70.1: *Distinguished Paper:* Achieving High Color Reproduction Accuracy in LCDs for Color-Critical Applications

Louis D. Silverstein[1], Syed F. Hashmi[2], Karl Lang[3], Elizabeth A. Krupinski[2], William Dallas[2] and Hans Roehrig[2]

[1]VCD Sciences, Inc., Scottsdale, Arizona USA

[2]Radiology Research Laboratory, University of Arizona, Tucson, Arizona USA

[3]Lumita, Inc., Madison, Wisconsin USA

Abstract

A methodology and associated software modules for calibration, characterization and profiling of color LCDs for color-critical applications such as medical imaging is described. Supporting analyses reveal very high color reproduction accuracy as determined by CIE DE2000 color differences for 210 test colors uniformly distributed in CIE Lab color space.

1. Introduction

Electronic color displays have become a part of everyday life. From television receivers to computer monitors to mobile communications and computing devices, modern color technology has forever changed the way we view information as well as our expectations regarding color reproduction accuracy. As color imaging technology migrates into critical imaging applications such as medical imaging, it becomes increasingly important to ensure the accuracy of color reproduction. In the case of medical imaging, for example, the color reproduction accuracy of display systems may affect the accuracy and/or efficiency of diagnostic decisions.

While radiology deals primarily with grayscale images and pseudo-color images, emerging medical imaging applications in telemedicine, such as telepathology and teledermatology, rely on true color information for interpretation and diagnosis. The ready availability of medical-grade color liquid crystal displays (LCDs) has resulted in a situation where even grayscale radiological images are increasingly displayed on calibrated color displays with a tone-reproduction curve conforming to the gray-scale standard display function (GSDF) specified by the committee for Digital Imaging and Communications in Medicine (DICOM).[1] Despite evidence that proper display calibration and color management may improve diagnostic performance and/or diagnostic efficiency, medical color displays are often deployed with little or no calibration and no means of system color management to ensure accurate and consistent reproduction color imagery.

Currently there is little guidance available for either the calibration or characterization of medical color displays. The DICOM working group has proposed to utilize color-device profiles conforming to the International Color Consortium (ICC) color profile specification standard.[2] Although the use of ICC profiles is now widespread for color management systems for graphics, photography and printing, the ICC profile specification basically provides an architecture, profile format and data structure for color management and color data interchange between different color imaging devices. It is not a general prescription for the measurement, characterization or calibration of color imaging technology nor does it specifically address the requirements and issues for color displays or medical applications. Moreover, there are many alternative approaches and options to consider when generating an ICC color profile. We consider the most appropriate protocols for color display calibration, characterization and profiling for the medical field to be an application-specific set of methods, procedures and measurement instrumentation.

A multitude of color calibration and color profiling techniques as well as associated hardware and software systems are available, but they have not been extensively evaluated for the medical imaging environment or other color-critical applications. In one recent study three commercially-available color display calibration/profiling software packages were evaluated along with three different, low-cost color sensors supported by these packages.[3] Such sensors are not true XYZ colorimeters, and the accuracy of their colorimetric estimates are generally dependent upon color-correction matrices derived from statistical error minimization methods (i.e., least-square fits or single-value decomposition). Color reproduction accuracy was based on a limited set of 25 test colors consisting of colors 1-17 of the Gretag Macbeth Color Checker Chart, the sRGB primaries and 5 neutrals. The results were highly variable and yielded average CIE DE2000[4] color-difference values ranging from 2.58 to 6.14. Another recent study examined six different displays and two calibration/profiling packages with low-cost sensors and found a similar range of color-difference values, although the older CIE 1976 ΔE^*_{ab} color-difference equations were used.[5] These levels of color reproduction accuracy and the accompanying variability are unacceptable for medical imaging. Moreover, both of these studies were based on a very limited sampling of color space and, in general, methods for assessing color reproduction accuracy have often lacked robustness and consistency. For these reasons, we have developed a rigorous methodology and associated software modules for calibration, characterization and profiling of color LCDs for medical imaging and other color-critical applications.

2. Methods

All colorimetric and photometric measurements were collected using a Photo Research PR-670 spectroradiometer. A Photo Research LRS-455 calibration standard source was available for periodic checks on the calibration of the PR-670. The PR-670 and all test displays were mounted on an optical bench with fixtures to assure centering and angular alignment of the test displays and facilitate positioning of the measurement instrument. Ambient illumination was set at 35 lx measured at the plane perpendicular to the display surface. This level of illumination was selected as a typical average level for reading rooms used for telemedicine and radiology displays. Figure 1 shows the physical measurement configuration.

