White Paper
An Introduction to Color for Medical Imaging

Geert Carrein
Johan Rostang
Bastian Piepers
Albert Xthona
What’s inside?

- A primer on color
- Human vision and color perception
- From DICOM gray JND to color JND
- Challenges for medical color displays
- The relevance of color for medical imaging
  - Absolute color
  - Prior segmentation to a certain color
  - Consistency in calculated colors
  - Image fusion of multiple modalities
Abstract

Over the past several years, we have seen a shift in medical displays from monochrome to color image presentation. During the early days of this transition, color was primarily used to offer the radiologist a much richer user interface or support the use of color annotations.

Since then the use of color in medical imaging has dramatically increased. Many of today’s emerging medical applications utilize color to denote a specific meaning, such as in ultrasound, PET-CT fusion and microscopy.

Based on currently available imaging technologies, it is clear that additional display requirements are necessary to ensure the accurate, and consistent, display of specific colors over time and space.

This paper will briefly introduce the theory of color and describe some of the challenges and considerations for medical color imaging with respect to diagnostic displays.
A primer on color
1. A primer on color

Visible light is generally described as a range of electromagnetic waves which can be perceived by the human eye. As seen in Figure 1, these visible electromagnetic waves occupy only a narrow part of the total electromagnetic spectrum. Each wave has a particular wavelength; the shorter the wavelength, the more energy it carries.

The visible blue waves contain the highest amount of energy, and feature the shortest wavelength (around 400 nm). The visible red waves produce the lowest amount of energy. They have a wavelength approximately 730 nm.

The human eye is most sensitive to green light. More precisely, the human eye is most sensitive to waves with a wavelength of about 555 nm.

Each individual wavelength within the visible range represents a pure color. When this wave enters the human eye, it causes a specific, pure color sensation. If the full spectrum of visible waves simultaneously hits the human eye, we experience this as white light. White light can be described as the composition of all individual wavelengths combined. Alternatively, certain objects are capable of splitting up white light back into its composition of individual pure spectral colors. Examples are the prism (Figure 2) and the rainbow (Figure 3).

A measurement device called a spectroradiometer is used to measure the spectral power distribution of the aforementioned visible spectrum. This spectrum can be used to calculate additional properties of a particular color.
Human vision and color perception
2. Human vision and color perception

2.1 Color Perception by the Human Eye

As mentioned in Section 1, the human eye is sensitive to a very narrow range of wavelengths within the electromagnetic spectrum, ranging from 380 nm to 780 nm. The retina of the human eye contains a large number of cones and rods. The cones are responsible for our color vision during daylight, while the rods contribute to the intensity of color we perceive and are also responsible for our night vision.

Most humans have cones that are sensitive to three different parts of the visible spectrum (Figure 4): long, medium and short wavelengths. These correspond to red, green and blue color sensations.

![Retinal Response for the Cones and Rods of the Human Eye](image)

The individual or simultaneous stimulation of these cones by light, in combination with the processing and interpretation of the human brain, forms the basis of our color vision.

**Note:** when light is absorbed by the cones, all individual wavelengths are absorbed, resulting in one single output signal to the optic nerve. During this process, all the spectral information of the original stimuli is lost. In other words, a similar output of the cones can be achieved by an infinite number of spectra. The human eye will perceive all these spectra as equal colors. This is a very important characteristic of the human eye as it allows us to create a similar color sensation with other light sources.

A well-known example of this is the image that you see on your display or television. The three light sources that stimulate your eye are the tiny red, green and blue pixels, which create the colors that you perceive. For instance, when you simultaneously increase the intensity of the red and the green pixels on the monitor, your brain will perceive this as yellow. By varying these amounts, you can match the color of an apple, although the emitted spectrum of the pixels is actually not at all equal to the spectrum reflected by an apple in real life.

Grassmann’s Law for additive color demonstrates that all colors inside a triangle formed by the three primary light sources can be formed by a linear combination of those three primaries.
2.2 Introduction to Colorimetry: The CIE 1931 XYZ Coordinate System

In section 2.1 Color Perception by the Human Eye, it was explained that human color vision is basically a three-dimensional phenomenon depending on the spectral sensitivity function of the three different cones inside the human eye.

