White Paper # 44

Level: Advanced

Date: April 2016

Revised: August 2022

Visualization of medical content on color display systems

1. Introduction

Since 1993, the International Color Consortium (ICC) has worked on standardization and evolution of color management architecture. The architecture relies on profiles which describe the color characteristics of a device in a reference color space.

The ICC[1] describes its profiles as *"… tools to translate color data created on one device into another device's native color space […] permitting tremendous flexibility to both users and vendors. For example, it allows users to be sure that their image will retain its color fidelity when moved between systems and applications"*.

This document focuses on results and recommendations for the correct use of ICC profiles for visualization of grayscale (GSDF [1]) and color medical images on color displays. These recommendations allow for visualization of medical content on color display systems, whereby DICOM GSDF images, pseudo color images and color accurate images can all be presented effectively on the same display. The results and recommendations in this document were first discussed in the ICC Medical Imaging Working Group (MIWG).

2. Calibration requirements, standards and guidelines for medical displays

Medical displays that are used for primary diagnosis need to achieve minimum performance levels to ensure that subtle details in radiological images are visible to the radiologist.

The National Electrical Manufacturers Association (NEMA) and The American College of Radiology (ACR) developed a Standard for Digital Imaging and Communications in Medicine (DICOM) [1]. Part 14 describes the Grayscale Standard Display Function (GSDF). This standard defines a way to take the existing

[1] http://color.org/abouticc.xalter

Characteristic Curve of a display system (i.e. the relationship between the Luminance Output for each Digital Driving Level or pixel value) and modify it to the Grayscale Standard Display Function.

GSDF has become the generally accepted worldwide standard for calibration of display systems for use in radiology. In many countries throughout the world calibration to GSDF is a requirement in local regulations, laws and guidelines for display systems that are used to diagnose and review radiological images. In some countries a distinction is made between displays used for diagnosis versus review of radiological images (see e.g. [2]) where specifications for primary diagnosis displays are set higher than for clinical review displays.

For medical displays systems used for applications other than radiology, the situation is less clear. There is no generally accepted standard today that describes how display systems for non-radiology modalities, such as ophthalmology, endoscopy, pathology, etc., need to be calibrated. Sometimes these displays are also calibrated to GSDF, but in other cases another standard such as sRGB is used. Typically the display system in such applications is closely coupled to the acquisition device. It is typically the manufacturer of the acquisition device that designs, integrates and tests the display, ensures that it has appropriate characteristics and obtains the required regulatory approvals.

There is no agreed upon standard for the specific color characteristics that are needed for radiology displays yet. There are several works in progress on this matter and one of these is described in more detail in section 3.2.2 of this whitepaper.

Displays that are not used by medical professionals to make diagnosis or review clinical images typically are not calibrated to GSDF. This is the case, for example, for displays used by referring physicians looking at patient records.

The recommendations in this whitepaper therefore are mainly appropriate for situations where displays are calibrated to DICOM GSDF such as displays used for primary diagnosis or review of radiological images.

3. Different uses of color in medical imaging

This section describes three typical uses of color in medical applications.

3.1. Absolute color reproduction for medical images

Depending on the specific field of medicine, requirements for the representation of colors may vary. For instance, for the interpretation of wound photographs, color is an indicator of the healing state of the wound and correct representation of colors is important, and dermatologist are demanding standardization [3].

The use of ICC profiles for achieving a good color reproduction across devices [4] is already a well-established practice in different fields including pathology [5], [6]. By connecting the acquisition device profile to the display profile, it is possible to achieve good color reproducibility thanks to the colorimetric rendering intents.

ICC MIWG is currently working on a definition of best practices for digital color photography in medicine[2], which will cover the use cases where high color fidelity is required.

3.2. Perceptually linear visualization of medical images

3.2.1. DICOM Grayscale Standard Display Function (GSDF)

A known issue with the distribution of images on hardcopy or softcopy media is that images are usually inconsistent and can have different perceptions [7]. This means that depending on the hardware, images will have different contrast values or luminance differences. DICOM [1] has proposed a standard, called GSDF, for the purpose of ensuring that grayscale radiology images are presented consistently on different devices.

GSDF is a relative calibration method which aims to linearize the perception of luminance of a display without reducing its luminance dynamic range. The perceived variation is expressed as Just Noticeable Difference (JND) and is based on Barten's Contrast Sensitivity Function [8], [9]. A good introduction to GSDF is available by Fetterly et al. [10].

This standard display calibration is applicable for any grayscale medical imaging modality, even if combined with pseudo colors (annotations, fusion of modalities like PET-CT...), and it has positive impact on diagnostic performances [11]. However, true color medical images like endoscopy or dermatology are out-of-scope.

3.2.2. Color Standard Display Function (CSDF)

In recent years, medical imaging data has been evolving from pure grayscale images to color images. As of this writing, color medical imaging has not been standardized, although there are several works in progress on this matter [12] [13].

As described in paragraph 3.1, certain medical disciplines require accurate color representation. However some modalities use colors to display numerical/quantitative information on top of grayscale images as illustrated by Figure 1. The exact color used to visualize quantitative information is less of importance as long as differences are easily perceivable and it is easy for the observer to visually determine what quantitative value is being represented by a specific color.

[2] http://color.org/groups/medical/photography_best_practices.xalter

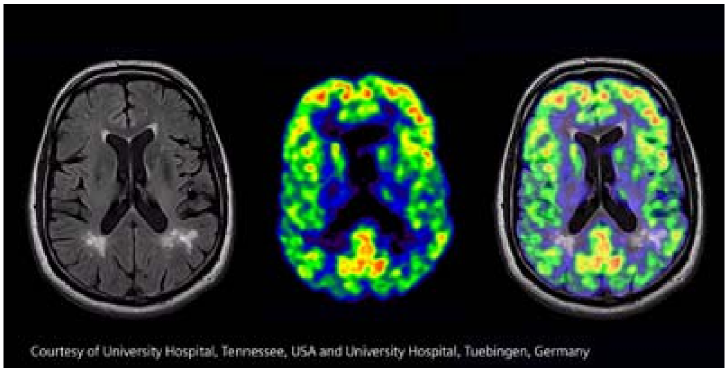


Figure 1: Example of PET-CT hybrid image. Positron Emission Tomography gives a quite accurate localization of metabolic activities represented in color and super projected on an X-ray Computed Tomography.

This quantitative imaging approach typically relies on color scales to represent calculated values. Figure 2 is an example of a commonly used "rainbow" color scale. This example associates colors with values from 0 to 1000. A large part of the color scale is covered by green, making it difficult to differentiate values from 350 to 650. On the other hand, only a thin band close to 750 is yellow, making this value clearly distinguishable. Also, depending on the quantitative value (and the color it corresponds to) it may be easy or difficult to perceive small differences in that quantitative value.

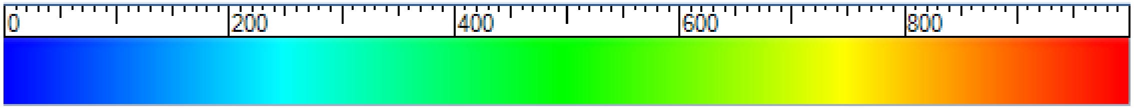


Figure 2: Rainbow color scale

Because the display is the component that in the end generates the colors, the choice of color scale and display hardware [14] affects the visual comparative analysis of pseudo-color images [15].

A perceptually linear color scale could help to optimize the visualization of the quantitative colors and reveal hidden details in the image. This can only be accomplished by taking into account the gamut of the display used for visualization.

The goal of a perceptually linear display calibration is that equal differences in RGB drive levels at the display input produce equal perceptual differences in the display output. For instance, DICOM Grayscale Standard Display Function (GSDF) is a perceptually linear calibration of the grayscale. Unfortunately, the definition of a Just Noticeable Difference (JND) given by DICOM [1] only considers luminance, and does not take chromaticity into account. For this reason, an extension of the GSDF to color cannot be achieved by using the same metric. CIEDE2000 [16], a commonly used color difference metric appears to be a good candidate for extending towards perceptually linear color behavior.

As a drawback, a completely perceptual linear display purely based on the CIEDE2000 metric would not be DICOM GSDF compliant on gray. Moreover, calibrating a display in such a way that it is perceived as being linear throughout its complete color gamut in terms of CIEDE2000 is not an easy task and obtaining a completely uniform color space requires reducing the display gamut and to decrease display luminance and display contrast. This means that the full hardware capabilities of the display cannot be used. Such a calibration is described in [17].

Recently a calibration using the CIEDE2000 color difference metric to make a display as perceptually linear as possible has been proposed without shrinking its gamut and preserving the DICOM GSDF calibration of the grayscale [18]. This calibration is called CSDF (Color Standard Display Function) and is positioned as an extension of GSDF towards color. CSDF is a possible candidate for a standardized color behavior for medical displays. The standardization process for CSDF has been started but currently CSDF is not yet an accepted standard and other candidates may be investigated as well. CSDF relies on several color sweeps through the RGB cube as depicted on Figure 3:

- the GSDF calibration of the grayscale (from Black $(0,0,0)_{RGB}$ to White $(1,1,1)_{RGB}$)
- the CIEDE2000 calibration of:
    - The different edges of the RGB cube.
    - Sweeps from Primary colors (Red $(1,0,0)_{RGB}$, Green $(0,1,0)_{RGB}$ and Blue $(0,0,1)_{RGB}$) to White $(1,1,1)_{RGB}$
    - Sweeps from Secondary colors (Cyan $(0,1,1)_{RGB}$, Magenta $(1,0,1)_{RGB}$ and Yellow $(1,1,0)_{RGB}$) to Black $(0,0,0)_{RGB}$

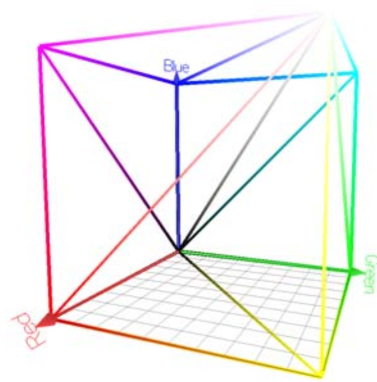


Figure 3: Color sweeps represented in the RGB cube that are made perceptually linear by CSDF. The Gray scale is linearized according to DICOM JND, and the others lines according to ΔE_{2000}.

The average CIEDE2000 step varies from one line to another to maintain the gamut integrity. The rest of the gamut colors are then adapted to ensure a smooth transition for the Gray GSDF to the CIEDE2000 calibrated colors.

4. *Proposed calibration method to achieve DICOM GSDF and CSDF calibration based on the ICC framework*

The ICC framework is widely used for correct visualization of color. The ICC MIWG is also currently working on guidelines for best practices for digital color photography in medicine, ICC White Paper 45, which will cover the use cases where high color fidelity is required.

This section focuses on how the ICC framework can be used to obtain DICOM GSDF and CSDF calibration, in addition to color accuracy.

The proposed calibration method of this section is capable of simultaneously allowing visualization of medical content on color display systems, whereby DICOM GSDF images, pseudo color images (CSDF) and color accurate images can all be presented effectively on the same display.

Medical display systems are usually able to perform automatic GSDF calibration and internally stabilize their brightness. In such situations the display continuously or periodically measures its own characteristics and the ambient light conditions by means of sensors and consequently adapts its behavior when necessary. For example, a medical display could alter internal look up tables and other settings to make sure it remains compliant with the DICOM GSDF standard. It is also possible that such a display reacts to changes in ambient light by changing its display luminance or by changing the calibration curve. For these reasons, the display behavior and therefore the “display profile” can change whenever the display adapts its internal calibration settings.

Therefore it is only possible to use a Color Management Module (CMM) with these self-calibrating displays if their ICC profiles are updated each time they change their behavior. In the case non-self-calibrating displays it is even more important that ICC profiles for these displays are regenerated at a sufficient frequency (see sections 8.4 and 9) since non-self-calibrating displays can have a behavior that fluctuates significantly over time.

In the ICC architecture, profiles connect source and destination data encodings (devices, or reference encodings, color space data, color names…). The most common usage is to connect profiles corresponding to an acquisition device and a rendering device via the Profile Connection Space (PCS). In the present case, the rendering device is a display; to which any acquisition device can be connected via the PCS.

The proposed method consists of creating two ICC profiles while characterizing a display. The first one describes its native color behavior and a second one describes the ideal calibration based on its properties. The proposed color management workflow is schematized on Figure 4.

The ICC framework defines several rendering intents. The method focuses on the Colorimetric intents as it aims to perfectly match the colors from the source profile on the display. As the gamuts described by source and destination profiles cover the same volume, there is no need for gamut mapping methods from Perceptual or Saturation rendering intents. It should be noted that in the ICC architecture, both Media-relative Colorimetric and ICC-Absolute Colorimetric intents have the same output in Display class profiles.

Both DICOM GSDF and CSDF are relative calibrations which aim at linearizing the perceptual differences between levels without shrinking the gamut of the display. It is therefore critical to accurately estimate the gamut of the device.

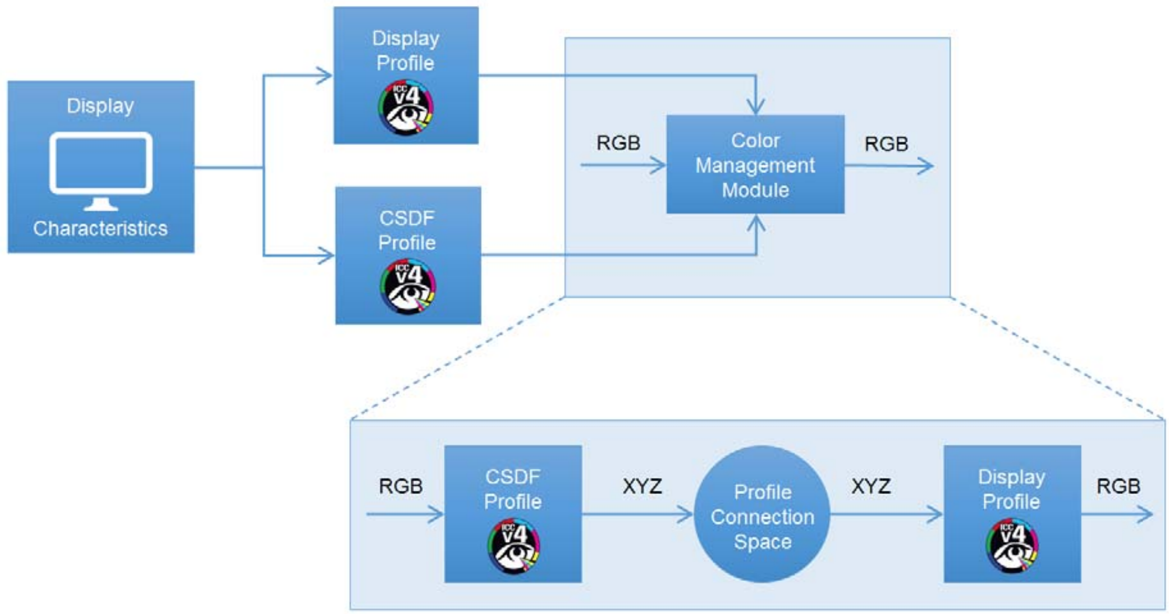


Figure 4: Workflow of ICC based CSDF display calibration.

5. Creation of the profiles

ICC specifies different type of profiles balancing between performance and memory foot-print. Annex A details the structure of each of them for the present use-case.

To achieve a DICOM GSDF only calibration, the use of monochrome profiles is possible. Monochrome profiles are designed to be used with monochromatic devices. They can be used to calibrate grayscale displays to DICOM GSDF at a low processing cost. They can also be used to calibrate color displays to DICOM GSDF. By combining a monochrome source profile and a color display profile, CMM will return RGB triplets where $R = G = B$. Thus, the color monitor will not display colors anymore.

Matrix-TRC profiles and N-component LUT based profiles are designed for color devices. Matrix profiles have a simple structure. They perform very well in describing theoretical display standards or models as the ones we use in this study (Annex B), but are unable to describe the internal constraints of a real LCD display (section A.2). In a lot of application this limitations is not really an issue, but it could be the case here.

LUT based profiles are much more powerful, but also more complex. The N-dimensional Color Look-up table (CLUT) is the core element of this structure, and is the one allowing describing a CSDF calibration in an ICC profile (see section A.3).

A final possibility is to use DeviceLink profiles. Unlike ordinary source or destination profiles, DeviceLink profiles do not describe a specific color space but the conversion from a source to a destination color space. In the present use case, it is possible to use a DeviceLink profile to describe the color transformation to be applied on a display to calibrate it. This does not influence the performance of the calibration process (see section B.2.4).

6. *Profile quality assessment methods*

When creating ICC profiles, it is important to control their correctness. Two complementary tests are proposed to validate profiles.

- The first test compares a given display model $L^*a^*b^*$ output with the corresponding ICC profile output to estimate the accuracy of the Device-To-PCS conversion of the profile (see section 6.1).
- The second test consists in Device-To-PCS followed by PCS-To-Device conversions using the same profile in order to perform a roundtrip. This second test estimates the invertibility of the profile, and combined with the test above, indirectly measure the accuracy of the PCS-To-Device conversion (section 6.2).

