of luminance. As luminance increases, the eye can detect smaller contrast differences. The sensitivity function is not, however, described by the CRT's power law function. Both empirical studies and theoretically based models of the human visual system have been used to define CRT calibration functions designed to optimize human contrast sensitivity.

Two empirically-based and one theoretically-based calibration functions were compared on a monochrome CRT. Imagery and Briggs targets (checkerboard patterns at varying contrast and resolution levels) were displayed under darkened ambient light conditions using each of the three calibration functions. Two different luminance ranges were evaluated, 0.1 -35 fL and 0.37-103 fL. Experienced imagery analysts rated the interpretability of the imagery using the National Imagery Interpretability Rating Scale (NIIRS). They also rated the discriminability of the Briggs targets. Results were analyzed to define the effects of the alternative calibrations on the perceived interpretability and quality of the displayed imagery and targets. Because contrast sensitivity varies with age, the effects of age on Briggs target ratings were also assessed. Finally, results of the current study were compared to those from two earlier studies in which the basic CRT power law function was evaluated.

Introduction

In the early 1990s, the American College of Radiology and the National Electrical Manufacturers Association formed a committee to develop a standard for Digital Imaging and Communications in Medicine (DICOM). The standard relates to the relationship between digital command level and display luminance.[1] The committee developed a monitor response function standard based on Barten's model[2] of the human visual system. In theory, this optimizes the ability of the human visual system to detect contrast differences in a monochrome image displayed on a CRT or other softcopy viewing device. This function will be referred to here as the NEMA/DICOM calibration function.

The NEMA/DICOM calibration was implemented in a video board used in the new National Imagery and Mapping Agency (NIMA) softcopy system, the Integrated Exploitation capability (IEC). The NEMA/DICOM calibration differed from the calibration currently used in a major current system, IDEX, and there was some concern that the NEMA/DICOM calibration might be inferior to the IDEX calibration. Specifically, the IDEX calibration assumes lower contrast sensitivity in low luminance conditions than does the NEMA/DICOM. This means that the IDEX calibration allocates more command levels at the low end of the command level range than does the NEMA/DICOM. An empirical calibration developed by Rogers and Carel[3] allocates even more command levels than the IDEX calibration at the low end and has been used in previous NIMA softcopy studies.

To assess the effects of these different calibration functions, an imagery analysts (IA) evaluation was performed. Briggs and delta-NIIRS ratings were made by imagery analysts. A high-end monochrome CRT was used and was run at two luminance ranges (0.1-35 fL and 0.37-103 fL). At each luminance range, look-up-tables (LUTs) based on each of the three calibration models were applied. Results were analyzed to define the effects of calibration alternatives on interpretability.

Background

Softcopy monitors inherently show a non-linear relationship between input command level and output luminance values. Figure 1 shows an example. The contrast sensitivity of the eye increases as luminance increases. This relationship is similar to, but not identical to, the typical monitor function. To maximize visual performance, it is desirable to modify the monitor function with a look-up-table (LUT) designed to account for the contrast sensitivity of the visual system.

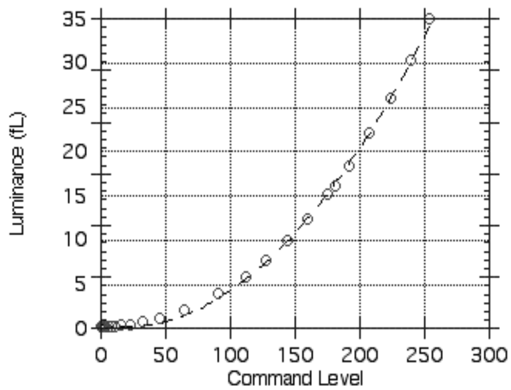

Figure 1. Typical command level/luminance function

Various investigators have attempted to develop an optimum correction function. A typical approach has been to design a test target and then to capture visual response data as a function of changes in display parameters. The IDEX calibration was developed on this basis.[3] Small target detection studies were run with the size of the target and background luminance varied. The luminance of the target was increased until it could be detected. Response data were analyzed to define modulation thresholds as a function of background and target luminance.

The function used for IDEX calibration is apparently based on this experimental data, but is not directly defined by the data. There appear to have been some empirical adjustments that are not, as far as these authors know, documented. The function is currently defined by a five-degree polynomial related to normalized command level and output luminance.