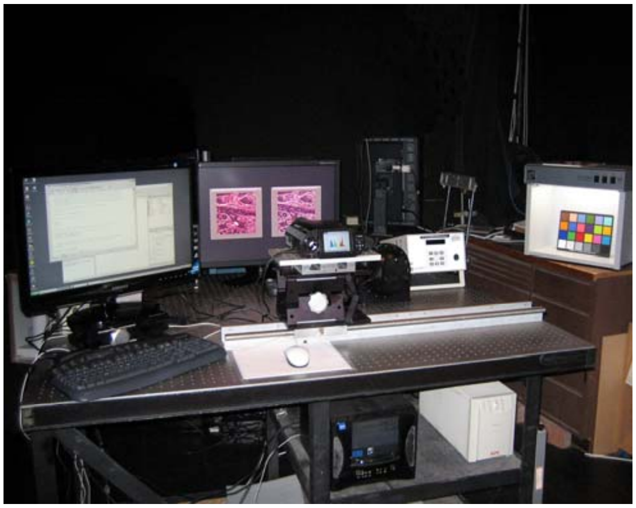

Figure 1. Physical measurement configuration.

Three test displays were utilized during the development and verification of our color management methods: two medical-grade color LCDs (NDS Surgical Imaging E3cHB and Eizo Radiforce RX-320) and one high-end professional graphics LCD (NEC LCD2690W2-BK). An additional mid-level LCD, the Dell U2410, was also investigated. The U2410 is a wide-gamut LCD notable for its internal color calibration features, which enable switching between several calibrations (sRGB, Adobe RGB and standard) via front-panel presets. All displays were driven with an NVIDIA Quadro FX3700 graphics card. Custom software developed using the MatLab programming was used to automate the generation of color test patterns and provided closed-loop control of the PR-670 to acquire and format colorimetric measurements. In addition, MatLab modules were developed to process the acquired measurements into a complete display characterization, generate an ICC version 2 or version 4 color display profile and perform an automated analysis of the color reproduction accuracy of the display.

In order to take advantage of the 10 – 12 bit look-up tables (LUTs) in each of the test LCDs for increasing the precision of tone-reproduction curve (TRC) and white-point adjustments, the initial calibration for each display was performed using the sensor and calibration software supplied with each display. The target calibration values were corrected by measurements made with the PR-670. The target calibration values were as follows: a TRC following the CIE L* standard; a white-point corresponding to D65; a peak luminance of approximately 300 cd/m2 for the NDS and Eizo LCDs and 150 cd/m2 for the NEC LCD; and a dynamic range of approximately 400:1 for all LCDs. After the initial display calibrations a MatLab characterization and profiling module automatically generated the appropriate color test patterns and measured the following parameters: XYZ of the RGB display primaries; XYZ of the white point and black point; and the tone-reproduction curve (TRC) for each color channel at 32 points. The characterization/profiling module then uses the Bradford chromatic adaptation matrix[2] to compute a chromatic adaptation transform (CHAD) relating the ICC standard D50 white point to the measured white point of the display[2], performs a black-level correction (described below), interpolates the TRCs up to the number of addressable levels using a cubic spline, scales and normalizes all measurements as appropriate for ICC profiles, and generates a fully-compliant ICC version 2 or version 4 color profile compatible with all major color-management modules (CMMs).

Properly measuring and accounting for the contributions of the display black level is important for color reproduction with LCDs. Since the LC panel serves as an array of filtered light valves which modulate illumination from a backlight module which is constantly emitting light, there is always some light leakage in the pixel off state. This light leakage combines with any reflected ambient illumination to determine the effective black level of the display. In order to properly implement black-level correction and encode it in a manner compatible with the ICC color profile structure, we developed a new method consisting of the following steps: 1) the tristimulus values (XYZ) of the RGB primaries, white point, and black level are all measured under the reference illumination; 2) the black level XYZ values are subtracted from the XYZ values of each of the RGB primaries and the measured white point; 3) the black level XYZ values are transformed into display RGB values using the 3x3 tristimulus matrix formed by the black-level corrected primaries; 4) the TRC for each color channel is adjusted and scaled by subtracting the measured luminance of level zero from each point in the TRC, adding the estimated residual R, G and B values for the black level to each point in the respective TRC and finally normalizing each TRC by its peak value.