Both tests are detailed below.

The first test is sufficient to show that Matrix-based profiles cannot be used to describe the target CSDF calibration (see Table 3). Other profiles architectures perform very well to this test.

The second test reveals the importance of the size of the CLUT in LUT-based profiles to correctly calibrate to CSDF (see Table 8 in section 0).

6.1. Fidelity test

To estimate how well the generated profiles would emulate the display model they were based on, the following fidelity test is performed:

1. A $18 * 18 * 18$ grid of RGB digital driving levels (DDL) is generated.
2. The DDLs are fed to both a display model (or measured on a physical display) and its corresponding ICC profile to perform a Device-to-PCS conversion.
3. If necessary, both values are converted to the $L^*a^*b^*$ color space and the perceptual color difference between the two values is calculated by using CIEDE2000.

A more detailed description of this test can be observed on Figure 5.

The goal of this test is to assess whether or not the color variations induced by the creation of the profile regarding the model it is based on will be significant and may introduce perceptually critical color differences. Generally, a color difference of less than one CIEDE2000 is considered to be indistinguishable by the human eye.

Fidelity results are assessed regarding the average and maximal color differences.

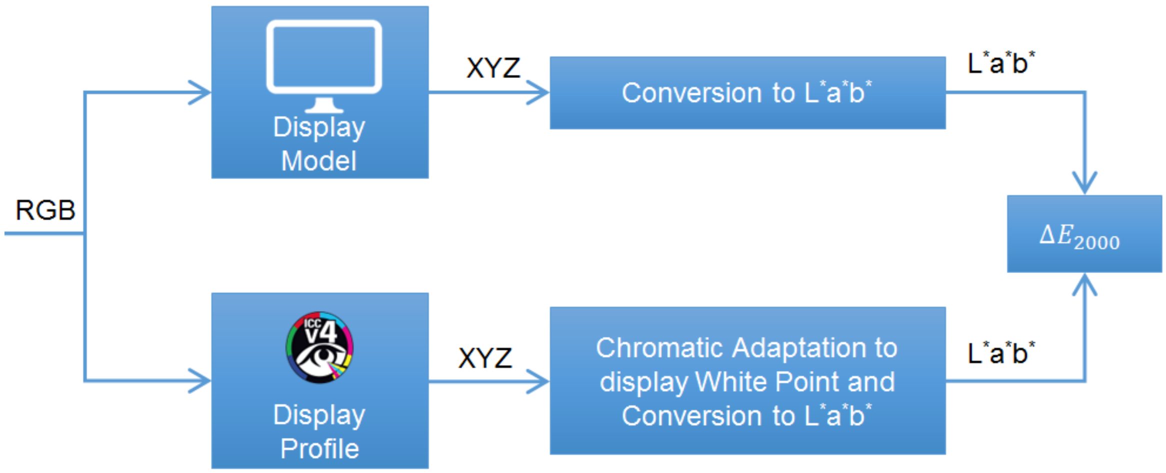


Figure 5: ICC Profile Model Fidelity test

6.2. Roundtrip test

Roundtrip test consists in connecting a profile with itself and assessing the error that is introduced by this match on an evenly spread set of $18 * 18 * 18$ points. For instance, a roundtrip test result showing a null error would assess that Device-to-PCS and PCS-to-Device conversions are exactly reversing each other, as they are theoretically meant to. Conversely, observing a large error on this test would imply that these conversions do not accurately match each other and that the profile itself induces errors in the color management process. However, the roundtrip test does not provide information about the specific cause of the conversion mismatches.

This test evaluates mismatches by calculating the CIEDE2000 perceptual color difference between two $L^*a^*b^*$ values. The first one is issued from the Device-to-PCS conversion of the profile. The second one is similarly obtained after prior application of additional Device-to-PCS and PCS-to-Device conversions as illustrated on Figure 6.

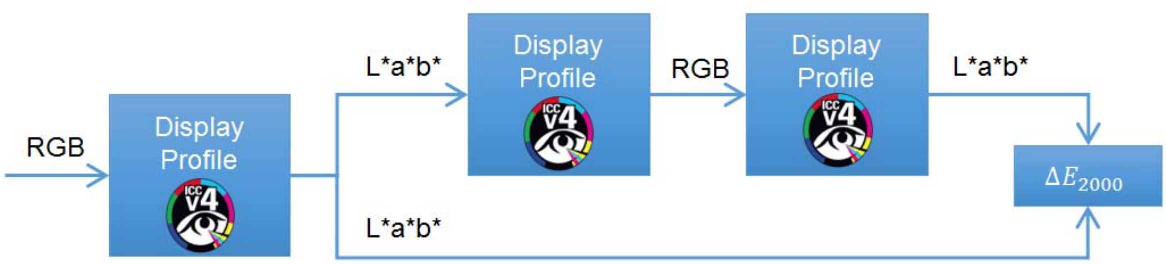


Figure 6: ICC profile roundtrip test

7. Calibration quality assessment methods

CSDF defines a different behavior for grayscale and saturated colors. For this reason they have to be evaluated separately, and a valid calibration must comply on both criteria. This means that for a display system to be CSDF compliant, both the metrics/graphs of sections 7.1 and 7.2 need to be generated and the results need to be within the described tolerances.

7.1. How to evaluate the quality of the grayscale calibration

As grayscale must comply with the DICOM GSDF standard, the quality assessment procedure for grayscale also matches the DICOM definition.

Quantitative assessment of luminance response is accomplished by using defined test patterns and luminance meters to measure the display device's luminance response for a limited number of values. The measurement protocol is similar to the one described in Annex C of the DICOM standard [1].

The grayscale compliance evaluations presented hereafter are based on luminance measurements of 18 evenly spread driving levels. These correspond to RGB triplets that can be represented as:

$$(R, G, B) = (0,0,0), (\frac{1}{17}, \frac{1}{17}, \frac{1}{17}), \ldots, (\frac{17}{17}, \frac{17}{17}, \frac{17}{17})$$

For each of them the relative difference from the theoretical Grayscale target to the observed luminance value, but also perceptual difference (JND) from one patch to the next one is evaluated.

Grayscale compliance of a system is summarized as the maximal error encountered on those 18 points. The lower the value, the better the calibration compliance score is. This score must fall within 10% ([19] section 4.3) for devices used for the interpretation of medical images (diagnostic).

Recommendations for display quality other than luminance response can be found in [19] and [20].

7.2. How to evaluate the quality of the color calibration

This section suggests a methodology for quantifying compliance/accuracy of the color component of the CSDF calibration. It is very important to stress that in this section only a metric is being described for specifically assessing the *color* aspects of CSDF calibration. As explained before, CSDF calibration also requires that the neutral grey diagonal of the display complies with DICOM GSDF (see section 7.1). For clarity reasons however, and to be able to clearly separate greyscale from color calibration performance, the reported results and graphs for "color" only refer to the metric described in 7.2.

In this recommendation, we quantify perceptual linearity of colors of a display based on the output obtained by sweeping primary and secondary colors. We define a series of 18 RGB triplets between Black and Red, with equal steps in the R channel value between them. Corresponding values are:

$$(\mathrm{R}, G, B) = (0,0,0), \left(\frac{1}{17}, 0,0\right), \ldots, \left(\frac{17}{17}, 0,0\right)$$

Likewise, evenly spread series of 18 RGB triplets are defined between Black and the other Primary colors (Green and Blue) as well as between Black and the Secondary colors (Cyan, Magenta, and Yellow). Corresponding sweep values from Black to Yellow, for example, are:

$$(R, G, B) = (0,0,0), \left(\frac{1}{17}, \frac{1}{17}, 0\right), \dots, \left(\frac{17}{17}, \frac{17}{17}, 0\right)$$

Leaving out the Black duplicates this results in 120 unique RGB triplets for which the corresponding display output is obtained as XYZ. Because discrimination between colors is less relevant at low luminance levels, we discard measurements corresponding to a driving level which results in a luminance below $5\ cd/m^2$ in the sweep between Black and White.

For example, if the triplet $(R, G, B) = (4/17, 4/17, 4/17)$ is the first measurement in the Black-To-White sweep presenting a Y value of at least $5\ cd/m^2$, then the measurements $(R, G, B) = (0,0,0)$, $\dots, (3/17,0,0)$ of the Black-To-Red sweep are discarded. The same logic is applied on the other sweeps.

All non-discarded measurements are then converted to $L^*a^*b^*$ values by taking the XYZ of full White as the reference White point. Next, we calculate CIEDE2000 between consecutive points in each of the six sweeps for the Primary and Secondary colors, resulting in six series of CIEDE2000 values noted Δ_i with i representing the color sweep (Red, Green, Blue Cyan, Magenta or Yellow). Each value $\Delta_{i,j}$ within the set Δ_i is then normalized by dividing them by the series average.

$$\overline{\Delta_\iota} = \sum_{j=1}^{N} \frac{\Delta_{i,j}}{N} \qquad \forall j,\ \ \delta_{i,j} = \frac{\Delta_{i,j}}{\overline{\Delta_\iota}}$$

For an ideal perceptually linear display, the resulting normalized curves would all be constant with value 100% ($\forall i\ \forall j,\ \ \delta_{i,j} = 1$). For each of the six sweeps, we quantify the perceptual linearity D_i as the maximum deviation from 100%.

$$D_i = \begin{cases} \max_j(\delta_{i,j}), & \max_j(\delta_{i,j}) - 1 \ \geq 1 - \min_j(\delta_{i,j}) \\ \min_j(\delta_{i,j}), & \max_j(\delta_{i,j}) - 1 \ < 1 - \min_j(\delta_{i,j}) \end{cases}$$

The overall perceptual linearity is quantified as the maximum deviation encountered in any of the six curves.

$$D = \max_i(D_i)$$

If the perceptual linearity metric value is within a predefined tolerance range, e.g. $\pm 15\%$ (i.e. $0.85 < D < 1.15$), the display calibration is considered to be perceptually linear.

The color compliance evaluations below are presented as the relative deviation from the target (i.e. values below 15% are compliant, values above are not).

The tolerance threshold is defined as a relative value because absolute values can vary a lot depending on the gamut of the device (e.g. Adobe RGB gamut presents superior $\Delta_{i,j}$ than sRGB, and thus the same variation of R, G or B would induce a larger CIEDE2000 on Adobe RGB than on sRGB).

The limit of 15% was selected based on what is achievable in practice and what is commonly used as a tolerance level for DICOM GSDF (see e.g. [19]). In case of DICOM GSDF most guidelines and local regulations require that for primary reading the maximum deviation shall be 10%, whereas for review applications this can be up to 15%. It is possible that based on future studies specific thresholds could be defined for different modalities

In parallel to the perceptual linearity of the colors, the DICOM GSDF compliance of the Grayscale must be controlled too, as GSDF is part of CSDF calibration. The method to assess the GSDF quality is described in section 7.1 and [1], [20].

7.3. Calibration smoothness

Green proposed in 2008 [21] a methodology for estimating the smoothness of a color transform from a transformed ramp. The color transform can be the result of the application of a colored 3D LUT or of ICC profiles. The method is represented by Figure 7.

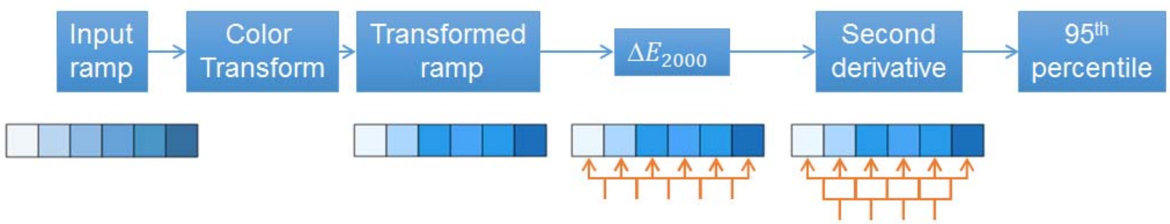


Figure 7: Green' smoothness workflow

For any input colored ramp with n pixels, the metric CIEDE2000 defined by the CIE is computed between CIELAB consecutive triplets of the ramp, resulting in $n-1$ CIEDE2000 values. From this resulting ramp, a second derivative is calculated by simply subtracting two consecutive elements of the CIEDE2000 ramp resulting in a set of $n-2$ values for which the 95^{th} percentile[3] is calculated representing the smoothness of the color transform of the input ramp.

In order to consider the entire calibration, Green's metric is applied on a large number of gradients through the RGB cube: The RGB cube is sampled to $50 * 50 * 50$ RGB triplets from which are built a total of 7500 ramps of 50 color shades.

Based on the 3D representation of the RGB cube, in the directions defined by the 3 main axes, 50-elements ramps are extracted as illustrated on Figure 8.

[3] From a study published in 2010 [29], the authors have shown that the optimum percentile level was determined to be 95th to best fit the subjective data from the measurement of the magnitude of tone jumps of 96 test gradations.

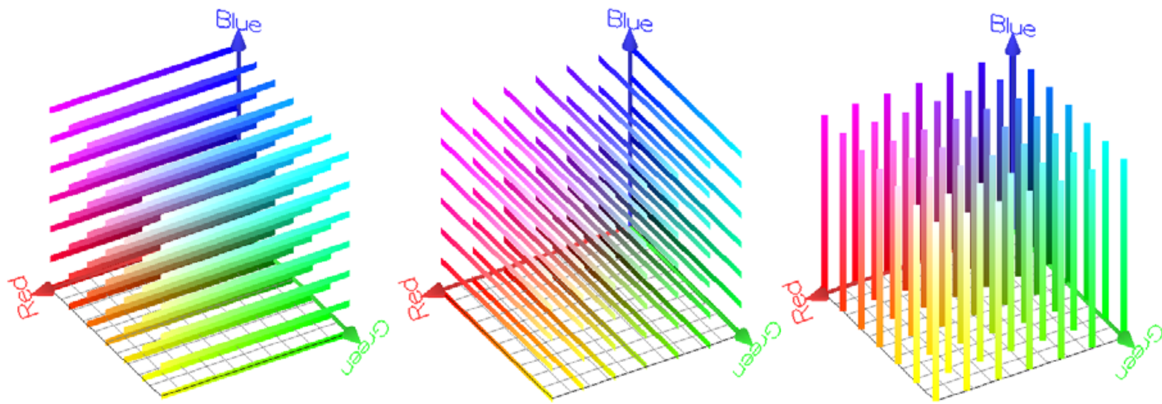


Figure 8: Examples of ramps used to evaluate the smoothness of a calibration. Here only 3 sets of $7 * 7$ ramps are represented while the tests involved a total of $3 * 50 * 50$ ramps

The smoothness of the color transform for each input ramp is computed. Simple statistics can be calculated based on the 7500 smoothness values obtained: average, standard deviation, minimum and maximum. A perfect smoothness would have a value of 0.

8. Impact of inaccurate profiling

Building ICC profiles relies on measurements which are sensitive to noise and ambient conditions. Since profiles must have a reasonable size and generation time, measuring every display color is not a viable approach. The content of a LUT-based profile may therefore rely on interpolation in the cases where all the color points of the 3D CLUTs were not necessarily measured.

Furthermore, a lot of display OSDs (on-screen-displays) make it possible for the user to select a display function in a collection of reference presets such as Gamma 2.2, Gamma 1.8, sRGB… In this situation, one could be tempted to use generic ICC profiles instead of characterizing the display in its actual configuration. However, the same preset on different displays can results in very different color rendering, and not even close to the standard they supposedly match [22].

Simulations presented in Annex B represent ideal situations where the display is perfectly characterized and its ICC profiles built on exact data (section 0 is an exception since this data is generated based on real measurements). This situation is barely realistic in practice.

The following paragraphs present the results of different simulations evaluating the impact of a misevaluation of different characteristics of the monitor or the ambient conditions.

8.1. Display luminance

A potential mismatch between profile luminance and the actual display may affect Grayscale compliance, since it is based on luminance. The display profile was fixed to $600\ cd/m^2$ and the influence of the difference between its luminance and the actual display luminance was assessed. Results are observable on Figure 9.