Because it is difficult to precisely measure the sensitivity function of the different cones, the CIE (Commission Internationale de l’Eclairage or International Commission on Illumination) defined a set of so-called “spectral color-matching functions” in 1931 that characterizes all the color sensations that a “standard” human observer can experience. These standard observer color-matching functions $x(\lambda)$, $y(\lambda)$, $z(\lambda)$ are represented in Figure 5.

The spectral stimulation of each of these functions using a light source with a spectral power distribution $I(\lambda)$, either direct or reflected from an object, results in one single $X$, $Y$, $Z$.

$$X = K \int_{380}^{780} I(\lambda) x(\lambda) \, d\lambda$$
$$Y = K \int_{380}^{780} I(\lambda) y(\lambda) \, d\lambda$$
$$Z = K \int_{380}^{780} I(\lambda) z(\lambda) \, d\lambda$$

Each XYZ color takes a unique position in a three-dimensional space defined by the X, Y and Z axes, and all colors that are possible in nature will fill up a volume in that color space. The CIE XYZ coordinate system defines $Y$ as the luminance. The XZ plane will contain all possible chromaticities at that luminance.

![Figure 5: The CIE Standard Observer Color Matching Functions](image-url)
Since it is difficult to work with the XYZ color space, the CIE also defined a two-dimensional projection of this space (Figure 6), the so-called xy chromaticity diagram (indicated by the small letters x and y).

The chromaticity diagram is easily derived from the three-dimensional XYZ color space by normalizing and projecting $x = X/(X+Y+Z)$ and $y = Y/(X+Y+Z)$. Due to this normalization, it is not necessary to keep the third coordinate $z = Z/(X+Y+Z)$, as the sum of the three always equals 1 or $x+y+z = 1$.

The two-dimensional projection of this xy chromaticity diagram is depicted in Figure 7. As you can see, the chromaticity plane xy has a horseshoe shape.

The colors located on the edge of the shape are the pure spectral colors, ranging from 380 nm to 700 nm as displayed in the colors of the rainbow or as the colors are split by a prism (see Section 1).

The straight line connecting the red and the blue is called the purple line. It does not exist in nature as a pure spectral color. The colors on this line are actually the result of stimulating the human eye simultaneously with red and blue. White is found in the center of this diagram.
From DICOM gray JND to color JND
3. From DICOM gray JND to color JND

3.1 Just Noticeable Difference (JND)

In medical imaging, it is very important that digital images are displayed in a consistent way on every monitor. Without standardization, it is highly likely that a digital image will look good on one display, producing an excellent diagnostic value, while appearing differently on another display, reducing the diagnostic value of that same image. In the case of medical grayscale images, one could implement a standard of maintaining equal luminance steps \( Y \) (see 2.2) between every two succeeding gray levels. However, experimental data has proven this to be an unsuccessful approach.

Medical grayscale imaging is based on the concept of Just Noticeable Difference or "JND." A JND is the smallest difference in luminance between two gray levels that the average observer can perceive on the display system.

As depicted in Table 1, a Just Noticeable Difference (JND) can be as small as \( \Delta Y = 0.0047 \, \text{cd/m}^2 \) or as large as 24 \, \text{cd/m}^2 along the Y axis. From Table 1, we can infer that the absolute nature of the XYZ system is far from perceptually uniform, and that equal luminance steps between every two succeeding gray levels is not a good standard of measurement.