An additional MatLab software module was developed to evaluate color reproduction accuracy using the display characterization and associated ICC profile. Sets of test colors were generated by taking the intersection of a uniform sampling of the CIE Lab color space axes with a reference color space. For the results reported in this paper, the reference color space used was the sRGB color space. Sets of 210, 538 and 839 test colors were created in this manner; however, we found little difference in the average color reproduction errors determined with the three color set sizes and have standardized on the set of 210 test colors. The CIE Lab values of the 210 test colors were transformed to display-relative XYZ using the CHAD from the display profile and then back to CIE Lab coordinates. The result is a set of reference values, which are those that a properly calibrated display should produce. The original CIE Lab values of the 210 test colors were then transformed using our display profile to display-relative RGB values. These were then realized as test patches on the center of the display. The XYZ values of the 210 test patches were automatically measured with the PR-670 and transformed to CIE Lab coordinates to form the measured color set. The CIE DE2000[4] color difference (and component difference of lightness, hue angle and chroma) were computed for each of the 210 pairs of reference and measured colors. The module generates summary statistics and histograms from the color-difference data.

3. Results

The use of our methodology for color control of LCDs resulted in excellent color reproduction accuracy (i.e., low CIE DE2000 color differences) for two of the LCDs utilized during the development and verification phase of the project: for the NEC 2690 the mean color difference for the 210 test colors was 0.72 (SD = 0.25, Max DE = 1.94) while the NDS E3cHB yielded a mean color difference of 0.71 (SD = 0.30, Max DE = 2.10). Color reproduction accuracy for the third test LCD, the Eizo Radiforce RX-320, was good but not at the level demonstrated by the NEC and NDS displays. For the RX-320 the mean DE2000 color difference for the 210 test colors was 1.41 (SD = .59, Max DE = 3.02). We were able to attribute the larger color errors for this display to imprecision in setting the white point and variability in tracking the neutral scale, both functions of the sensor and calibration software supplied with this high-end, medical-grade color LCD. To put these results in a functional perspective, many researchers in the color science and image processing technical communities assume a de facto equivalency of a unit ΔE^*_{ab} to a unit just-noticeable-difference (JND).[6] While it is not possible to establish a single JND value of color difference for all viewing contexts and all color difference equations, it is reasonable to assume that a DE2000 value of one is approximately equal to a JND of perceived color difference.

In order to provide more detail on the pattern of color-difference results, Figure 2 shows of a set of histograms revealing the distribution of DE2000 values and component lightness, hue angle and chroma differences for the NEC 2690 and NDS E3cHB displays. It is apparent that most of the error can be attributed to lightness differences while hue shift and chroma differences contribute very little to the DE2000 total color difference.

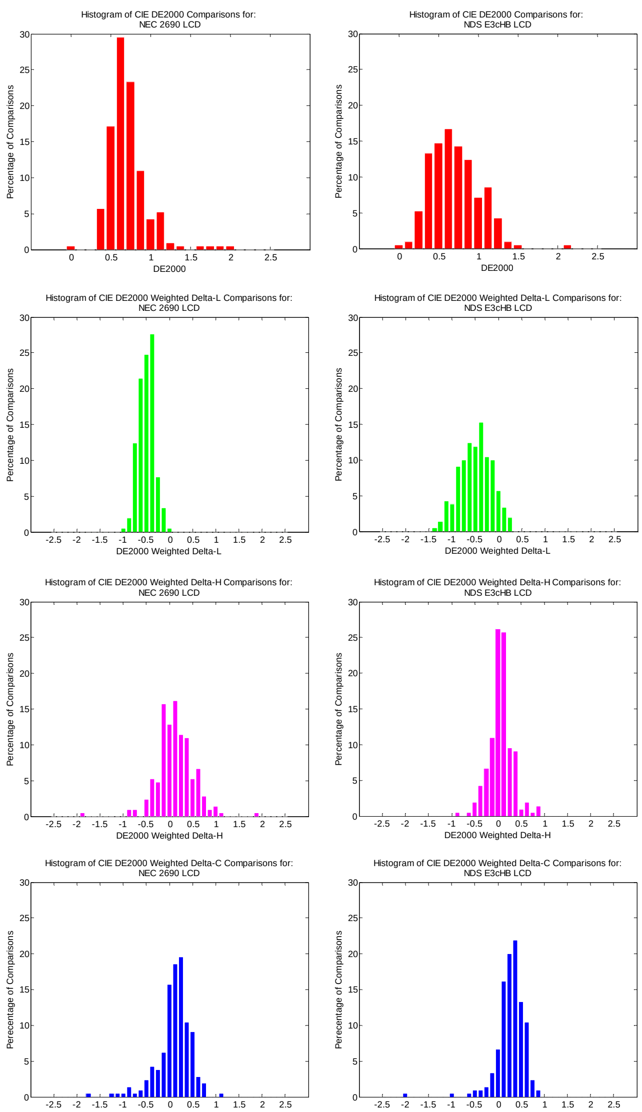

Figure 3. Histograms of DE2000 color differences, lightness differences, hue-angle differences and chroma differences for the NEC 2690 LCD (left) and NDS E3cHB LCD (right) .