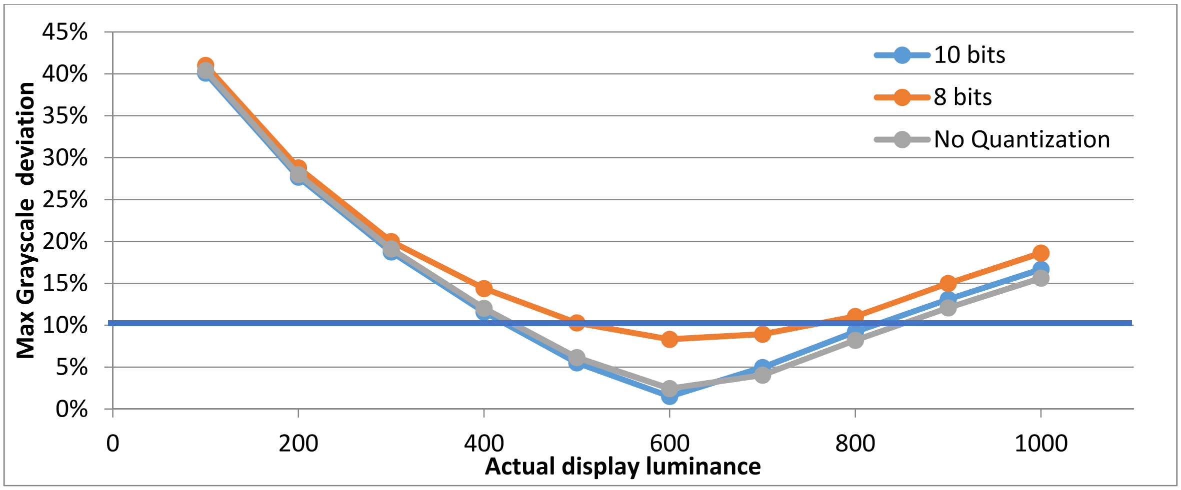


Figure 9: Influence on Grayscale compliance of Luminance mismatch between an sRGB profile describing a luminance of $600\ cd/m^2$ relatively to the actual display.

Misestimating a display luminance in its ICC profile leads to large deviations from Grayscale calibration: up to 14% error for a $200\ cd/m^2$ overestimation and 11% error for a $200\ cd/m^2$ underestimation.

For 8 bit systems, profile luminance misestimating invalidates the grayscale component of the calibration from a $100\ cd/m^2$ overestimation and approximately a $150\ cd/m^2$ underestimation. On the other hand, results remain compliant for both theoretical and 10 bit values if overestimation and underestimation does not respectively surpass $175\ cd/m^2$ and $215\ cd/m^2$.

With the presented method, a display's Grayscale calibration will thus remain valid if the ICC profile describing the display does not encompass a luminance value that deviates largely from the actual one.

Tests have been repeated with different display profile architectures and different display native behaviors without observing noticeable differences.

It is also interesting to notice that misestimating the luminance of a display has no impact on the color component of the Color compliance, as depicted by Figure 10. There is no visible difference between the theoretical values and 10 bit quantization.

It is possible to observe some variations of luminance on short term because of temperature variations within the display. Backlight efficiency depends on the lamps temperature. Liquid Crystals are also sensitive to temperature.

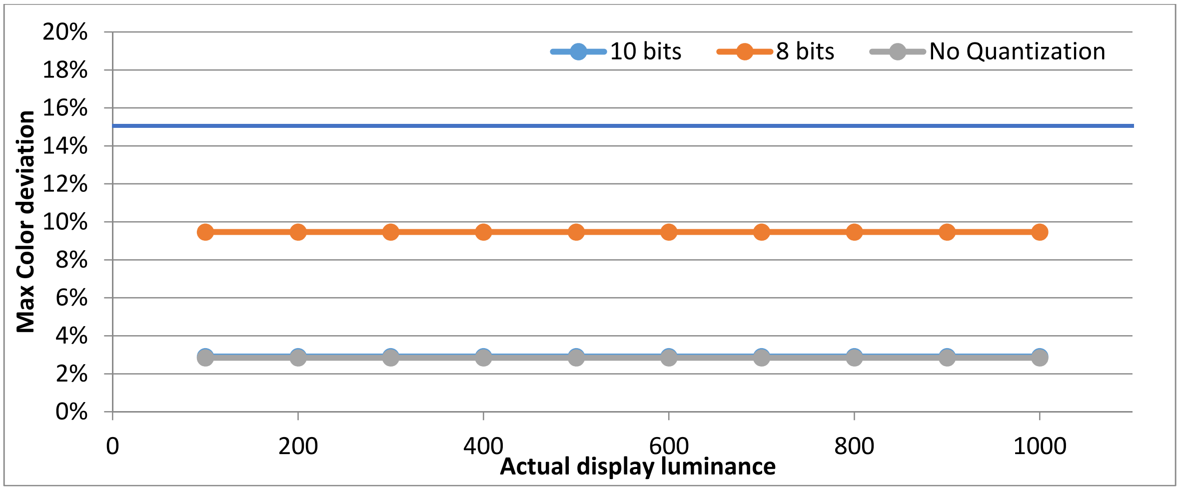


Figure 10: Influence on Color compliance of Luminance mismatch between an sRGB profile describing a luminance of $600\ cd/m^2$ relatively to the actual display.

Figure 11 shows how a display's luminance evolves from the moment it is turned on.

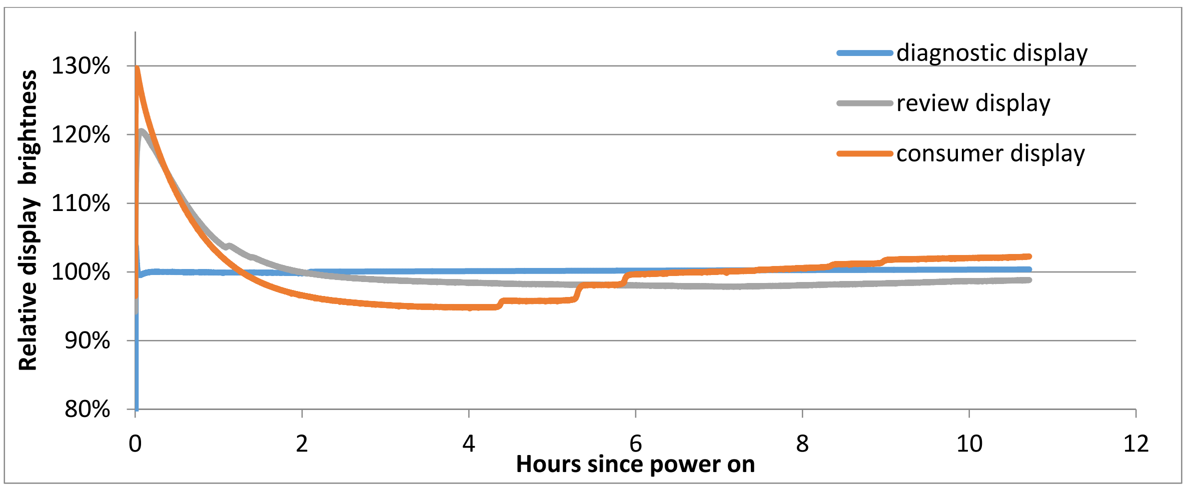


Figure 11: Short term evolution of a display luminance from start up. Diagnostic display is equipped with front and back sensors for real time stabilization. Clinical Review only has a back sensor. Consumer display is not stabilized at all. The different curves have been normalized to their average for an easier comparison. Some of the data presented here comes from [23].

Warming up creates an important overshoot during the display's first 2 hours of use, making it un-calibrated until the luminance has reached a normal level if the display cannot compensate it.

8.2. Display contrast

As with luminance inaccuracies, errors profiling contrast induce error in calibration accuracy.

Results for grayscale are assessed by evaluating Grayscale compliance for several differences between display profile contrast and actual display model contrast. Results are presented in Figure 12.

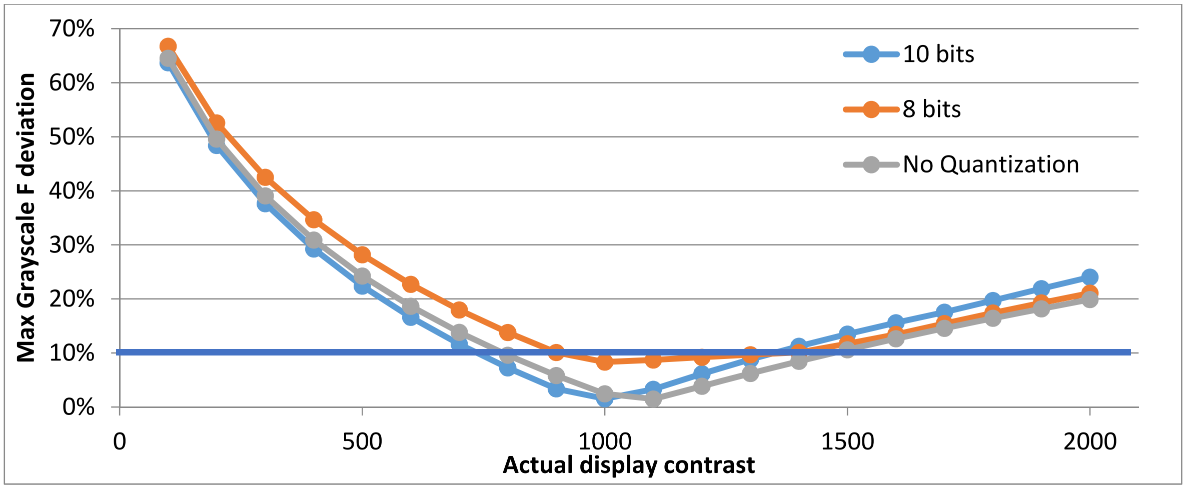


Figure 12: Influence on Grayscale compliance of contrast mismatch between an sRGB profile describing a $1000\colon 1$ contrast ratio relatively and the actual display on GSDF deviation.

Measurement devices used to quantify a display's luminance are usually much more accurate on bright levels than they are on dim ones. Even low end devices would not return an error of more than 10 to $20\ cd/m^2$ while measuring luminance close to $600 cd/m^2$. As it is observable on Figure 9, this kind of error, which would already be considered as huge, would not impact Grayscale compliance significantly.

However, contrast ratio is both much more sensitive than luminance to small variations and has a bigger impact on the calibration results. For instance, the reference display model has a luminance of $600\ cd/m^2$ and a contrast ratio of $1000\colon 1$, which means that the Black point luminance of this display is $0.6\ cd/m^2$. Measurement devices are much more likely to return an erroneous value for the dimmest luminance level of a display. In this case, even an $0.2\ cd/m^2$ error would lead to a drop from $1000\colon 1$ to $750\colon 1$ contrast ratio, which would invalidate the Grayscale calibration.

Perceptual linearity of colors is evaluated in a similar fashion, and the influence of contrast differences on the color component of the calibration are shown in Figure 13.

Contrast overestimation by the profile has a much larger influence on perceptual linearity of colors than a corresponding underestimation. For instance, if the profile's black point luminance is twice as high as the actual black luminance, deviation from color calibration raises from 3% to 8% on 10 bits systems. Contrarily, if the profile's black point luminance is twice as low as the actual one, deviation only raises from 3% to 5%.

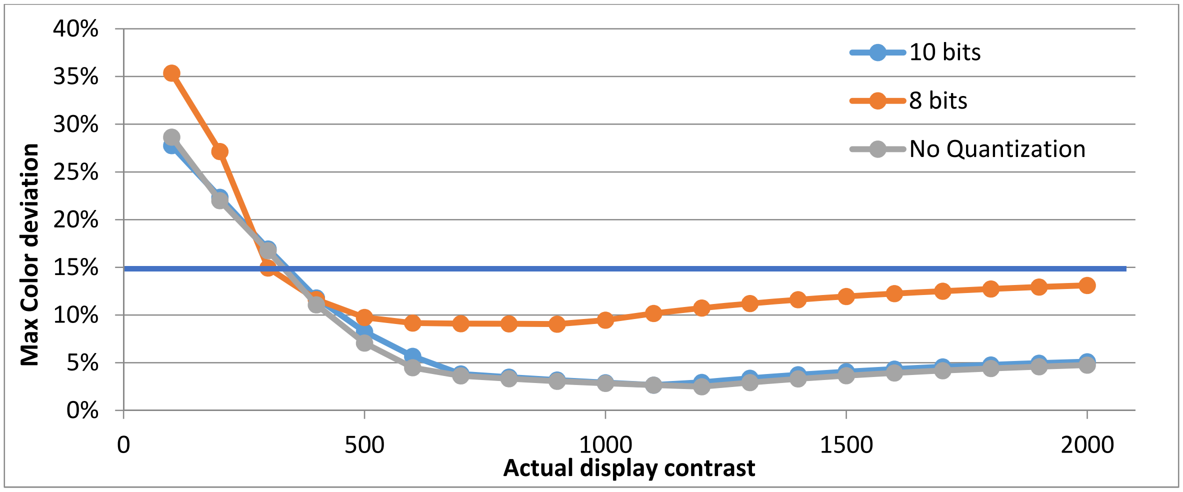


Figure 13: Influence on Color compliance of contrast mismatch between an sRGB profile describing a $1000{:}\,1$ contrast ratio relatively and the actual display contrast

8.3. Display function (gamma)

The display function is typically the parameter that can be tuned from the display settings menu. Such menus usually propose to choose among a limited number of presets. Those presets are often common to different devices, and sRGB and Gamma 2.2 are available in almost every display. One could be tempted to use the display OSD in combination with a pre-created ICC profile.

Unfortunately, display presets are usually not accurate enough to allow such practices [22].

Figure 14 illustrates the fact that Grayscale compliance is highly sensitive to imprecisions of the display functions contained in an ICC profile. An error of 0.05 in the estimation of the Gamma is indeed enough to invalidate the Grayscale calibration.

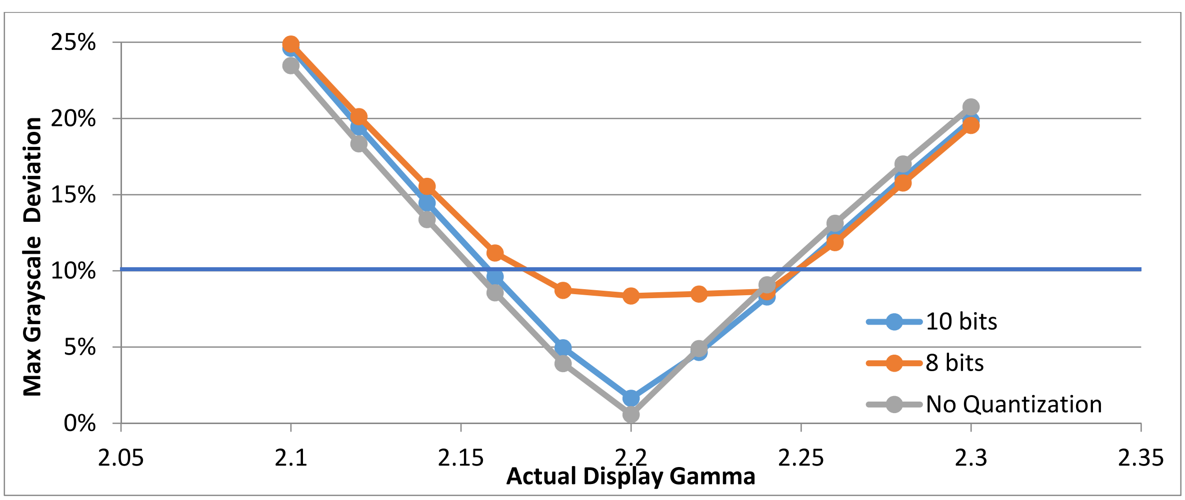


Figure 14: Influence of display function mismatch between a gamma2.2 profile and the actual display on Grayscale compliance

On the other hand, the accuracy of the display function seems to be less critical for the calibration of the colors (see Figure 15) but remains a disturbing factor.

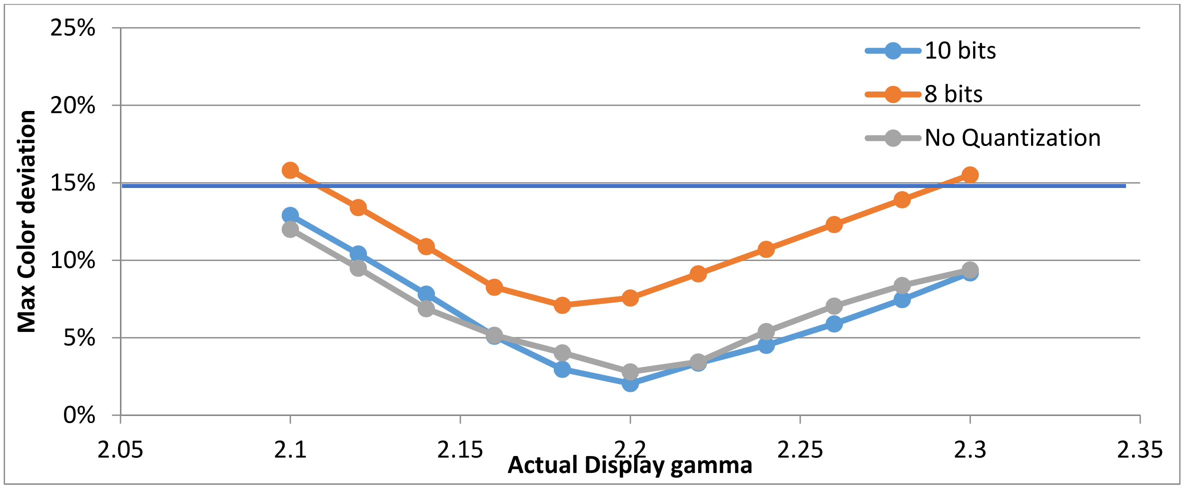


Figure 15: Influence of display function mismatch between a gamma2.2 profile and the actual display on Color compliance

8.4. Display age

Because of the degradation of the materials composing the display, the colors it emits are susceptible to change in both chrominance and luminance over the lifetime of the device. Avanaki et al [24] have studied the effects of both of these variations on the interpretation of digital pathology images.