<table>
<thead>
<tr>
<th>JND</th>
<th>( L [\text{cd/m}^2] )</th>
<th>JND</th>
<th>( L [\text{cd/m}^2] )</th>
<th>JND</th>
<th>( L [\text{cd/m}^2] )</th>
<th>JND</th>
<th>( L [\text{cd/m}^2] )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0500</td>
<td>2</td>
<td>0.0547</td>
<td>3</td>
<td>0.0594</td>
<td>4</td>
<td>0.0643</td>
</tr>
<tr>
<td>5</td>
<td>0.0696</td>
<td>6</td>
<td>0.0750</td>
<td>7</td>
<td>0.0807</td>
<td>8</td>
<td>0.0866</td>
</tr>
<tr>
<td>9</td>
<td>0.0927</td>
<td>10</td>
<td>0.0991</td>
<td>11</td>
<td>0.1056</td>
<td>12</td>
<td>0.1124</td>
</tr>
<tr>
<td>1001</td>
<td>3461.6330</td>
<td>1002</td>
<td>3484.1910</td>
<td>1003</td>
<td>3506.8970</td>
<td>1004</td>
<td>3529.7500</td>
</tr>
<tr>
<td>1005</td>
<td>3552.7520</td>
<td>1006</td>
<td>3575.9030</td>
<td>1007</td>
<td>3599.2060</td>
<td>1008</td>
<td>3622.6610</td>
</tr>
<tr>
<td>1009</td>
<td>3646.2680</td>
<td>1010</td>
<td>3670.0300</td>
<td>1011</td>
<td>3693.9460</td>
<td>1012</td>
<td>3718.0180</td>
</tr>
</tbody>
</table>

Table 1: Grayscale Standard Display Function: Luminance versus JND Index

When calibrating a medical display to a luminance of \( Y = 500 \, \text{cd/m}^2 \) and a black point close to zero, i.e., \( Y = 0.5 \, \text{cd/m}^2 \), basically all gray levels are redistributed according to the DICOM Grayscale Standard Display Function (GSDF). The full DICOM GSDF function is depicted in Figure 8. One can easily observe that, due to its logarithmic scale, equal steps on the luminance axis do not create equal JNDS.
The primary intention of the DICOM GSDF calibration is to redistribute the range of available luminance in such a way that equal driving-level steps create equal perceptual differences. This is demonstrated in Figure 9, such that between every succeeding gray level, we have one JND difference. Each gray level can be clearly distinguished and every transition looks perceptually the same.

Prior to the DICOM GSDF, radiologists relied on differences in luminance between adjacent parts of the screens to make a judgment. At that time, certain parts of the image were compressed in terms of JNDS, where other parts were stretched so that they consumed much of the available grayscale resolution without generating more visible differences. This situation was resolved by the DICOM GSDF. Furthermore, this guarantees that a digital medical grayscale image is always displayed in a consistent way, and that it has the maximum possible diagnostic value.
3.2 Color JND

Just as in the case of the DICOM GSDF, where equal amounts of luminance (Y) increases did not result into equal visual differences, equal color differences in X, Y and Z will not result in equal perceived color differences.

Many people have studied visual color differences. David MacAdam, an American physicist and color scientist who studied colorimetry and color order, developed a simple way of representing them. The xy chromaticity diagram (Figure 10) depicts a number of ellipses. The human eye is not capable of distinguishing any perceptual difference between the color points inside each ellipse, thus confirming the notion that equal increases in xy do not yield equal perceptual differences. In the blue region, for instance, the human eye is capable of discerning very small changes in xy as a perceptual difference. (Note that the actual ellipses in Figure 10 are enlarged by a factor of 10).

In an attempt to create a more perceptually uniform color space, the CIE defined the L*a*b* color space in 1976. This color space is based on nonlinear compressed CIE XYZ color space coordinates. The three components describe the lightness of a color (L*), its position between magenta and green (a*) and its position between blue and yellow (b*). Uniform changes of the components in the L*a*b* color space aim to correspond to uniform changes in perceived color. Consequently, the perceived color difference can be approximated by taking the Euclidean distance between two color points, which was the first definition for the color difference formula \( \Delta E_{ab}^* \) in L*a*b* color space.

\[
\Delta E_{ab}^* = (L_2^* - L_1^*)^2 + (a_2^* - a_1^*)^2 + (b_2^* - b_1^*)^2
\]

However, the L*a*b* color space was ultimately deemed not as perceptually uniform as expected, which forced the CIE to adapt, in 1994 (\( \Delta E_{94}^* \)) and again in 2000 (\( \Delta E_{00}^* \)), the color definition formula.