A second calibration function was defined in a study performed by Rogers and Carel.[4] Sine wave targets were displayed and adjusted until the bars could be seen. Target size, spatial frequency, luminance, and surround luminance were all varied. An eleven-term regression equation was developed to predict modulation thresholds. The threshold data are cascaded to develop a calibration function.

The NEMA/DICOM model [1] takes a different approach. Barten,[2] using theory and reference to previous empirical data, developed a model of the human visual system. The model takes into account neural and photon noise, lateral inhibition, optical MTF of the eye, integration angle, and viewing angle. Results of empirical studies were used to demonstrate the validity of the model. The NEMA/DICOM calibration defines modulation thresholds for certain viewing assumptions (two-degree target, four-cycles/degree spatial frequency). Using the user defined maximum and minimum luminance values (Lmax and Lmin), the model calculates threshold modulation values and cascades the values to define a desired calibration function.

A comparison of the IDEX and NEMA/DICOM models showed the IDEX model to be more conservative at low luminance levels. The IDEX model assumes higher Cm thresholds and thus devotes more command levels to the lower luminance levels. Predictions generated by the Rogers and Carel model were reviewed to define a calibration that might be even more conservative than the IDEX calibration. Assuming a background luminance at 25% of Lmax, a one-degree target subtense (vs. the two degrees assumed by NEMA/DICAOM) and four cycles/degree spatial frequency (same as NEMA/DICOM) results in more conservative Cm values at low luminance. Figure 2 shows a comparison of the modulation thresholds defined by the three. The NEMA/DICOM model is most optimistic regarding modulation thresholds, the Rogers and Carel least optimistic.

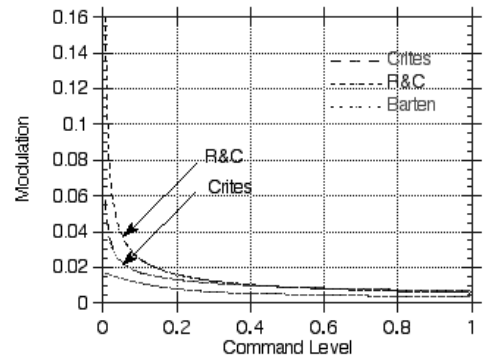

Figure 2. Comparison of modulation thresholds.

Previous studies [5,6] have compared the IDEX calibration to the basic CRT power law functions. The principal effect of the IDEX calibration has been to produce more uniform Briggs scores as a function of target brightness. Overall Briggs scores and NIIRS rating differences were quite small (less than 0.1 NIIRS, less than 3 levels for the Briggs C-7 target).

Method

A high quality monochrome monitor was calibrated to two luminance levels. Briggs targets, visible images, and radar images were run through the three calibration LUTs and displayed. Imagery analysts provided Briggs ratings and delta-NIIRS ratings on the displayed imagery. Results were analyzed to determine the effects of the alternative calibrations.

Rating Scales

The 15-level Briggs targets [7] were used in this study as opposed to the normal 8-level targets in order to achieve better sensitivity across the display dynamic range. Figure 3 shows an example of the 17 checkerboards within each of the 15 Briggs targets in a target set. The figure also shows the score associated with the boards. Boards are scored from 10 to 90. Target sets vary in terms of the command level difference between the dark and light squares in each checkerboard. Target sets are defined with command level differences of 1,3,7 and 15 counts (C-1, C-3, C-7, and C15).

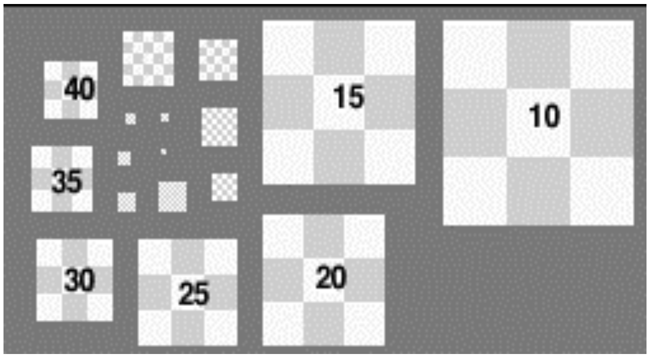

Figure 3. Example of Briggs target.

Briggs ratings require the observer to define the smallest checkerboard (in each set of 17) where the squares can be separately distinguished. The smallest target receives a score of 91, the largest a score of 6. The smallest board is then rated in terms of square definition with ratings ranging from 1 (well-defined squares) to 5 ("blobs"). These ratings are subtracted from the size scores so that the range of scores is 1 to 90. If the largest pattern in each set can not be resolved, a score of 0 is given.