Figure 16 summarizes the variations observed while testing non-stabilized and stabilized displays of different types. By referring at section 8.1 and Figure 16, it appears that aging is crucial in the grayscale compliance of the calibration.

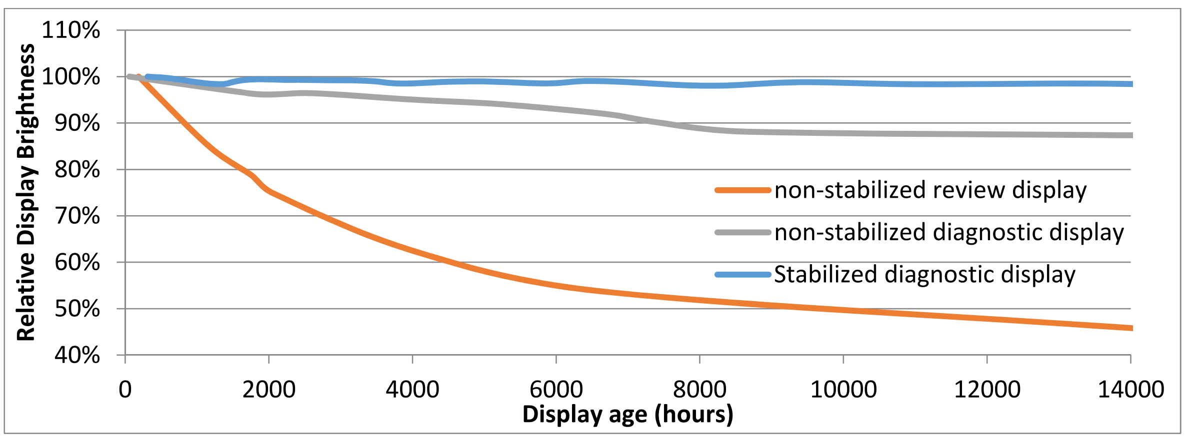


Figure 16: Long term evolution of the maximum luminance of different non-stabilized displays compared to stabilized displays

Please notice that different display systems can have a large difference in terms of performance and stability. While the luminance variation is almost only related to the decreasing efficiency of the backlight (CCFL or LED, where typically LED backlights are more stable over time [25]), and can be compensated by giving more power to the light sources, the color shift is more difficult to anticipate as it depends on the evolution of the optical properties of different layers of diffusers and filters. To evaluate the impact of this color shift on the calibration, a medical display has been characterized over its entire lifetime.

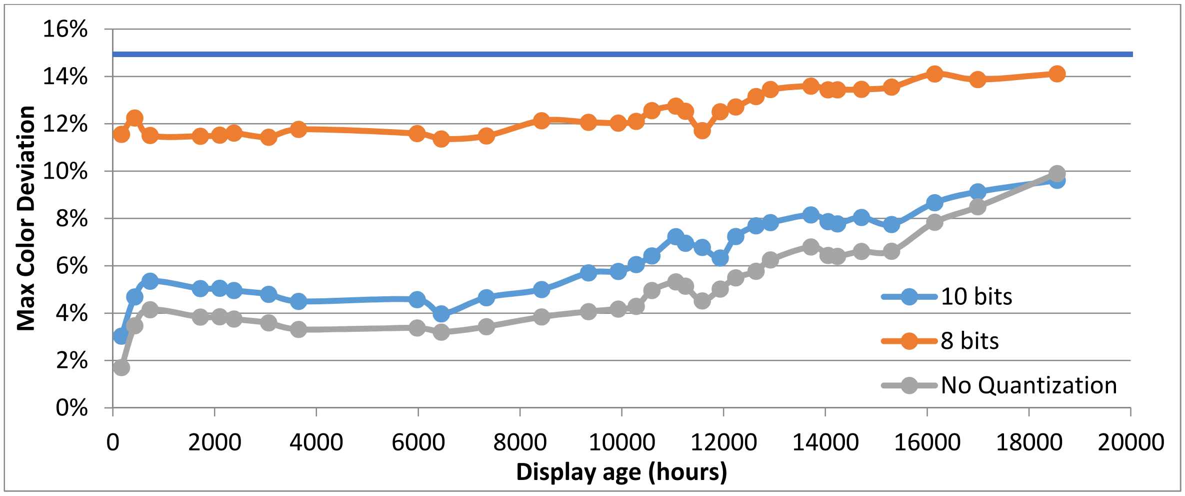


Figure 17: Effect of the aging of a diagnostic display on the Color compliance of a calibration calculated at its production.

A calibration has been calculated based on the first measurements, and evaluated over the complete dataset. Figure 17 presents the results of these tests and shows that aging has a limited impact on the color compliance of the calibration.

8.5. Ambient light

The present study also takes into account ambient light chromaticity by considering a lighting color temperature of $5000K$ (D50).

Impact of the ambient light is modelled as an offset applied on the XYZ output of the display model. This offset is defined as follows with I being the illuminance level (in lux) and 0.01 the reflection coefficient of the display.

$$Y_{amb} = 0.01 * I$$

This offset differs from X and Z channel, according to the proportion of X Y and Z of White D50 $(0.96422,\ 1.0,\ 0.82521)$:

$$X_{amb} = 0.96422 * Y_{amb}$$

$$Z_{amb} = 0.82521 * Y_{amb}$$

8.5.1. Why considering the ambient light?

Ambient light partly reflects on any surface, including displays. The proportion of reflected light mainly depends on the material and reflecting surfaces geometries. This is usually characterized by a Reflection Coefficient associated with the display.

On medical displays, this coefficient is usually higher than on consumer level displays because of the presence of a front glass adding two more interfaces (air-glass and glass-air) on top of the air-panel interface, creating even more reflections.

For this reason we decided to use a reflection coefficient of 0.01 in our simulations. In other words, we consider that the display reflects 1% of the ambient light. Figure

18 shows how the additional light from the reflection can adversely affect the Grayscale part of a calibration if ambient light's effect is not taken into account.

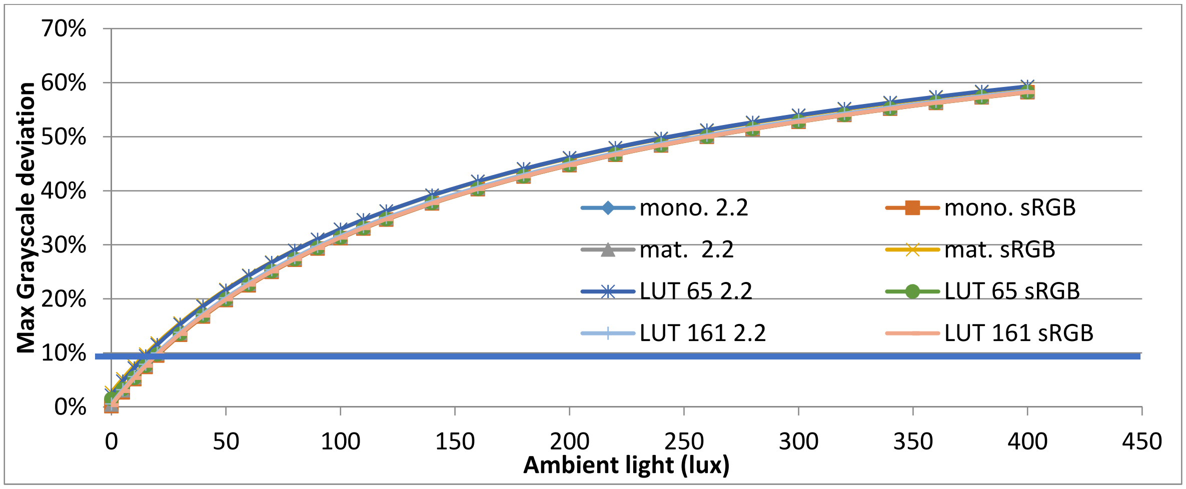


Figure 18: Effect of the ambient light on the Grayscale compliance of a calibration evaluated with different type of display profiles.

It appears on Figure 18 that the effect of ambient light on the calibration is not linear. For this reason, it is important to detail its impact in different environments.

8.5.2. Diagnostic rooms

Diagnostic reading rooms are already used when establishing a diagnostic from quantitative imaging modalities, X-rays and other grayscale modalities where lighting conditions are controlled and illumination maintained low (2 to $10\ lux$ for x-rays, 15 to $60\ lux$ for CT and MR) [19]. In these conditions, knowing precisely the ambient light has its importance. Figure 19 shows how Grayscale compliance varies with the ambient light while profiles were built considering an illumination of $5\ lux$.

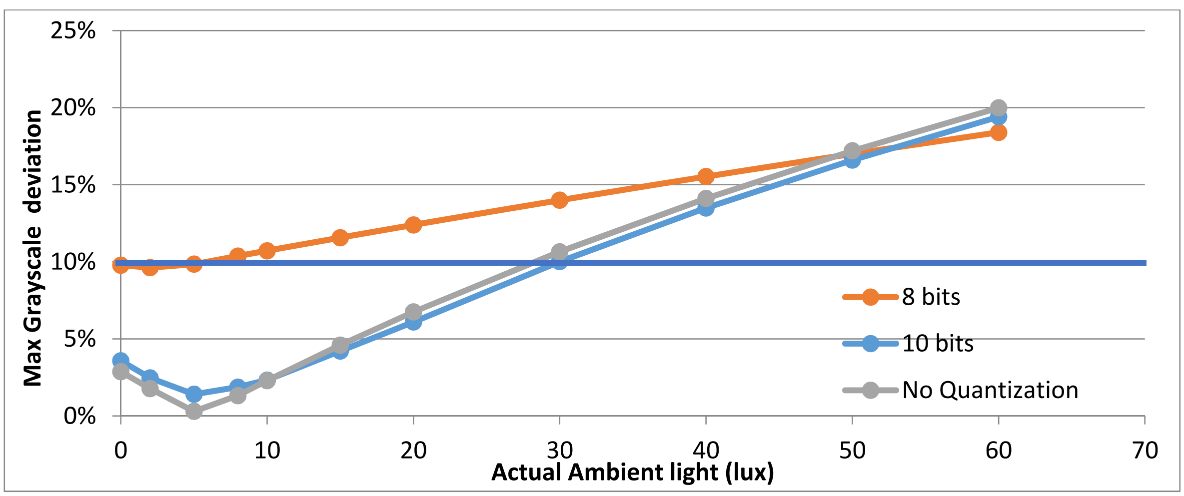


Figure 19: Effect of the ambient light on Grayscale compliance when ICC profiles used for calibration are built for an illumination of $5lux$ with an $18 * 18 * 18$ color LUT

If the lighting conditions are correctly controlled (no windows…) it is possible to assess a correct calibration by having a single estimation of the ambient light at the profile generation time and monitoring ambient light afterwards may not be required.

Figure 20 presents the impact of ambient light on the calibration of colors.

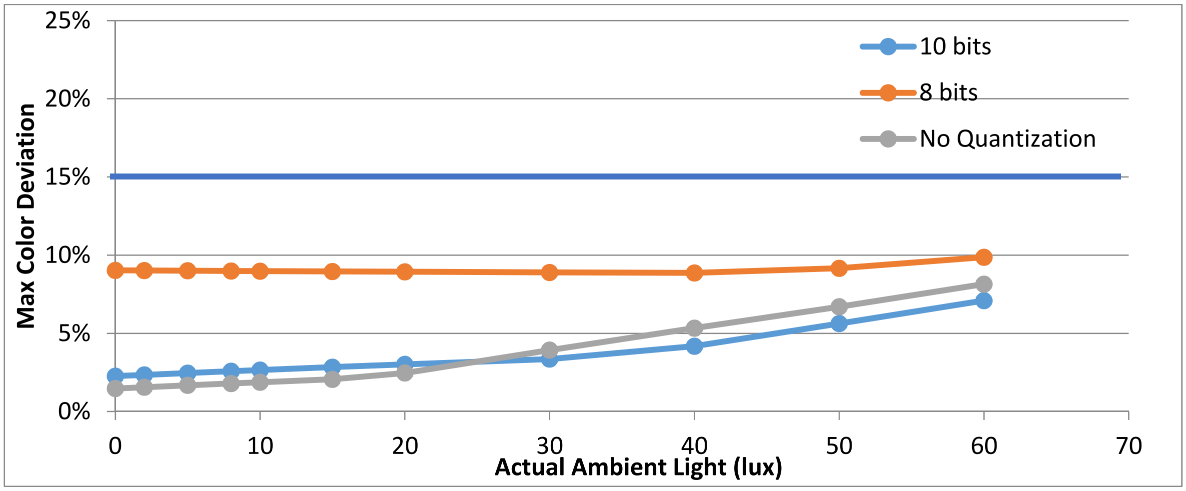


Figure 20: Effect of the ambient light on Color compliance when ICC profiles used for calibration are built for an illumination of $5lux$ with an $18 * 18 * 18$ color LUT

It is clear here that Color calibration does not suffer from an approximate estimation of the ambient light during the calibration process, at least for low illumination levels.

8.5.3. Staff offices

While quantitative imaging modalities are to be examined in dedicated reading rooms with reduced ambient light, pathology diagnostics are usually established in physician offices, where lighting conditions are not controlled and can vary from 50 to $180\ lux$ [19]. In such conditions, it is much more difficult to control the office's illumination as it highly depends on external parameters such as the weather which might abruptly and unpredictably change. It is therefore necessary to continuously measure ambient light and regenerate calibration profiles several times a day.

Figure 22 shows that higher relative variations of ambient light between the profile and the display it describes, have a larger influence on Color compliance for staff offices than they do for diagnostic rooms. For instance, a 50% underestimation of the ambient light led to 9%, 3% and 2% maximal Color deviations in diagnostic rooms, respectively for 8 bit, 10 bit, and floating point precisions. In the case of staff offices, a 45% underestimation already leads to 16%, 10% and 9% Color deviations for 8 bit, 10 bit, and floating point precisions, respectively.

These variations of color calibration compliance remain rather limited in 10 bit systems compared to 8 bit architectures and could be considered as acceptable. However, this is not the case for Grayscale calibration, as misestimating the ambient light by 30% is enough to make the calibration incompliant in an office (Figure 21), where the illumination is susceptible to drastically change throughout the day.

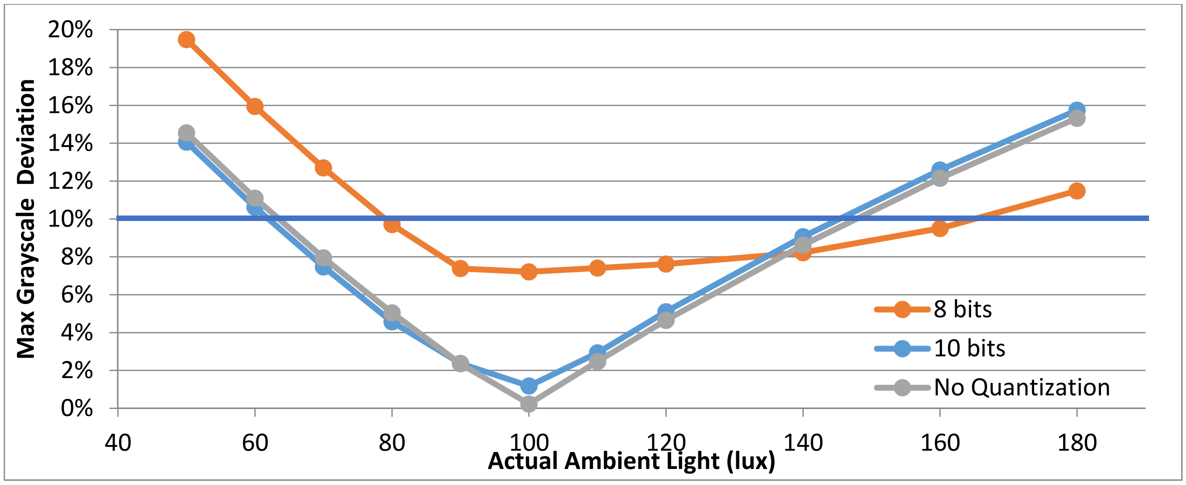


Figure 21: Effect of the ambient light on Grayscale compliance when ICC profiles used for calibration are built for an illumination of $100 lux$ with an $18 * 18 * 18$ color LUT

In order to preserve an accurate Grayscale calibration using this method, ICC profiles would have to be regularly recreated according to the office's ambient light variations. If the presented method were used to obtain the most accurate Grayscale calibration, staff offices would be improper for diagnostic purposes.