The histograms reveal the general distribution of color errors, but they do not describe the pattern or direction of those errors in color space. In order to do this we generated vector plots in CIE Lab space. The base of each vector is the position of the input reference color while the tip of each vector provides the position of the color actually reproduced by the display. The length of each vector gives an indication of the magnitude of the error between reference and reproduced colors. Figures 3 and 4 show the CIE Lab vector plots for the NEC 2690 and NDS E3cHB displays, respectively.

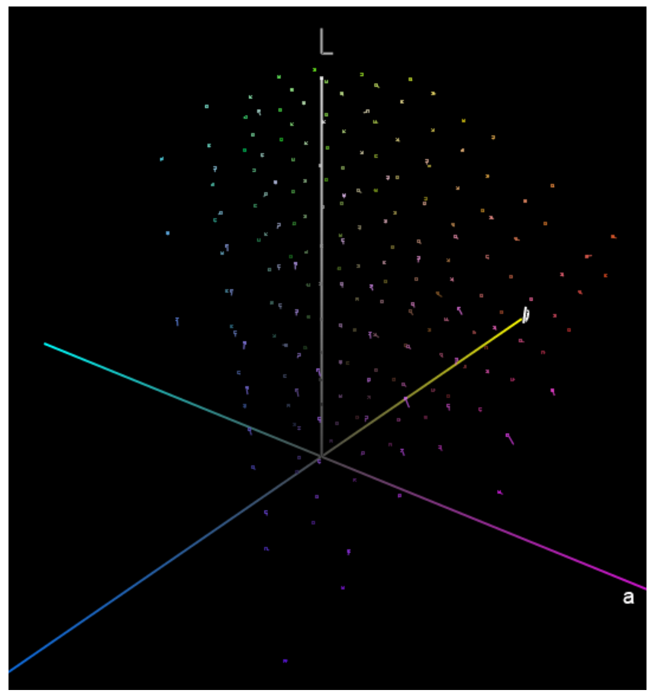

Figure 2. CIE Lab color vector plot showing reference test color points and how they are reproduced by the color-managed NEC 2690 LCD.

The final display we investigated using our methodology is the Dell U2410. As mentioned above, the U2410 is a mid-level, wide-gamut LCD notable for its internal color calibration features, which enable switching between several calibrations (sRGB, Adobe RGB and standard) via front-panel presets. Although the U2410 contains an internal 12-bit LUT, this feature is used only for processing of the preset internal calibrations and is not generally accessible to the user. We performed two separate analyses using the Dell U2410. The first analysis utilized the standard preset mode and we compared color reproduction accuracy for the Dell-supplied ICC profile for this display with our methodology and with a display color profile created with a DataColor Spyder 3 Elite sensor/software package.[7] The Spyder 3 Elite sensor and software were used only for display characterization and creation of an ICC V4.0 profile for the Dell U2410. For the second analysis we evaluated the efficacy of the sRGB calibration preset of the U2410. In this case we reproduced the 210 sRGB test colors using an unmanaged, direct pass through to the U2410 and assessed the resulting color reproduction accuracy.

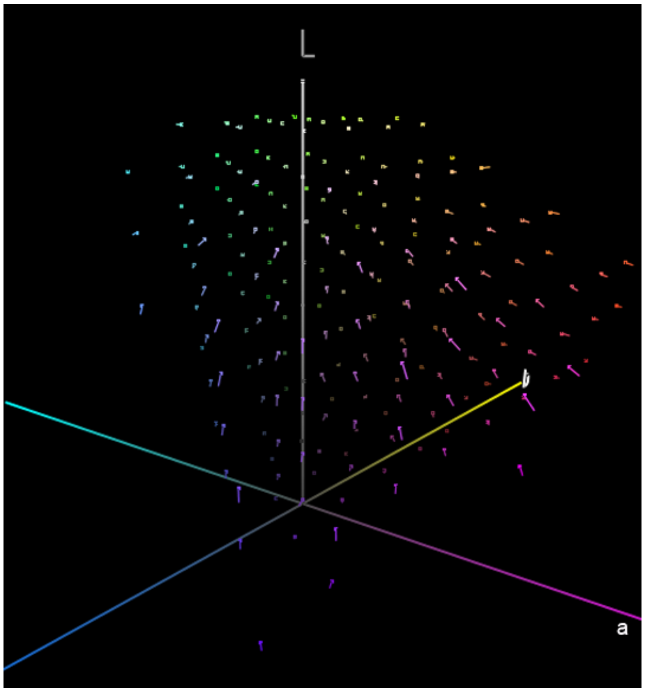

Figure 4. CIE Lab vector plot showing reference test color points and how they are reproduced by the color-managed NDS E3cHB LCD.