NIIRS ratings [8] are made on a 0 to 9 scale. Each level is defined by six criteria or interpretation tasks; e.g., identify individual rail cars by type. The tasks are of increasing difficulty from 0 to 9 and thus require increasingly better image quality. Separate NIIRS exist for each image type (visible, IR, radar). In the present study, NIIRS difference ratings (delta-NIIRS) were made at the decimal level.

Imagery

The Briggs C-1, C-3, and C-7 targets were used in the evaluation. They were rated at both 1x and 2x magnification. Fifteen visible and ten radar images were selected for the evaluation. The visible images were selected in part on the basis of their histograms. Five were normally distributed, five were skewed to the left (dark), and five were skewed to the right (bright). The radar images all tended to be skewed to the left. The IDEX calibration was used as the standard, images with the other two calibrations were delta-NIIRS rated relative to the standard.

Monitor Calibration

The monitor used was a two mega-pixel monochrome monitor with a pixel density of 100 ppi. Using full field targets, the monitor was calibrated to a dynamic range of 0.1 to 35 fL, the standard IDEX calibration. To assess the applicability of the IDEX calibration to other than the standard IDEX dynamic range, the monitor was also calibrated to a range of 0.37 to 103 fL. The goal had been to match the two calibrations in terms of dynamic range but this was not possible. The dynamic range was 25.4 dB for the Lmax of 35 fL and 24.4 dB for the Lmax of 103 fL. The three LUTs applied to each Lmax set-up were defined according to the respective calibration functions. In the case of the IDEX calibration, this simply involved normalizing the command level and luminance range and running the defining equation. In the case of the NEMA/DICOM and Rogers and Carel calibration, modulation thresholds were computed and cascaded to generate the final LUT.

Evaluation Procedure

It was originally intended to run 12 analysts through the evaluation. Because contrast discrimination is known to degrade with age, the goal was to split the analyst sample equally across three age groups, 20-30, 30-40, and 40-50. After running several analysts, it was discovered that there was a software/hardware interaction that changed the image mapping depending on the location of the cursor in the image. Accordingly, it was necessary to begin again. In the second repetition, only ten analysts were available, five in the 40-50 age group, four in the 30-40, and one in the 20-30. The sample was therefore defined as 20-39 and 40-50.

In each case, the evaluation began with Briggs ratings at 1x and 2x magnification on the 0.1-35 fL calibration. Delta-NIIRS ratings were then made at 2x magnification. The process was repeated at the 0.37 to 103 fL calibration.

Data Analysis

Analysis began with outlier analysis. Analyses of variance were performed on all of the rating data. Variables for the Briggs data included calibration, target contrast, and target brightness. The effects of analyst age were also tested. For the delta-NIIRS ratings, variables included calibration, image type, and histogram skew.

Results

Results were analyzed to determine the performance of the three calibration models as a function of maximum luminance (Lmax). Potential interactions with analyst age and image histogram distribution were also investigated. No outliers were found in the Briggs rating data. Rater-group correlations ranged from 0.85 to 0.96. For the delta-NIIRS ratings, one IA reported no differences and one reported very large differences. Data were run with and without these two IAs; final results were reported without the data from these two IAs.

Briggs Ratings

Overall Briggs scores are shown in Figure 4. The effect of Lmax was not statistically significant and none of the calibrations differed significantly from each other. Figure 5 shows results as a function of target contrast level. Again, none of the calibration differences were statistically significant. Rating variability for the C-1 target, however, was significantly greater than for the C-3 and C-7 targets.

The effects of target brightness level are shown in Fig. 6 for the C-3 target. Target 1 is the darkest target and target 15 the brightest. Results generally follow the same pattern, except that the Rogers and Carel calibration does not show the same fall-off for the darkest target as the other two calibrations. This is consistent with the form of the calibration. The Rogers and Carel model assumes a higher Cm threshold than do the other two models.