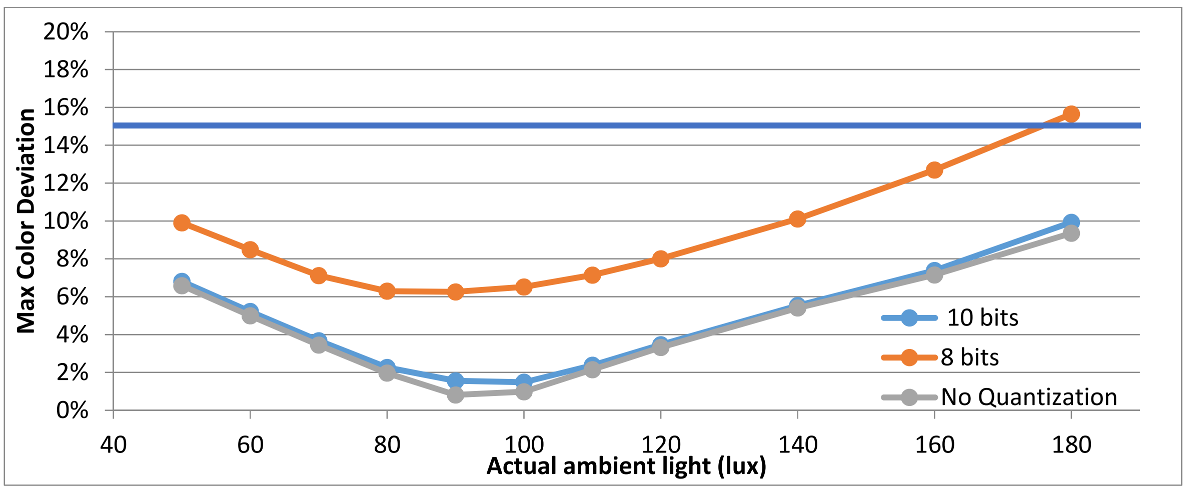


Figure 22: Effect of the ambient light on Color compliance when ICC profiles used for calibration are built for an illumination of $100 lux$ with an $18 * 18 * 18$ color LUT

8.5.4. Operating rooms

Color Management and display calibration is also a concern for surgery. Reviewing scans and radios in an operating room happens and in this case DICOM GSDF calibration must also be respected.

AAPM estimates that operating room illumination usually varies from $300\ lux$ to $400\ lux$. This is quite high and can produce important reflections, especially on displays equipped with a front glass.

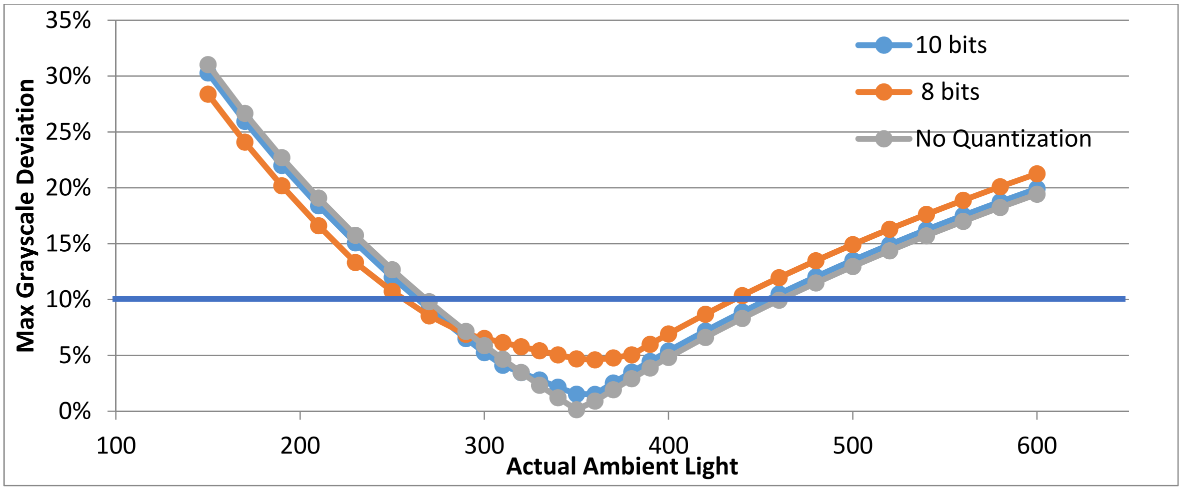


Figure 23: Effect of the ambient light on Grayscale compliance when ICC profiles used for calibration are built for an illumination of $350lux$ with an $18 * 18 * 18$ color LUT

Figure 23 and Figure 24 show that ambient light variation in these conditions has a lower impact on the calibration compliance than it does in diagnostic rooms or staff offices. Relative variation appears to be similar: an 42% underestimation of ambient light by the profile leads to 15.5%, 9.3% and 8.6% Color deviations for 8 bit, 10 bit and floating point precisions.

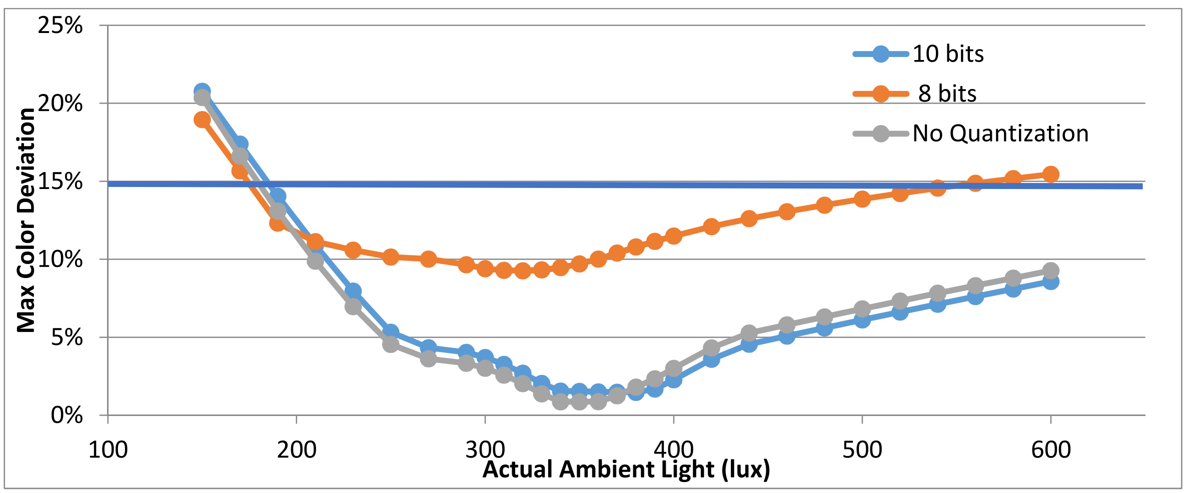


Figure 24: Effect of the ambient light on Color compliance when ICC profiles used for calibration are built for an illumination of $350lux$ with an $18 * 18 * 18$ color LUT

However, influence of absolute variation has far less impact than in other use cases, but the presence of very powerful directional light sources can be a concern for the quality of the color calibration.

9. Recommendations

The following recommendations allow for visualization of medical content on color display systems, whereby DICOM GSDF images, pseudo color images and color accurate images can all be presented effectively on the same display.

For display systems which already have embedded DICOM GSDF / CSDF calibration and stabilization, it is recommended to use the same profile for source and destination.

For non-calibrated displays, the following recommendations are provided with the goal to stay within 10% tolerance of the Grayscale target and within 15% tolerance of the Color target:

- System configuration:
    - Only use ICC profiles that have been created for the specific display. Generic profiles do not offer sufficient accuracy, even if the display can be set to a reference state.
    - Every time a display setting is changed (e.g. display luminance or contrast settings), new source and destination profiles need to be created and used. Use at least $10\ bit$ connections from application to software when a most accurate calibration is needed, since $8\ bit$ ones are clearly not sufficient for these use cases. It is to be noted that $10\ bit$ connections can be achieved by means of $8\ bit$ display controllers that use appropriate dithering algorithms.
    - Display luminance and contrast should be stabilized to the value that was used when creating the profile, since luminance and contrast deviations result into reduced calibration accuracy (See Figure 9 and Figure 12).
    - If the luminance cannot be stabilized, a “warming-up” period should be respected before the display can be used. A period of $2\ hours$ (see Figure 11) is recommended, but this time may be reduced if the stability and warm-up of the display is known and reproducible.

- ICC Profile and CMM:
    - Both source and destination profiles must take the ambient light into account.
    - Both source and destination profiles should be LUT based profiles using XYZ color space as PCS as described in section A.3. As explained in section 5 and B.2.4, it is also possible to use DeviceLink profiles.
    - For DICOM GSDF calibration of grayscale display, the use of monochrome profile is possible, and recommended.
    - For CSDF calibration, the CLUT of the source profile (describing the calibration) must have a size of at least $11 * 11 * 11$ points to be compliant (see Figure 30), but using at least $33 * 33 * 33$ points is recommended for a more accurate calibration. The display profile can be matrix-based, but we recommend using a more accurate LUT-based profile as depicted in section A.3.
    - Special attention must be given to PCS-To-Device conversion of the Black point. This is critical to achieve an acceptable calibration. See section 8.2.

- Calibration process:
    - The calibration compliance must be verified <u>at least</u> every 50 calendar days since typical display behavior changes over time as Figure 16 shows. If the compliance test fails, the whole calibration process has to be repeated. This means renewing display measurements and regenerating the display profile

based on these measurements. More frequent measurements are possible and could guide determining when recalibration is needed.
- o Ambient light must be stable. Otherwise, the calibration process must be repeated every time the ambient light conditions change (see Figure 21).

10. Patent statement

The International Color Consortium (ICC) draws attention to the fact that it is claimed that compliance with section 3 (on CSDF) of this White Paper can involve the use of a patent. ICC takes no position concerning the evidence, validity and scope of this patent right. The holder of this patent right has assured the ICC that he is willing to negotiate licences under reasonable and non-discriminatory terms and conditions with applicants throughout the world. In this respect, the statement of the holder of this patent right is registered with ICC. Information may be obtained from:

Barco N.V.
Healthcare Department
Beneluxpark 21, 8500 Kortrijk
Belgium

Whether specific implementations of a particular recommendation listed in the white paper are covered by the subject matter disclosed in these patents/patent applications is to be evaluated by any interested party.
The Patent Holder does not claim that these patent/patent applications are required for the implementation of a particular recommendation listed in the white paper. In case of doubt, any entity is welcome to contact Barco patent department for further clarifications on the matter.

Annex A. Detailed structure of the ICC profiles

Version 4.3 of the ICC specification [26] makes it possible to use different architectures to build ICC profiles.

A.1. Monochrome profiles

ICC defined monochrome profiles to describe grayscale devices. As DICOM GSDF is a calibration of grayscale systems, it makes sense to use monochrome DICOM profiles, and let the CMM return RGB triplets where R = G = B.

The monochrome profiles are very simple. They also present the advantage of requiring the same tags whether they are input, output or display profiles. However, as their name suggest they are only suitable for grayscale devices.

Apart from the copyright and description tags, there are:

- Media White Point Tag:

This tag contains the White point of the device, normalized and chromatically adapted to the PCS illuminant. For a display profile, this is equivalent to the PCS illuminant itself. The capture device White point of an input profile is “the encoding maximum White for the capture encoding”.

- Gray TRC Tag:

This tag contains the Gray Tone Reproduction Curve, representing the conversion from the device Digital Driving Level to the achromatic channel of the PCS. This curve can be composed of up to 4096 points, or being a predefined parametric curve.

The display profile then contains an accurate description of the “native” display function, while the input profile describes the exact DICOM target for the given display luminance and contrast.

- Chromatic adaptation Tag:

This tag contains a linear Bradford chromatic adaptation matrix corresponding to the adaptation from the actual illuminant to the PCS adopted White Chromaticity as represented by the equation hereafter.

$$\begin{bmatrix} X_{PCS} \\ Y_{PCS} \\ Z_{PCS} \end{bmatrix} = \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & a_{22} \end{bmatrix} * \begin{bmatrix} X_{SRC} \\ Y_{SRC} \\ Z_{SRC} \end{bmatrix}$$

In addition to those required tags, the luminance tag was included because of the necessity to ensure the input profile contains a DICOM calibration corresponding to the luminance of the display, as the calibration depends on both the maximum and minimum luminance of the device.

A.2. Three-component Matrix-TRC-based profiles

This profile architecture assumes the conversion from device color space to PCS is a simple linear combination of their respective channels as shown on Figure 25. It can be understood as a set of three tone curves modeling the non-linearity of the response of each input channel.

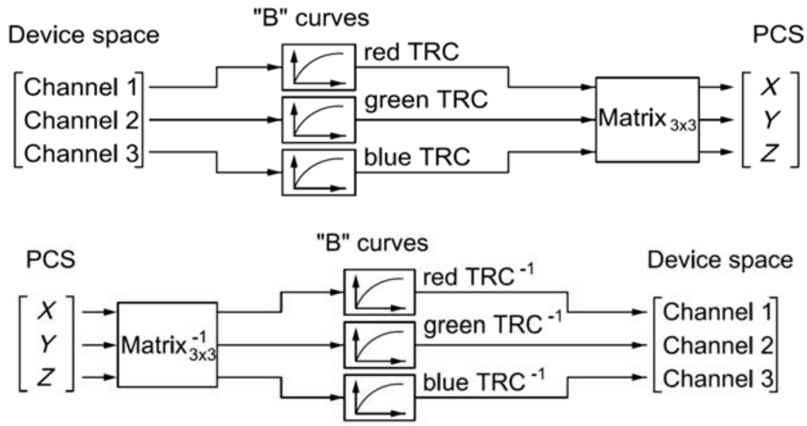


Figure 25: Model of conversion from device space to PCS and from PCS to device space as it is used in display matrix-based profiles.

Those curves are contained in the three Red, Green and Blue TRC Tags. Similarly the tags Red, Green and Blue Column matrix represent the three columns of the conversion matrix. They also are the values of the 3 primaries of the device expressed in the PCS.

Matrix-based profiles perform very well in describing theoretical display standards or models as the ones we use in this study, but are unable to describe the internal constraints of a real LCD display. For instance, this architecture cannot deal with cross-talk in between the sub-pixels of a Liquid Crystal Panel. On an actual display, the Red TRC depends on the levels of Green and Blue as shown on Figure 26, and the simplicity of this model does not allow this phenomenon to be taken into account.

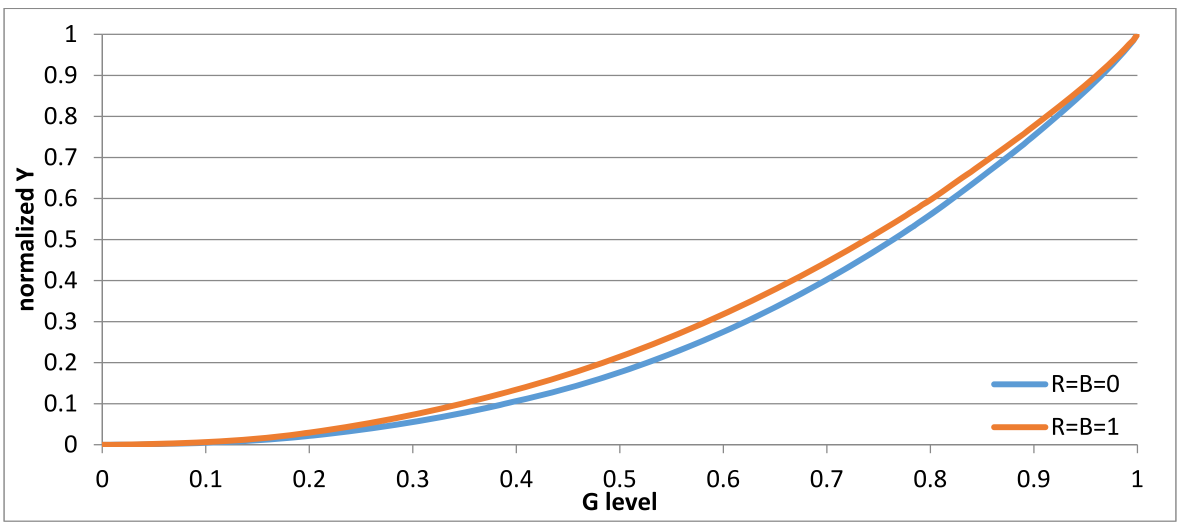


Figure 26: Comparison of the Green Luminance curve of a diagnostic color display, when Red and Blue Chanel are both set to 0% and 100%. The curves have been normalized for better readability. The horizontal axis represents the Green ddl, and the vertical axis is the normalized luminance (Y-Ymin)/(Ymax-Ymin).

This approximation may be acceptable depending on the use-case.

For the same reason, this profile architecture cannot be used to represent the CSDF calibration as the light output of the display for a given RGB triplet is no longer directly proportional to the light of the three primaries in this case.

For instance, CIEDE2000 calibration of the Black-to-Green sweep can be represented by the Green TRC tag, but this TRC is also applied on the Magenta-to-White sweep whereas its calibration is completely different. Figure 27 illustrates this by presenting an example of how the Green channel is impacted by the CSDF on different parallel color sweeps in the RGB cube.