Investigation of the Dell U2410 standard preset calibration revealed that the ICC V2,0 profile supplied by Dell for this monitor resulted in poor color reproduction accuracy (mean DE2000 color difference = 4.25, SD = .97, Max DE = 6.34). The color reproduction accuracy improved dramatically when we utilized our methodology for display characterization and profiling (mean DE2000 color difference = .92, SD = .34, Max DE = 2.53). Interestingly, the low-cost Spyder 3 Elite sensor/software package also produced a profile yielding excellent color reproduction accuracy (mean DE2000 color difference = .44, SD = .24, Max DE = 1.69). We believe the surprisingly good results for the Spyder 3 Elite, especially given results previously reported in the literature,[3,5] are probably the result of two fortuitous conditions for the Dell 2410: a matching of sensor color filters and display primaries; and a very well-behaved gamma TRC shaped in the U2410 12-bit LUT that fits well with the gamma fitting model applied in the Spyder 3 Elite software. We would not expect these conditions to generalize across a broad range of displays and calibration protocols.

Finally, we report on the efficacy of the sRGB preset mode of the Dell U2410. It was found that in this mode the U2410 provided reasonably good color reproduction accuracy (mean DE2000 color difference = 1.65, SD = 1.24, Max DE = 7.56) for the 210 sRGB test colors considering that no external calibration, characterization or color profiling are required.

4. Discussion

We have described the development and verification of a robust protocol for the calibration, characterization and profiling of color LCDs for color-critical applications such as medical imaging. However, this represents only the first phase of our research efforts. For medical applications such as telepathology and teledermatology in which we expect color to affect diagnostic accuracy and/or efficiency, we seek to determine the degree to which color reproduction accuracy influences the task performance of diagnosticians. We are currently deploying the methodology described in this paper in a psychophysical Receiver Operating Characteristic study.[8] Six pathologists are reviewing 250 benign and malignant regions of interest from "virtual" breast biopsy slides on each of two displays using a counterbalanced experimental design. One display is a calibrated/color managed LCD and the other is a matched, uncalibrated LCD without the benefit of color management. The task is to determine if the image is benign or malignant and report decision confidence. Decision times are being recorded. The hypotheses being tested are that decision accuracy and decision times will be better with color images presented on a calibrated, color-managed display.

5. Acknowledgements

The research reported in this paper was supported by NIH/ARRA Grant 1R01EB007311-01A2 to the Radiology Research Laboratory, University of Arizona. This work was conducted while the first author, Louis D. Silverstein, was a Visiting Scholar at the Radiology Research Laboratory, University of Arizona. The authors are grateful to NDS Surgical Imaging and Eizo Nanao Corporation for providing displays for the experiments.

6. References

[1] National Electronics Manufacturers Association, "Digital Imaging and Communications in Medicine (DICOM), Part 14: Gray-Scale Standard Display Function" (2001).

[2] International Color Consortium, "Specification ICC.1:2004-10, Image technology colour management - Architecture, profile format, and data structure" (2006).

[3] H. Roehrig, K. Rehm, L. D. Silverstein, W. J. Dallas, J. Fan and E. A. Krupinski, "Color calibration and color-managed medical displays: does the calibration method matter?" Proceedings of SPIE Medical Imaging 2010: Image Perception, Observer Performance, and Technology Assessment, Vol. 7627, Paper 19 (2010).

[4] M. R. Luo, G. Cui and B. Rigg, "The development of the CIE 2000 colour-difference formula: CIEDE2000." Color Research and Application, 26, 340-350 (2001).

[5] W-C. Cheng and A. Badano, "Virtual Display: A platform for evaluating display color calibration kits," Society for Information Display Digest of Technical Papers, in press.

[6] H. R. Kang, "Color Technology for Electronic Imaging Devices," Bellingham, WA: SPIE Optical Engineering Press (1997)

[7] http://spyder.datacolor.com/product-mc-s3elite.php

[8] R.S. Weinstein, A.R. Graham et al., "Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future." Human Pathology, 40, 1057-1069 (2009).