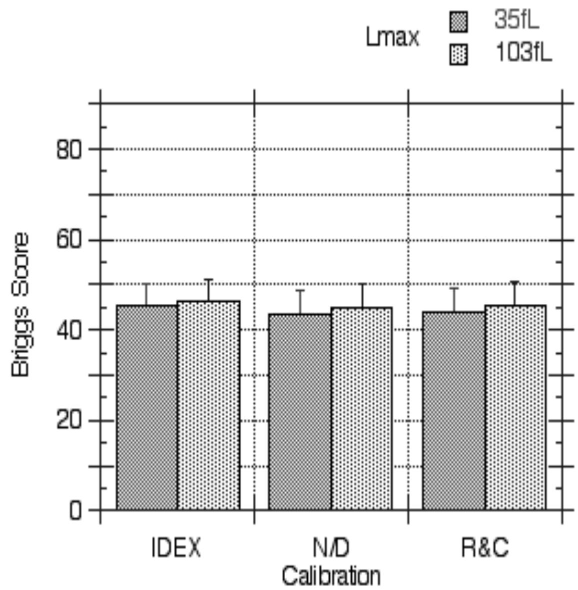

Figure 4. Overall Briggs ratings.

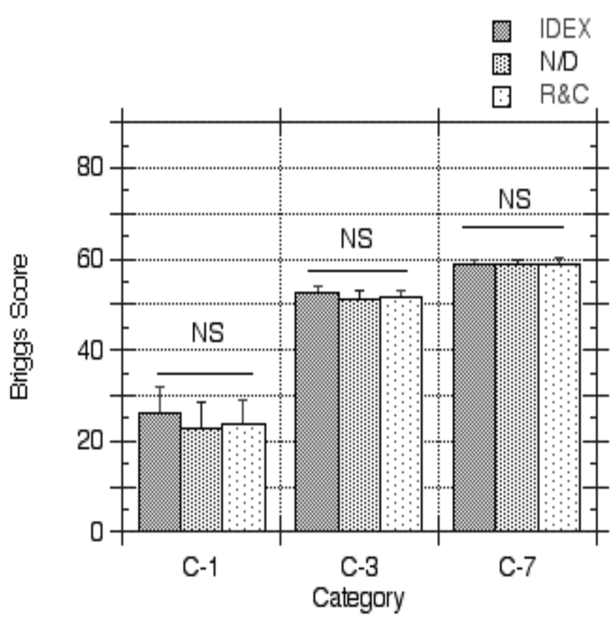

Figure 5. Briggs results as a function of target contrast.

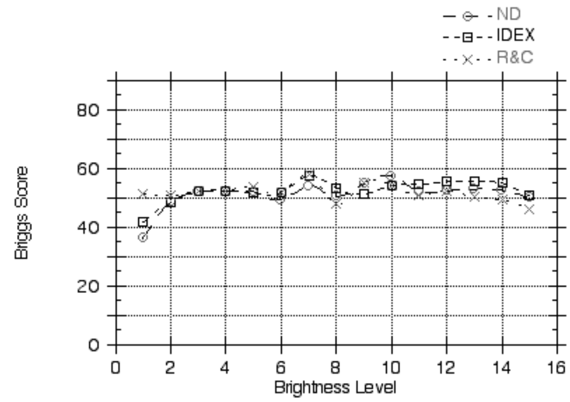

Figure 6. Effects of target brightness, C-3 target.

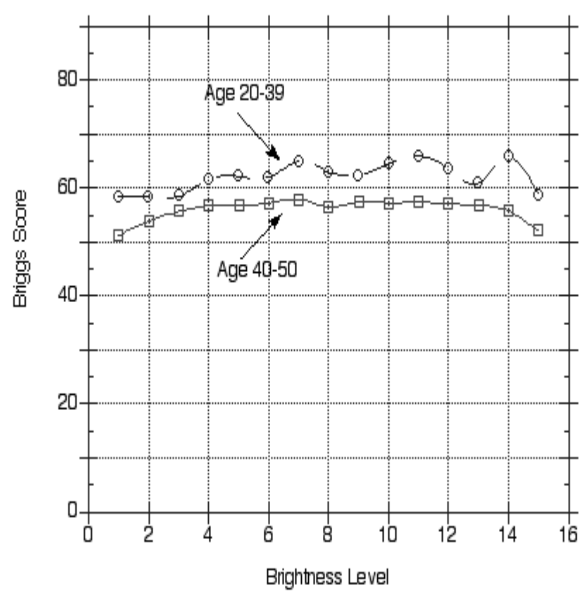

Figure 7. Effects of age, C-7 target

The effects of age group are shown in Figure 7 for the C-7 target. There was a significant target contrast by age group interaction. The difference between the two age groups decreased as contrast increased. The correlation between age and Briggs scores accounts for 39% of the variance observed in Briggs scores. With the under-30 data point removed, age accounts for 60% of the observed variance.