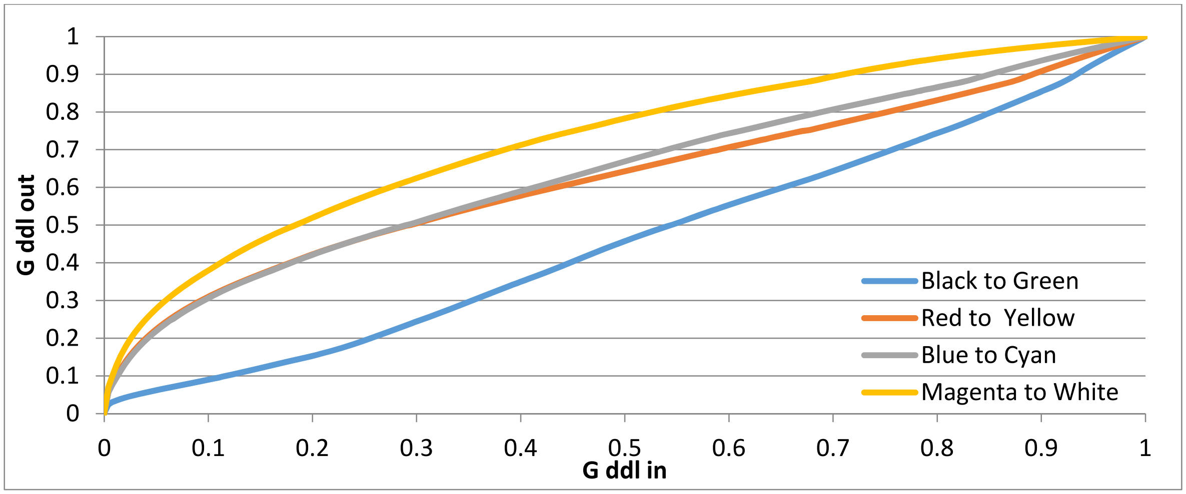


Figure 27: Green-to-Green 1D LUT for CSDF calibration of different color sweeps. All these sweeps are defined in RGB by a changing G value from 0 to 1 and with R and B constant.

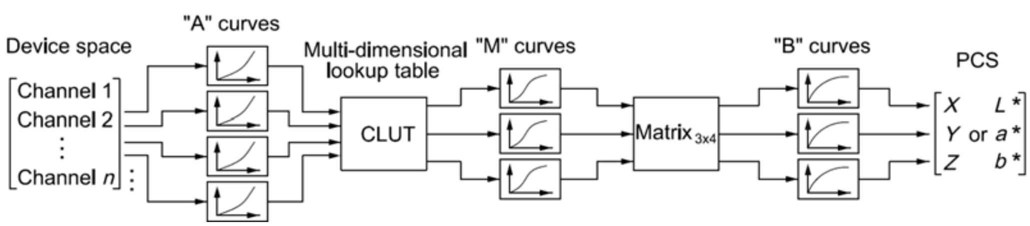
A.3. N-component LUT-based profiles

LUT based profiles are far more complex than the previously described architecture (see Figure 28). LUT profiles have N-dimension tables with entries for every combination (or a range large enough to allow interpolation) of input values and their corresponding PCS values. There is one table per direction (PCS-To-Device and Device-To-PCS) and per rendering intent (Perceptual, Saturation and Colorimetric).

Not all six tables are required for every profile. Firstly, having a single rendering intent is enough to build a profile, and only display profiles require the two directions of conversion. Nevertheless, this is enough to make these profiles larger, but also more accurate in their description of the color behavior of a device.

Several other elements can be combined with the LUT to make the device characterization even more accurate.

A, B and M curves behave just like TRC described in section A.2. The CLUT is organized as an i-dimensional array with a variable number of grid points in each dimension, where i is the number of input channels in the transform. Each grid point value is an o-integer array, where o is the number of output channels.



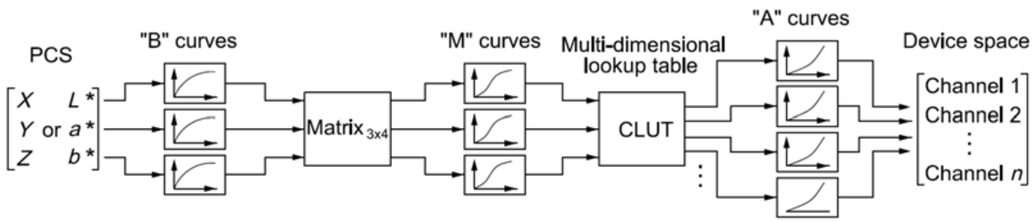


Figure 28: Device-to-PCS and PCS-to-Device conversion workflows for LUT based profiles. The different elements arround the Color LUT (CLUT) can be used to create a nonlinear repartition of the input values of the LUT, or set to identity.

During our experiments, both i and o were equal to 3. We used only cubic LUTs (same size on every dimension) and only a few sizes have been tested.

Profiles were constructed using the XYZ PCS.

- A-curves are unused in both Device-to-PCS and PCS-to-Device conversions and were thus set to be identity tone curves.
- CLUT stages are used for RGB-to-RGB conversions. The tables may contain individual point corrections to make the profile more faithful to the display it represents.
- M-curves are used to apply inverse RGB companding in Device-to-PCS conversion and RGB companding in PCS-to-Device conversion. This handles the RGB-XYZ non linearity and makes RGB linear.
- The matrix is used to finish the linear RGB-to-XYZ conversion and thus contains RGB reference primaries as XYZ values, written in column order. For encoding reasons, these values are all divided by 2.
- B-curves, as A-curves, were unused in both tags and were set to identity curves.

In the case of the aforementioned display models, identity 3D CLUTs were used in the CLUT stage since there is no physical display color point correction to apply on them. For CSDF profiles, a 3D color correction LUT is calculated and encompassed in the CLUT stage of the pipeline. CSDF profiles are always built from a DICOM model to ensure a good DICOM calibration by storing the GSDF in the profile M-curves.

A.4. DeviceLink profiles

In the classical workflow presented in sectionA.2, the color space of the input device is transformed to the color space of the output device via the device-independent color space (PCS) by connecting two different profiles (a source profile and a destination profile). A DeviceLink profile is a special kind of ICC profile that converts the color space of the input device directly into the color space of the output device without any intermediate step.

DeviceLink profiles contain a single table similar to the one presented at the top of Figure 28. The DeviceLink profile can be built similarly to the N-component LUT based profile; except that every elements related to PCS can be removed. In the end, only the RGB-To-RGB conversion is preserved in the CLUT element, and if necessary the 1-dimensional RGB-To-RGB LUT ensuring GSDF calibration can be stored in the B-Curves.

DeviceLink profiles present as main drawback a lack of flexibility. Indeed each profile corresponds to a single very precise situation. While the classical workflow allows for example to use the same source profile when a display's internal calibration state is changed from sRGB to DICOM, and only update the display profile, Here it is necessary to update the DeviceLink profile, and so to recalculate the RGB-To-RGB calibration LUT.

Annex B. Application of the calibration method and results

The calibration method has been tested with all of the profile models presented in Annex A for both source and destination profiles, though one architecture is designed for monochromatic devices and can only be used for GSDF calibration purpose.

The different profiles do not represent physical displays, but follow some simple models:

- Gamma 1.8
- Gamma 2.2
- Gamma 3.5
- sRGB
- DICOM GSDF

For all of them, a Luminance of $Y_{max} = 600\ cd/m^2$ and a contrast of $1000:1$ ($Y_{min} = 0.6\ cd/m^2$) are considered since these are typical values. Color coordinates of White, Red, Green and Blue primaries follow the sRGB standard [27] as summarized in Table 1.

Color	Y (cd/m^2)	x	y
White	600	0.3127	0.329
Black	0.60	0.3127	0.329
Red	128.08	0.64	0.33
Green	429.26	0.30	0.60
Blue	43.86	0.15	0.06

Table 1: Color coordinates of Black Point, White Point and Primary Colors common to the different display models used during this study.

This choice was made in order to ensure a good reproducibility of the experiments, but the method has been tested with physical displays in section 0.

Figure 29 shows the resulting luminance response curves of those models.

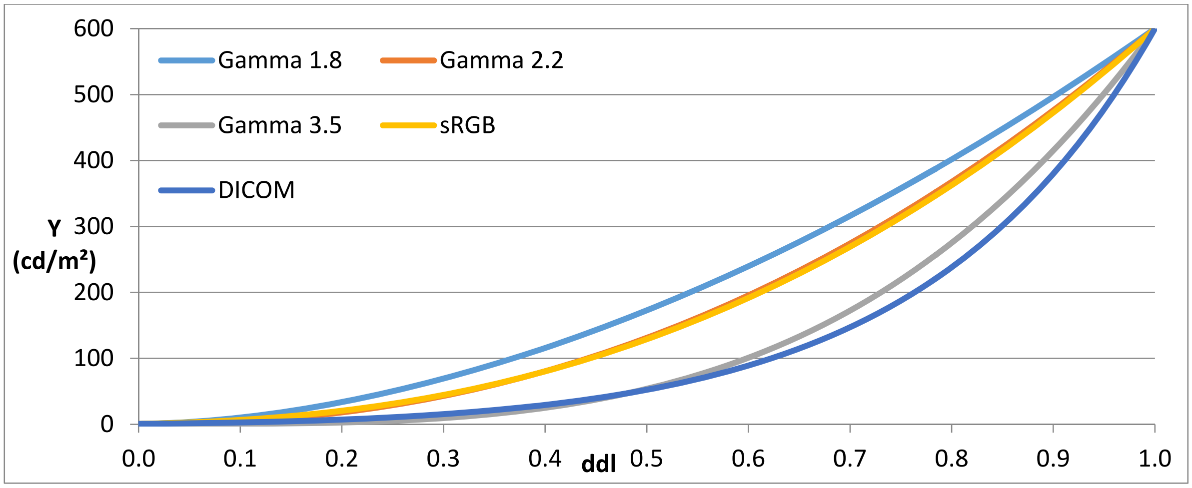


Figure 29: Luminance curves of the different display models considered during this study.

B.1. Profile quality assessment

Created profiles were tested in following the methods presented in section 6 to assess their quality before using them for calibration purpose.

B.1.1. Fidelity test

B.1.1.1. Monochrome profiles results

In case of a DICOM GSDF calibration for pure grayscale modalities, the use of grayscale monitors and monochrome profiles is possible. For this reason, monochrome profiles are also evaluated with the difference that the previously presented input test sample is reduced to RGB triplets with $R = G = B$. Results are presented in Table 2 and reveal a very good accuracy of monochrome profiles.

Display Model	Average Profile Fidelity	Worst Profile Fidelity
sRGB	0.0182519	0.031728
Gamma 3.5	0.0145193	0.031728
Gamma 2.2	0.0180893	0.031728
Gamma 1.8	0.0195657	0.031728
DICOM	0.0149173	0.031728

Table 2: Monochrome profiles fidelity results as a function of the display model (CIEDE2000).

B.1.1.2. Matrix-TRC profiles results

In this paragraph we estimate the accuracy of Matrix based profiles. Results summarized in Table 3 are perfectly acceptable for every profile except CSDF.

This confirms the assumption of section A.2 about the impossibility to describe complex color systems with such profiles. According to the results presented here this profile architecture will not be considered anymore in the next sections.

Display Model	Average Profile Fidelity	Worst Profile Fidelity
sRGB	0.0068204	0.034245
Gamma 3.5	0.005613	0.034245
Gamma 2.2	0.00671118	0.034245
Gamma 1.8	0.00734149	0.034245
DICOM	0.00562321	0.034245
CSDF	7.09157	24.9025

Table 3: Matrix-TRC profiles fidelity as a function of the display model (CIEDE2000).

B.1.1.3. Reference LUT-based profiles results

Since CLUTs that are used for creating profiles based on display models are completely linear, their sizes do not impact the results of the tests for reference profiles. This parameter will thus not be used in the interpretation of the reference results presented in Table 4.

Display Model	Average Profile Fidelity	Worst Profile Fidelity
sRGB	0.0205527	0.0784756
Gamma 3.5	0.0175035	0.0701783
Gamma 2.2	0.0205474	0.0793607
Gamma 1.8	0.0217619	0.0844731
DICOM	0.0172126	0.0631233

Table 4: Reference Profiles Fidelity results as a function of the display model (CIEDE2000).

For these profiles, induced color differences are far below the perceptual limit and thus even the maximal measured difference would not be perceivable in real conditions.

B.1.1.4. CSDF LUT-based profiles results

Fidelity test results for CSDF profiles are presented in Table 5.

CLUT size	Average Model Fidelity	Worst Model Fidelity
11	0.0199861	0.106663
18	0.0199088	0.111355
33	0.0199288	0.111493
65	0.0199219	0.107764

Table 5: CSDF Profiles Fidelity results as a function of the CLUT size (CIEDE2000).

As it is observable in the table, the chosen CLUT size does influence Fidelity test results for CSDF profiles. However, this influence appears to be relatively minor. Similarly to the reference profiles, color differences induced by profile generation are not perceptually significant.

B.1.2. Roundtrip test

B.1.2.1. Monochrome profiles results

The results are presented in Table 6 and show a maximal error of 0.213944 CIEDE2000 and a mean error of 0.0126455 CIEDE2000 for Gamma 3.5 profiles, while the others present values comparable to the reference LUT-based profiles presented in section 0.

Profile	Roundtrip Mean	Roundtrip Max
sRGB	0.00170167	0.00031189
Gamma 3.5	0.213944	0.0126455
Gamma 2.2	0.00548937	0.000646326
Gamma 1.8	0.000117576	0.00113504
DICOM	0.0014838	0.000130997

Table 6: Monochrome profiles roundtrip results with luminance of 600 cd/m² and contrast ratio of 1000 (CIEDE2000)

B.1.2.2. Display LUT-based profiles results

According to [28], when using transforms containing CLUTs larger than $2 * 2 * 2$, accuracy requirements stipulate that "*round tripping color differences in CIELAB ΔE^*_{ab} should be less than 1 mean and less than 3 maximum*". However, reference profiles use identity RGB-to-RGB CLUTs, thus comparable to a $2 * 2 * 2$ CLUT. In that case, still according to [28], *"color differences should be less than 0.5 mean and less than 1 max"*.

Results for reference profiles are presented in Table 7. Roundtrip tests for created profiles show maximal errors of 0.0963472 CIEDE2000 and mean errors below 0.00389718 CIEDE2000. These are far below the thresholds given above.

Profile	Roundtrip Mean	Roundtrip Max
sRGB	0.00152663	0.022604
Gamma 3.5	0.00389718	0.0963472
Gamma 2.2	0.000834971	0.00580855
Gamma 1.8	0.000662107	0.00549094
DICOM	0.0015788	0.0280541

Table 7: Reference profiles roundtrip results with luminance of 600 cd/m² and contrast ratio of 1000 (CIEDE2000)

B.1.2.3. CSDF LUT-based profiles results

The chosen CLUT size is of major influence on the roundtrip quality of these profiles. Extreme values in function of the LUT size are shown in Table 8.

The maximal observed color differences can thus be very high for small accuracy profiles such as the ones based on $11 * 11 * 11$ CLUTs. Conversely, the mean color differences, while still being greatly influenced by the LUT size parameter, still show good overall results.

CLUT size	Roundtrip Mean	Roundtrip Max
11	0.349	6.921
18	0.176	4.938
33	0.052	3.901
38	0.038	2.970
65	0.015	0.820

Table 8: LUT-based CSDF profiles maximal roundtrip errors relatively to the chosen LUT size (CIEDE2000). A color scale is applied on the values to emphasize the results compared to the thresholds defined in [28].

The high maximal values obtained here are due to the difficulty to properly reverse the CSDF 3D LUT, especially near White. Due to the fact that the CIEDE2000 calibration makes colors brighter than they are with DICOM GSDF, the LUT tends to group the colors close to White as it is observable on Figure 27 where the Green-To-Green 1D LUT corresponding to the Magenta-to-White color scale extracted from the 3D LUT appears. This effect can be compensated by using a bigger CLUT.

B.2. Calibration quality assessment

Results were assessed by connecting generated CSDF profiles with their corresponding display profile (regarding luminance & contrast) using the Colorimetric intent.

Compliance of the calibration on grayscale and colors were assessed separately, since grayscale and colors do not share the same metrics.

To evaluate the influence of the bit depth, a quantization step is applied on the output of the ICC framework, before the resulting RGB triplet is fed to the display model.

Nowadays, a vast majority of display systems supports 8bits only input signals, but some high end devices propose to use 10 bits signals. For medical applications, 8bits does not guarantee the best image quality [29] while using more than 10 bits is not necessary as the human visual system is only able to distinguish up to 900 shades of gray, even on high luminance displays [30].

However being able to observe images with 10bits precision requires the complete video chain to be compatible. Of course the display itself must support 10 bits input signals (typically provided by DisplayPort connection), but also the workstation (the Graphic Process Unit and its driver) has to support 10 bits output, the software used to read the images must be able to render 10 bits, and the image itself must be encoded on 10bits (or more). If only one of those components is limited to 8 bits, the final result would be viewed with 8 bits quantization. Windows 7 and above, OSX 10.11 (El Capitan), and Linux are all able to support 10 bits color output with compatible hardware.

B.2.1. Grayscale calibration quality assessment

B.2.1.1. Monochrome calibration

Here are presented the results of simulations obtained when calibrating grayscale displays to DICOM GSDF by using monochrome profiles. Results are summarized in Table 9.

Destination Profile	Grayscale compliance	Grayscale compliance 10 bits	Grayscale Compliance 8 bits
sRGB	0.145%	1.522%	8.326%
Gamma 3.5	0.138%	1.190%	6.823%
Gamma 2.2	0.145%	1.661%	8.375%
Gamma 1.8	0.454%	2.163%	12.636%

Table 9: Grayscale compliance results for monochrome profiles generated with luminance of $600\ cd/m^2$ and a contrast ratio of $1000{:}1$

This method seems to be reliable, at least for 10 bits systems. 8 bits quantization is not necessarily non-compliant except when calibrating a Gamma 1.8 (12.636%). In that case the deviation is quite high and the conditions are not optimum.

B.2.1.2. Color calibration

Here colors displays are calibrated to CSDF by using LUT-based profiles. Grayscale calibration compliance within CSDF is assessed according to the DICOM standard, as explained in section 7.1. Results are summarized in Table 10.

Destination Profile	Grayscale compliance	Grayscale compliance 10 bits	Grayscale compliance 8 bits
sRGB	2.455%	1.522%	8.326%
Gamma 3.5	0.442%	1.187%	6.827%
Gamma 2.2	0.573%	1.649%	8.352%
Gamma 1.8	0.786%	2.227%	12.772%
DICOM	2.483%	1.976%	0.079%

Table 10: Grayscale compliance results for profiles generated with luminance of $600\ cd/m^2$ and a contrast ratio of $1000{:}1$

Without applying quantization, results show good compliance scores. In this case, deviation amplitude appears to be inversely correlated to the display model's gamma. 10 bit compliance results are much better than the 8 bit ones with deviation ranging from 1.187% for the Gamma 3.5 reference to 2.227% for the Gamma 1.8 profile, which showed the worst compliance score with 8 bit quantization.

It is interesting to note that the quantization does not necessarily make the simulated calibration worse. This is due to the fact that the ICC framework introduces some errors in the process (These errors are estimated during the roundtrip test). Quantization, by rounding the output of the ICC framework can correct or reduce the Framework imprecision. This is especially visible when applying the color calibration on a DICOM compliant display model.

Grayscale compliance test relies on 18 gray levels, evenly spread from Black-to-White:

$$(R,G,B)_{in} = \left(\frac{i}{17}, \frac{i}{17}, \frac{i}{17}\right) \quad i \in [\![0;17]\!]$$

Transformed by the profiles connection, these triplets become:

$$(R,G,B)_{out} = \left(\frac{i}{17} + \varepsilon_r, \frac{i}{17} + \varepsilon_g, \frac{i}{17} + \varepsilon_b\right) \quad i \in [\![0;17]\!]$$

Then quantization is applied:

$$(R,G,B)_{quantized} = \left(\frac{round\left(\left(\frac{i}{17}+\varepsilon_r\right)*q\right)}{q}, \frac{round\left(\left(\frac{i}{17}+\varepsilon_g\right)*q\right)}{q}, \frac{round\left(\left(\frac{i}{17}+\varepsilon_b\right)*q\right)}{q}\right)$$

$$i \in [\![0;17]\!]$$

$$q = 255 \text{ } or \text{ } 1023$$

If $-1/2q \leq \varepsilon_r < 1/2q$ (and similarly for ε_g and ε_b), the imprecision introduced by the connection is simply erased by the quantization.

This is exactly what happens when applying the calibration on a DICOM display. The DICOM GSDF grayscale described in the source profile is exactly identical to the one in the destination profile. Connecting those two profiles and transforming gray levels with this connection results in a very accurate roundtrip, introducing minor imprecisions on each RGB triplet along the grayscale. Imprecisions are within the range $-1/2q \leq \varepsilon < 1/2q$ resulting in $(R,G,B)_{8bit} = (R,G,B)_{in}$. As the display is already DICOM calibrated, an evenly spread set of input RGB results in a perfect theoretical DICOM GSDF compliance. However, this is strictly specific to the definition of the quality assessment method, and does not reflect the final image quality which depends on all the existing levels and not only 18 of them.

B.2.2. Color calibration quality assessment

Color compliance is assessed with an arbitrary tolerance of 15% deviation for $6 * 18$ color samples. The simulated results of the proposed calibration method are summarized in Table 11.

Destination Profile	Source profile	Color max deviation	Color max deviation	Color max deviation

	LUT size		10 bits	8 bits
sRGB	11	10.593%	12.201%	15.334%
	18	1.832%	2.624%	9.030%
	33	2.838%	2.901%	9.465%
	65	1.940%	2.624%	9.030%
Gamma 3.5	11	10.564%	10.397%	10.244%
	18	1.804%	2.142%	7.447%
	33	2.825%	3.318%	7.447%
	65	1.940%	2.142%	7.447%
Gamma 2.2	11	10.547%	12.050%	21.534%
	18	1.770%	1.675%	7.570%
	33	2.800%	2.049%	7.570%
	65	1.929%	1.677%	7.570%
Gamma 1.8	11	10.573%	13.768%	19.495%
	18	1.792%	3.220%	14.757%
	33	2.790%	4.030%	14.757%
	65	1.934%	3.220%	14.757%
DICOM	11	10.587%	10.701%	13.129%
	18	1.804%	2.030%	7.306%
	33	2.837%	3.215%	7.306%
	65	1.951%	1.850%	7.306%

Table 11: Color compliance obtained by using different display models and different size of CLUT in the source profile

B.2.2.1. Without quantization

Without quantization, Color compliance scores are below the tolerance limit for all of the tested LUT sizes. However, compliance scores of the $11 * 11 * 11$ profiles clearly demonstrate that this particular size is unsuited when accurate calibration is needed, with Color deviation spanning from 10.547% to 10.593%.

Since compliance is evaluated with 18 points samples along RGB sweeps, the $18 * 18 * 18$ results depicts the accuracy of Color calibration with a minimal influence of interpolation (see Figure 30). With the tested set of display models and $18 * 18 * 18$ CLUTs, Color deviation ranges from 1.770% to 1.832%.

Results for $33 * 33 * 33$ and $65 * 65 * 65$ CLUT sizes illustrate Color compliance scores on a larger grid, thus encompassing interpolation induced error but giving a better idea of the accuracy of the calibration on the whole gamut. That is why observed deviations for these two sizes are superior to the ones observed with $18 * 18 * 18$. For $33 * 33 * 33$ CLUT size, deviation scores show a minimal value of 2.790% and a maximal value of 2.838%. Using more entries in the CLUT enhances the accuracy of the calibration, as depicted by the $65 * 65 * 65$ results that ranges from 1.929% minimum to 1.951% maximum deviations.

B.2.2.2. With 8 bit quantization

With 8 bit quantization applied, results obtained with $11 * 11 * 11$ LUT based profiles show a very high maximal color deviation from CSDF color targets, ranging from 10.244% to 21.534%. This corroborates the assessment that $11 * 11 * 11$ entries LUTs are unsuited for accurate calibration targets, especially on a 8 bit system.

Results for all the other profiles show very similar maximum deviation values: around 9% for sRGB reference profiles and 7.5% for Gamma & DICOM profiles.

Since LUT size does not seem to influence these values much, it may be deducible that this is the maximal accuracy obtainable with the presented architecture on 8 bits.

B.2.2.3. With 10 bit quantization

As for Grayscale results, using 10 bit quantization instead of a 8 bit quantization effectively reduce the maximal observed deviation. Beneficial influence of a larger bit depth is not consequently significant when using $11 * 11 * 11$ LUTs, with the lowest maximal deviation being 10.397% (which is even higher than the 8 bit minimum deviation) and an overall maximum deviation of 13.768%. That makes the results Color compliant for the whole set of tested profiles, despite giving globally poor results.

Results for $18 * 18 * 18$ and $65 * 65 * 65$ LUTs show very similar Color compliance results. For $18 * 18 * 18$ LUTs, maximum deviations span from 1.675% to 3.220%. For $65 * 65 * 65$ LUTs, it ranges from 1.677% to the same 3.220% maximum.

Since the influence of interpolation errors is reduced with higher LUT sizes (because interpolated values are closest), it is minimal with $65 * 65 * 65$ LUT size. The surprisingly good results obtained with $18 * 18 * 18$ are explained in section B.2.2.4.

With $33 * 33 * 33$ LUTs, maximum deviations show a minimal value of 2.049% and a maximal value of 4.0300%. This specific size presents relative deviations which are much more impacted by interpolation than its $18 * 18 * 18$ and $65 * 65 * 65$ counterparts. However, $33 * 33 * 33$ produces still very good compliance scores.

B.2.2.4. Conclusion

The chosen CSDF Color validation method induces different interpolation errors for every LUT-size. Since there are arbitrarily 18 samples used on RGB sweeps and because a 3D LUT containing these samples as internal nodes will return non-interpolated values, all LUT sizes having $N * 18 - (N - 1)$ side points will return better results than other chosen sizes would (with N being an integer factor) without guaranteeing a better calibration accuracy on daily use.

Figure 30 depicts observed relative Color compliance of a profile connection using a CSDF profile as source and a destination profile as a function of CLUT size.

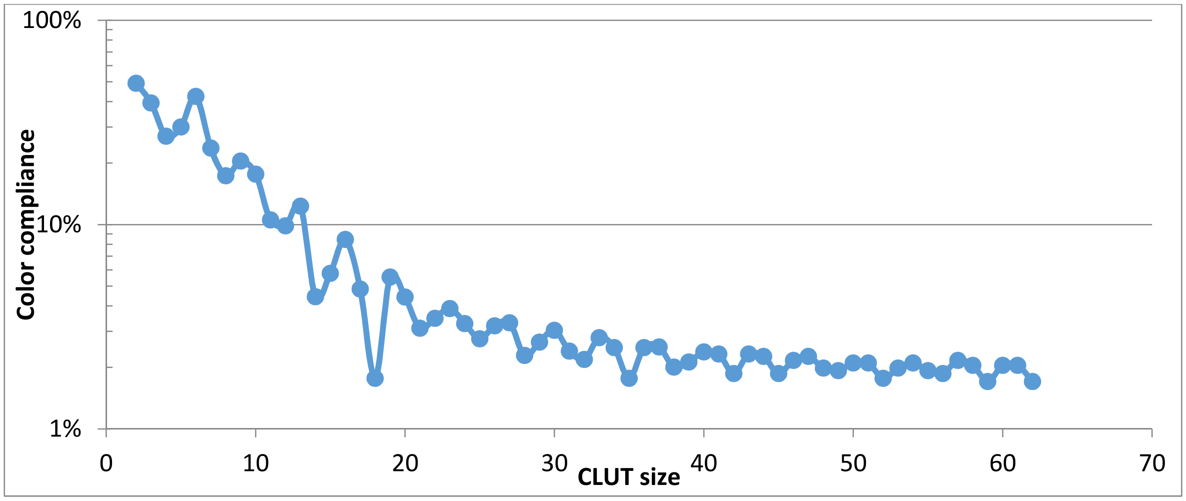


Figure 30: Observed Color compliance as a function of the source profile CLUT size when used with $600\ cd/m^2$ with contrast ratio of 1000:1 in logarithmic scale on vertical axes.

It is clearly observed that results obtained with a number of CLUT side points matching the aforementioned equation are the most compliant. Therefore minimal deviation is observed for 18, 35 and 52 CLUT side points when tested on 18 points. Nevertheless, we can observe from the trend of the graph, but also from the profile validations that a LUT size of at least 32 is advised.

B.2.3. Calibration smoothness

The resulting smoothness corresponding to the different display models and calibrations mentioned in the present document are presented in Table 12.

Calibration	Average	Standard deviation	Minimum	Maximum
sRGB	0.1842	0.1497	0.0267	1.0289
Gamma 3.5	0.2548	0.1722	0.052	0.9767
Gamma 2.2	0.2175	0.1894	0.0245	1.3444
Gamma 1.8	0.2122	0.2091	0.0252	2.3454
GSDF	0.2202	0.1522	0.0457	1.3098
CSDF	0.1955	0.125	0.0429	1.1279

Table 12: Smoothness of different display models, without quantization or ICC color transform

One can notice that CSDF calibration presents the lowest average value (0.1955) after sRGB (0.1842), revealing a pretty smooth calibration, but also the best standard deviation (0.125) which means the smoothness is more homogeneous over the entire gamut.

The smoothness of the CSDF calibration applied on different displays has also been studied and is presented in Table 13.

Display model	Average	Standard deviation	Minimum	Maximum
sRGB	0.187	0.1499	0.027	1.0392
Gamma 3.5	0.1958	0.1249	0.0432	1.1333
Gamma 2.2	0.1958	0.1249	0.0436	1.1306
Gamma 1.8	0.1958	0.1249	0.0435	1.131
GSDF	0.1961	0.1248	0.0433	1.1291

Table 13: Smoothness of different display models calibrated to CSDF by using ICC profiles, without quantization

Applying the calibration on a sRGB display seems to result in a better average smoothness (0.187) compared to Gamma models (0.1958) and GSDF (0.1961). And while it presents the highest standard deviation announcing homogeneity (i.e. some parts of the color gamut will present sharper transitions), CSDF calibration of sRGB display has the lowest maximum score and is - in terms of smoothness - the best configuration.

Table 14 focuses on the effects on smoothness of quantization and of the size of the CLUT in the profiles used perform for the color transform. Only the calibration of the sRGB display is presented in the table, but similar trends have been observed with the other models.

CLUT Size	Average smoothness	Average smoothness 10 bits	Average smoothness 8 bits
11	0.1941	0.2204	0.4338
18	0.1907	0.2136	0.4334
33	0.187	0.208	0.4342
65	0.1855	0.203	0.4268

Table 14: Effect of the quantization and the profiles CLUT size on the average smoothness of a color calibration applied on a sRGB display model.

Quantization has a huge impact on the final calibration smoothness. While quantizing to 10 bits slightly deteriorate the smoothness of a system, passing from 10 bits to 8 bits more than doubles the average smoothness value. This makes other consideration such as CLUT size, but also the display model on which to apply the calibration of far less importance.

B.2.4. Using DeviceLink profiles

Using DeviceLink profiles to calibrate a system is possible as explained in section A.4. This method has also been evaluated here and results for grayscale and color compliance are

presented respectively in Table 15 and Table 15: Grayscale compliance results with DeviceLink profiles generated with luminance of $600\ cd/m^2$ and a contrast ratio of $1000:1$

.

Display type	Grayscale compliance	Grayscale compliance 10 bits	Grayscale Compliance 8 bits
sRGB	0.110%	1.522%	8.326%
Gamma 3.5	0.110%	1.186%	6.828%
Gamma 2.2	0.094%	1.662%	8.375%
Gamma 1.8	0.103%	3.760%	12.772%
DICOM	0.106%	1.441%	0.079%

Table 15: Grayscale compliance results with DeviceLink profiles generated with luminance of $600\ cd/m^2$ and a contrast ratio of $1000:1$

Display type	Devicelink CLUT size	Color max deviation	Color max deviation 10 bits	Color max deviation 8 bits
sRGB	11	10.686%	12.200%	15.334%
	18	1.644%	2.624%	9.031%
	33	2.717%	2.901%	9.465%
	65	1.826%	2.624%	9.031%
Gamma 3.5	11	10.681%	10.397%	10.245%
	18	1.662%	2.142%	7.447%
	33	2.702%	3.318%	7.447%
	65	1.806%	2.142%	7.447%
Gamma 2.2	11	10.705%	10.416%	21.510%
	18	1.642%	2.044%	7.588%
	33	2.714%	2.044%	7.588%
	65	1.805%	2.044%	7.588%
Gamma 1.8	11	10.705%	13.761%	19.496%
	18	1.645%	3.226%	14.757%
	33	2.723%	4.030%	14.757%
	65	1.803%	3.226%	14.757%
DICOM	11	10.663%	10.593%	13.130%
	18	1.651%	2.027%	7.306%
	33	2.693%	3.537%	7.306%
	65	1.805%	2.027%	7.306%

Table 16: Color compliance obtained by using different display models and different size of CLUT in the DeviceLink profile

Using DeviceLink profiles ensures a better conservation of the Black Point resulting in a better Grayscale compliance without quantization. However, when taking the quantization into account, this benefit compared to the classical framework is lost making them essentially similar in performance.

Regarding the presented results, there is no reason to recommend the use of one system or another. Decision to use DeviceLink profile or not is at the user discretion.

B.3. Experimental validation

B.3.1. Medical grade display

The described method was experimentally validated on a medical grade display set to three different display functions: gamma 2.2, gamma 1.8, and DICOM GSDF. Measurements were performed using a Konica Minolta CA-210 on an evenly spread set of 18 color points as it was previously described.