An attempt was made to predict Briggs results on the basis of command level differences after processing the data through the three LUTs. The attempt was not successful. It appears that other factors may also have influenced results. Absolute luminance levels and the relationship between contrast thresholds and luminance may have contributed.

Delta-NIIRS Ratings

Figure 8 shows delta-NIIRS ratings for the NEMA/DICOM and Rogers and Carel calibrations relative to the IDEX calibration. There is no statistically significant difference between the NEMA/DICOM and IDEX calibrations; ratings are significantly lower on the Rogers and Carel calibration.

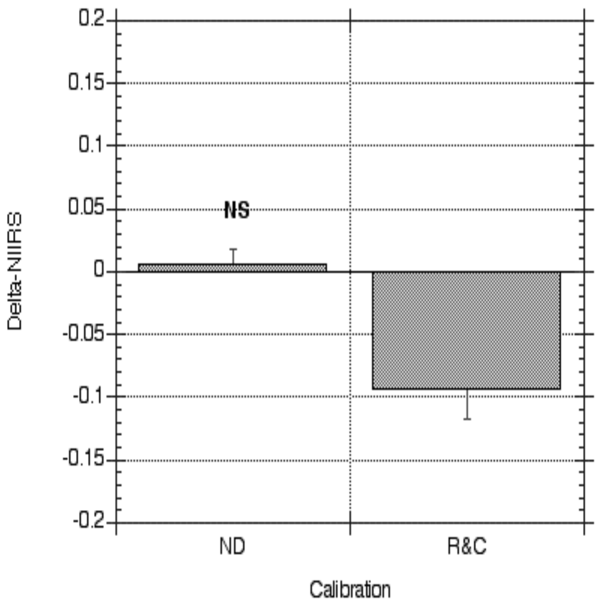

Figure 8. Delta-NIIRS ratings for NEMA/DICOM and Rogers and Carel calibrations relative to IDEX calibration.

The impact of varying Lmax is shown in Figure 9. Again, the NEMA/DICOM calibration does not differ. significantly from the IDEX. The effect of image type and histogram skew are shown in Figure 10. NEMA/DICOM ratings are significantly higher than IDEX for radar imagery but not for visible. Ratings for the Rogers and Carel calibration are significantly lower than IDEX for all image types.

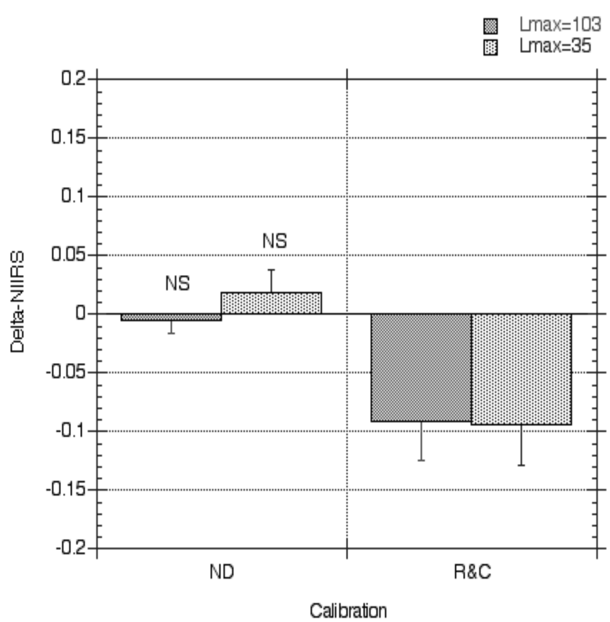

Figure 9. Effect of Lmax on delta-NIIRS ratings.

Conclusions and Recommendations

None of the calibrations showed statistically significant overall differences in Briggs ratings. Results appeared at least loosely correlated with the modulation thresholds defined for each of the calibrations. The Rogers and Carel function showed larger thresholds at low luminance levels; this appeared to be reflected in the Briggs scores.