The exact values of the display primary colors, but also Black and White are given in Table 17.

Color	Y(cd/m^2)	x	y
White	462	0.305	0.334
Black	0.46	0.262	0.273
Red	82.11	0.643	0.327
Green	324.9	0.319	0.622
Blue	48.15	0.150	0.081

Table 17: Measured luminance and chromaticity values used for the display model

Based on these measurements, display models used for experimental validation have been generated. These models thus have a luminance of $462 cd/m^2$ and a contrast ratio of $1004:1$. It is important to stress that these models were generated based on a limited number of measurements (see Table 17). Therefore it is to be expected that the generated models will not perfectly match the true display behavior (see section 8 for a more detailed description on the effects of this non-perfect modeling). These models then were used to calibrate the display systems.

Simulation results are presented in Table 18 and corresponding measurement results are presented in Table 19.

Profile	Color max deviation 10 bits	Color max deviation 8 bits	Grayscale max deviation 10 bits	Grayscale max deviation 8 bits
Gamma 2.2	3.023%	6.564%	3.252%	4.750%
Gamma 1.8	3.078%	14.031%	3.025%	11.068%
DICOM	2.680%	8.179%	3.132%	5.266%

Table 18: Simulated calibration compliance on the 3 tested configurations with CLUTs of $33 * 33 * 33$ points

As could be expected, experimental results show larger deviations than the simulation results. This is normal since in case of the experimental results the actual display behavior was measured while assessing calibration compliance, while in case of simulation results the assumption is that the display correctly follows the theoretical display model.

Profile	Color max deviation 10 bits	Color max deviation 8 bits	Grayscale max deviation 10 bits	Grayscale max deviation 8 bits
Gamma 2.2	6.051%	6.662%	3.152%	4.706%
Gamma 1.8	6.123%	9.418%	1.561%	10.325%
DICOM	6.243%	6.577%	2.344%	5.714%

Table 19: Measured calibration compliance on the 3 tested displays with CLUTs of $33 * 33 * 33$ points

Especially for $10\, bits$ color signals, the experimental results show larger Color deviations than the simulation results for a $10\, bits$ color signal. On the other hand, measurement results obtained with a $8\, bit$ per channel system are much closer to the simulated results. The reason is that the inaccuracies introduced by the quantization when using $8\, bits$ channels are larger than the inaccuracies due to non-perfect display modeling.

B.3.2. Consumer off-the-shelf display

The same protocol as above has been repeated on a consumer off-the-shelf (COTS) display. This one was set to gamma 2.2, and only supported 8 bit input. Here again, every measurements have been done with a Konica Minolta CA-210 after having respected a warm-up period of 3 hours. The measured values of the display Black, White and primary colors are given in Table 20 and its contrast is $1114{:}\,1$.

Color	Y(cd/m^2)	x	y
White	200.5	0.313	0.3262
Black	0.18	0.263	0.250
Red	46.17	0.632	0.334
Green	137.9	0.312	0.643
Blue	17.23	0.148	0.065

Table 20: COTS display measured luminance and chromaticity values used for the display model

The ICC profiles created for these experiments were LUT-based profiles with $33 * 33 * 33$ CLUT in the case of CSDF profile. The display profiles have a purely linear CLUT, and no attempt to improve the profile fidelity by introducing some corrections in there, as it is suggested in section A.3.

At a first attempt of calibrating this display to CSDF, we trusted the preset and generated a display profile having perfect gamma 2.2 TRC. This ends up with a very bad calibration compliance presented in the first row of Table 21 and confirming the results of section 8.3.

A second calibration has been executed, this time after having measured 256 gray levels on the display to model more accurately the real display TRC. Here the observed Grayscale compliance is in line with the simulation, and the observed Color compliance is even slightly better than expected as presented in the second row of Table 21.

Display TRC	Simulated		Measured	
	Color max deviation	Grayscale max deviation	Color max deviation	Grayscale max deviation
Assumed Gamma 2.2	6.568%	6.117%	12.646%	18.848%
Measured Gray TRC	12.059%	6.446%	9.528%	6.879%

Table 21: Simulated and Measured calibration compliances on the COTS display with CLUTs of 33*33*33 points with 8bit quantization

Observations also match pretty well to the predictions presented in Table 10 and Table 11 for 8 bits systems with similar display functions and CLUT sizes.

These measurements suggest that calibrating a display to with the presented method using only a LUT-based profile without correction or matrix-based profile (which is equivalent when following the recommendations of sections A.2 and A.3) is possible. A $33 * 33 * 33$ CLUT for the CSDF profile is enough to obtain a compliant calibration, but the observed deviation are already pretty high, and the calibration would have to be repeated regularly to maintain the display calibrated.

It also appears that using 8bits system is possible as it produces compliance results just below the rejection threshold. However, the chances of passing this threshold because of some variations of the usage conditions are high.

Annex C. Relationship to dRGB

Michael Flynn (Henry Ford Health System) proposed a new color space called medical RGB (dRGB) which tries to merge the DICOM GSDF and sRGB color space[4]. dRGB is also one of the ICC MIWG projects.

dRGB is not only a color space but also a complete framework giving specifications for medical display performances and calibration. It also includes the use of ICC profiles to perform dRGB to PCS conversions.

When linking this to the Color Space draft [31], being worked out in the context of AAPM TG196; one can say that the present document covers use cases 1A, 1C and 2C presented on Figure 31.

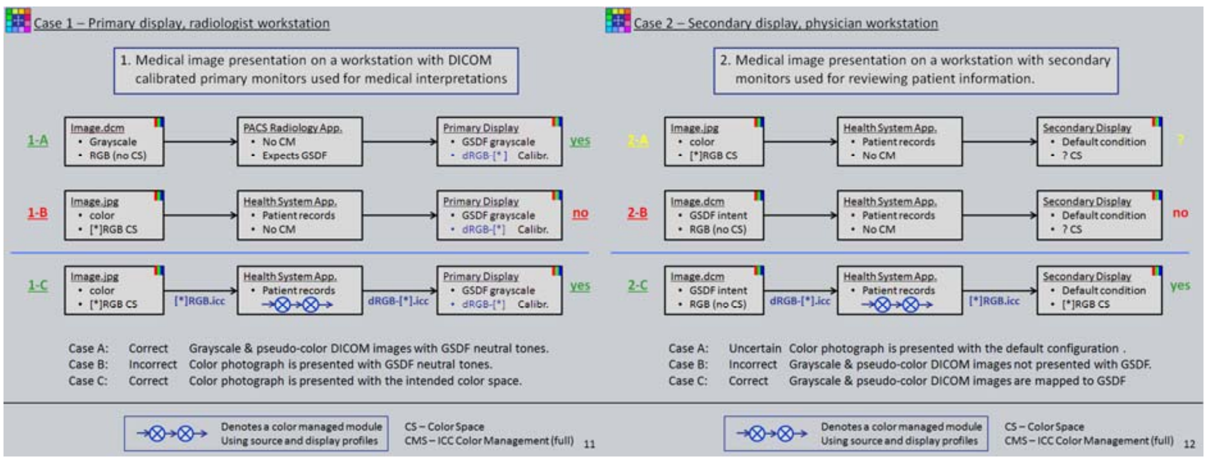


Figure 31: Grayscale and color medical images as described by Michael Flynn in [31].

---

4 http://www.color.org/groups/medical/Flynn.pdf

Bibliography

[1] NEMA, Digital imaging and communications in medicine (DICOM), part 14: Grayscale Standard Display Function, vol. PS 3.14, National Electrical Manufacturers Association, 2001.

[2] Department of Health, Queensland Govenrment, AU, “Medical imaging monitors specification guidelines,” July 2010. [Online]. Available: https://www.health.qld.gov.au/qhcss/documents/monitor-specifications.pdf. [Accessed 10 April 2016].

[3] A. A. Marghoob for the International Skin Imaging Collaboration Melanoma Project Working Groups, “Standards in Dermatologic Imaging,” *JAMA Dermatology,* vol. 151, no. 18, pp. 819-821, 2015.

[4] L. Silverstein, S. Hashmi, K. Lang and E. Krupinski, “Paradigm for achieving color-reproduction accuracy in LCDs for medical imaging,” *Journal of the Society for Information Display,* vol. 20, pp. 53-62, January 2012.

[5] W. C. Revie, M. Shires, P. Jackson, D. Brettle, R. A Cochrane and D. A Treanor, “Color Management in Digital Pathology,” *Analytical Cellular Pathology,* 2014.

[6] E. Krupinski, L. Silverstein, S. Hashmi, A. Graham, R. Weinstein and H. Roehrig, “Impact of Color Calibration on Breast Biopsy Whole Slide Image Interpretation Accuracy and Efficiency,” in *Breast Imaging*, vol. 8539, H. Fujita, T. Hara and C. Muramatsu, Eds., Springer International Publishing, 2014, pp. 744-748.

[7] E. Krupinski, “The importance of perception research in medical imaging,” *Radiation medecine,* vol. 18, no. 6, pp. 329-334, 2000.

[8] P. Barten, “Physical model for the contrast sensitivity of the human eye,” in *SPIE*, 1992.

[9] P. Barten, “Spatio-temporal model for the contrast sensitivity of the human eye and its temporal aspects,” in *SPIE*, 1993.

[10] K. A. Fetterly, H. R. Blume, M. J. Flynn and E. Samei, “Introduction to grayscale calibration and related aspects of medical imaging grade liquid crystal displays.,” *Journal of digital imaging,* vol. 21, no. 2, pp. 193-207, June 2008.

[11] E. A. Krupinski and H. Roehrig, “The influence of a perceptually linearized display on observer performance and visual search.,” *Academic radiology,* vol. 7, no. 1, pp. 8-13, 2000.

[12] A. Badano, C. Revie, A. Casertano, W.-C. Cheng, P. Green, T. Kimpe, E. Krupinski, C. Sisson, S. Skrøvseth, D. Treanor, P. Boynton, D. Clunie, M. Flynn, T. Heki, S. Hewitt, H. Homma, A. Masia, T. Matsui, B. Nagy, M. Nishibori, J. Penczek, T. Schopf, Y. Yagi and H. Yokoi, “Consistency and Standardization of Color in Medical Imaging: a Consensus Report,” *Journal of Digital Imaging,* vol. 28, no. 1, pp. 41-52, February 2015.

[13] DICOM, *DICOM supplement 100: Color Softcopy Presentation State Storage SOP Classes,* 2005.

[14] T. Kimpe, J. Rostang, A. Avanaki, K. Espig, A. Xthona, I. Cocuranu, A. Parwani and L. Pantanowitz, “Does the choice of display system influence perception and visibility of clinically relevant features in digital pathology images?,” in *Proc. SPIE 9041, Medical Imaging 2014: Digital Pathology 904109*, 2014.

[15] S. Zabala-Travers, M. Choi, W.-C. Cheng and A. Badano, “Effect of color visualization and display hardware on the visual assessment of pseudocolor medical images,” *Medical Physics,* vol. 42, no. 6, pp. 2942-2954, June 2015.

[16] CIE, “Improvement to industrial colour-difference evaluation.,” 2001.

[17] I. Lissner and P. Urban, “Toward a Unified Color Space for Perception-Based Image Processing,” *IEEE Transactions on Image Processing,* vol. 21, no. 3, pp. 1153-1168, March 2012.

[18] T. Kimpe, J. Rostang, G. Van Hoey and A. Xthona, “WE-D-204-04: Color Standard Display Function (CSDF): A Proposed Extension of DICOM GSDF,” *Medical Physics,* vol. 42, no. 6, pp. 3670-3671, 2015.

[19] E. Samei, A. Badano, D. Chakraborty, K. Compton, C. Cornelius, K. Corrigan, M. Flynn, B. Hemminger, N. Hangiandreou, J. Johnson, D. Moxley-Stevens, W. Pavlicek, H. Roehrig, L. Rutz, S. Shepard, R. Uzenoff, J. Wang and C. Willis, Assessment of display performance for medical imaging systems, American Association of Physicists in Medicine, 2005.

[20] IEC 62563-1:2009, “Medical electrical equipment, Medical image display systems, Part 1: Evaluation methods,” IEC, Geneva, 2009.

[21] P. J. Green, “A smoothness metric for colour transforms,” in *Color Imaging XIII: Processing, Hardcopy, and Applications*, 2008.

[22] A. Avanaki, K. Espig, A. Xthona and T. Kimpe, “WE-D-9A-07: SRGB Displays: A Good Choice for Medical Color Images?,” *Medical Physics,* vol. 41, no. 6, pp. 501-501, jun 2014.

[23] N. Odlum, G. Spalla, N. Van Assche, B. Vandenberghe, R. Jacobs, M. Quirynen and C. Marchessoux, “Preliminary display comparison for dental diagnostic applications,” in *SPIE Medical imaging*, 2012.

[24] A. R. Avanaki, K. S. Espig, S. Sawhney, L. Pantanowitz, A. V. Parwani, A. Xthona and T. R. L. Kimpe, “Aging display’s effect on interpretation of digital pathology slides,” in *SPIE Medical imaging*, 2015.

[25] D. S. Hirschorn, E. A. Krupinski and M. J. Flynn, “PACS displays: how to select the right display technology,” *Journal of the American College of Radiology,* vol. 11, no. 12, pp. 1270-1276, December 2014.

[26] ICC, *Specification ICC.1:2010 (Profile version 4.3.0.0),* 2010.

[27] IEC61966-2-1:1999, “Multimedia systems and equipment - Colour measurement and management - Part 2-1: Colour management - Default RGB colour space - sRGB.,” IEC, 1999.

[28] P. Green, Color Management : Understanding and Using ICC Profiles - Phil Green, Michael Kriss, P. Green and M. Kriss, Eds., Wiley, 2010, p. 314.

[29] T. Hiwasa, J. Morishita, S. Hatanaka, M. Ohki, F. Toyofuku and Y. Higashida, “Need for liquid–crystal display monitors having the capability of rendering higher than 8 bits in display-bit depth,” *Radiological Physics and Technology,* vol. 2, no. 1, pp. 104-111, 2009.

[30] T. Kimpe and T. Tuytschaever, “Increasing the Number of Gray Shades in Medical Display Systems, How Much is Enough?,” *Journal of Digital Imaging,* vol. 20, no. 4, pp. 422-432, 2006.

[31] C. Revie, S. Smiley, P. Green, H. Reinl, M. Barbieri, J. Chang, M. Flynn, M. Marks, C. Weerasinghe, B. Hulsken, J. P. Van de Capelle, R. Horn, C. Bai, M. Derhak, V. Lovell, M. Maria, M. Mahy, K. Xiao, H. Fukuda, M. Chang, J. Pescatore, J. Vogh, W. Li, T. Lianza, P.-C. Hung, M. Yamaguchi, A. Masia, D. Orf and P. Bautista, “Meeting minutes international Color Consortium MIWG,” Boston, 2014.

[32] Y. J. Kim, Y. Bang and H.-K. Choh, “Gradient approach to quantify the gradation smoothness for output media,” *Journal of Electronic Imaging,* vol. 19, p. 1012, 2010.

Table of Content

1. Introduction **1**

2. Calibration requirements, standards and guidelines for medical displays **1**

3. Different uses of color in medical imaging **2**

3.1. Absolute color reproduction for medical images 2

3.2. Perceptually linear visualization of medical images 3

3.2.1. DICOM Grayscale Standard Display Function (GSDF) 3

3.2.2. Color Standard Display Function (CSDF) 3

4. Proposed calibration method to achieve DICOM GSDF and CSDF calibration based on the ICC framework **5**

5. Creation of the profiles **7**

6. Profile quality assessment methods **8**

6.1. Fidelity test 8

6.2. Roundtrip test 9

7. Calibration quality assessment methods **9**

7.1. How to evaluate the quality of the grayscale calibration 10

7.2. How to evaluate the quality of the color calibration 10

7.3. Calibration smoothness 12

8. Impact of inaccurate profiling **13**

8.1. Display luminance 13

8.2. Display contrast 15

8.3. Display function (gamma) 17

8.4. Display age 18

8.5. Ambient light 19

8.5.1. Why considering the ambient light? 19

8.5.2. Diagnostic rooms 20

8.5.3. Staff offices 21

8.5.4. Operating rooms 22

9. Recommendations **23**

Annex A. Detailed structure of the ICC profiles **26**

A.1. Monochrome profiles 26

A.2. Three-component Matrix-TRC-based profiles 26

A.3. N-component LUT-based profiles 28
A.4. DeviceLink profiles 29
Annex B. Application of the calibration method and results 31
B.1. Profile quality assessment 32
B.1.1. Fidelity test 32
B.1.2. Roundtrip test 34
B.2. Calibration quality assessment 35
B.2.1. Grayscale calibration quality assessment 36
B.2.2. Color calibration quality assessment 37
B.2.3. Calibration smoothness 40
B.2.4. Using DeviceLink profiles 41
B.3. Experimental validation 43
B.3.1. Medical grade display 43
B.3.2. Consumer off-the-shelf display 44
Annex C. Relationship to dRGB 46