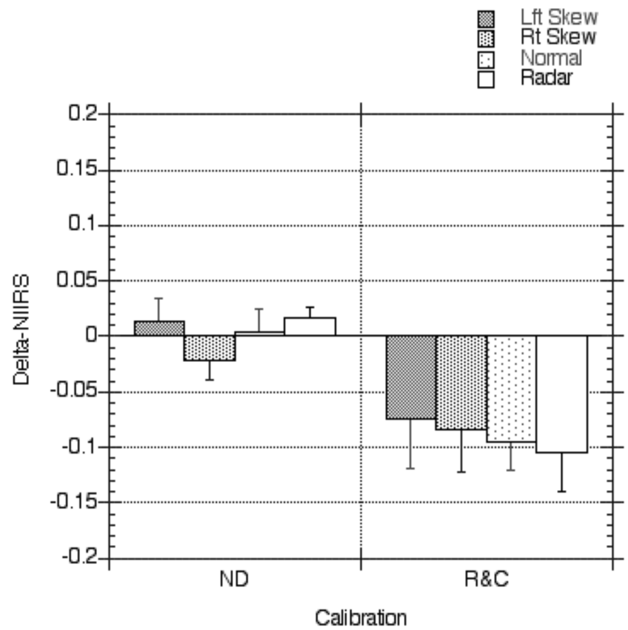

Figure 10. Effect of image type and skew on delta-NIIRS.

Rating variability was significantly greater for the C-1 target as opposed to the C-3 and C-7, this may in part have been related to the age group effect and interaction observed. Briggs scores were significantly higher for the 20-40 year old group as opposed to the 40-50 year olds. The difference was greatest for the C-1 target and least with the C-7 target (although still significant). There was not a significant interaction between calibration and age group, indicating that the findings are applicable to both groups.

For the delta-NIIRS ratings, the NEMA/DICOM and IDEX calibrations did not differ significantly on an overall basis. The NEMA/DICOM calibration provided statistically significant higher ratings for radar imagery, the difference for visible was not significant. The Rogers and Carel calibration showed results, which were significantly lower than the IDEX (and NEMA/DICOM) calibration.

Based on these findings, it is recommended that the NEMA/DICOM calibration be used for IEC applications where the calibration has been integrated into the video driver. Where this is not the case, the NEMA/DICOM calibration is recommended because it is tied to a model of the human visual system. For those situations where the IDEX calibration is in place, no change is recommended. The small difference between the NEMA/DICOM and IDEX calibrations does not warrant any resource expenditure.

Acknowledgements

The study reported here was funded and carried out at the National Imagery and Mapping Agency. The authors would like to thank Mr. Michael Parcell for his support, as well as the imagery analysts who participated in this study.

References

1. National Electrical Manufacturers Association (NEMA), Digital Imaging and Communications in Medicine (DICOM), Part 14 "Grayscale Display Function Standard, Supplement 28, 28 Jan 1998
2. Barten, Peter G. J., "Physical Model for the Contrast Sensitivity of the Human Eye", *SPIE* Vol **1666**, Human Vision, Visual Processing, and Digital Display III, 1992
3. Crites, Clay W. et al, Visual Tonal Discrimination Characteristics and their Effect on Display Tonal Calibration, General Electric Corporation, Space Division, 10 July 1986
4. Rogers, J. G., and Carel, W. L., Development of Design Criteria for Sensor Displays, Hughes Aircraft Co, Culver City CA, 1973
5. Leachtenauer, J. and Salvaggio, N. L., "NIIRS Prediction: Use of the Briggs Target", *ASPRS/ASCM Annual Convention and Exhibition Technical papers*, Volume 1, remote Sensing and Photogrammetry, American Society for Photogrammetry and remote Sensing and American Congress on Surveying and mapping, Baltimore MD, April 22-25, 1996, pp282-291
6. Leachtenauer, J. and Salvaggio, N., "Color Monitor Calibration for Display of Multispectral Imagery", *Society for Information Display, International Symposium, Digest of Technical Papers*, Boston MA, May 13-15, 1997, PP 1037-1040
7. Briggs, S. J. Manual Digital Test Target BTP#4, d180-25066-1, Boeing Aerospace Co., Seattle WA, January 1979
8. Leachtenauer, J. "National Imagery Interpretability Rating Scales: Overview and Product Description", *ASPRS/ACSM Annual Convention and Exhibition, Technical Papers*, Volume 1, Remote Sensing and Photogrammetry, American Society for Photogrammetry and Remote Sensing and American Congress on Surveying and Mapping, Baltimore MD, April 1996, pp. 272-281.

Biography

Jon Leachtenauer received his BA and MS degrees from Syracuse University. For the past 40 years, he has been engaged in human factors research on image and display quality. He recently established a consulting firm (J/M Leachtenauer Associates Inc.) and is a consultant within the Office of Technology, National Imagery and Mapping Agency. He consults in the areas of evaluation design, image and display quality and utility, and scales development.