Although similar challenges exist in distinguishing certain just noticeable color differences in medical color images, a standard color DICOM function has not yet been defined. A solution is urgently needed, especially in cases where color information holds certain diagnostic value. Once CIE updates its color difference formulas to quantify perceptual color differences, these can be used to assess if a color difference is just noticeable or not. Furthermore, new standards will serve as a useful tool to validate the perceptual color uniformity of a display system. It is also inevitable that, in the coming years, the color behavior of medical displays will also be standardized, with the principle of color JND and the color difference formulas playing an important role.
Challenges for medical color displays
4. Challenges for medical color displays

4.1 Stability

With the introduction of the DICOM GSDF calibration on monochrome displays, we learned a very important lesson: where initial calibration tools performed a software calibration only, monitor stability over short or longer periods of time has proven to be more of a challenge. It makes absolutely no sense to perform any kind of calibration on a display that is not stable. Therefore, medical displays include hardware technology to keep the monitor stable over short and longer periods of time.

Most of today’s implemented technologies maintain display stability for luminance changes only. These include stabilization circuits for the backlight only, or for more advanced front sensor technologies that stabilize the entire display optical stack for luminance changes.

To improve the color stability of modern medical color displays, we need to tackle the additional complexity of stabilizing the three components of a color. The color stability of a display is influenced by the luminance and the spectrum of the backlight and the aging of the color filters. It is clear that backlight-only stabilization cannot provide this capability. Instead, more advanced and precise color measurement devices mounted on the front of the display are needed to solve this color stability problem.

4.2 Accuracy

To achieve good image quality and consistency between multiple displays, it is important that the sensor be accurate in measuring colors. One might think that calibration of a color is just three times as complex (X, Y, Z) as calibrating a monochrome display (Y only).

In reality, the situation is far more complicated. In a monochrome system, the DICOM function is typically modeled with 32-, 64-, 256- or 1024-bit measurements. Needless to say, the calibration time needed is proportional to the number of measurement samples taken during the calibration. More advanced calibration algorithms offer fast calibration, which uses a lower amount of measurements and relies on interpolation to recalibrate the DICOM function. As luminance is a one-dimensional entity, this interpolation can simply be done on a one-dimensional function.

As explained in Section 2.2, all possible colors within the XYZ color space form a volume. The colors that can be represented on a medical color display lie within a sub-volume, called the color gamut of the display (Figure 11). To calibrate the color display, it is not enough to model the three axes with 32-, 64-, 256- or 1024-bit measurements; we need to describe the complete gamut with 32³, 64³, 256³ or 1024³ measurements. However, this would be a very time-consuming process, making it unacceptable.

Color calibration is generally described as the process of defining a transformation from the actual color response of the display to a known state. The majority of the color calibration algorithms use a 3D look-up table (LUT) with a limited number of samples to describe this transformation. For every input color, the corresponding corrected color is looked up in the 3D LUT. Colors that are located in between the grid points of the 3D color LUT are interpolated. It is clear that the more samples the color LUT contains, the finer the interpolation result will be.

![Figure 11: Color Gamut of a Display as a Sub Volume of the XYZ Color Space](image)
Adding a well-dimensioned 3D LUT and a high performance interpolation technique to the display system is however no guarantee for achieving cutting-edge color calibration. Color calibration is all about, as stated before, defining a transformation to a known state. If this transformation is not calculated with the necessary precision, the color response of the display will not accurately resemble that known state. A popular known state is sRGB. The precision of the calibration can be assessed by calculating the color differences (see color difference formulas in Section 3.2) between the color response of the display system and sRGB.

Although the calibration to a known state is a valid and valuable method for many applications, another approach is also conceivable. In its newest platform, Barco combines the DICOM GSDF with a new function called CSDF (Color Standard Display Function). The relevance of CSDF for medical color imaging will be explained in Section 5.

Similar to the philosophy of DICOM GSDF (see Section 3.1), the intention of the CSDF is to create a perceptually uniform display for color. To make the display perceptually uniform for color, an advanced algorithm has been developed that optimizes the entire color response of the display so that the change in number of color JNDs for every color step in every direction of the color cube is almost equal. The CIE 2000 definition of the color difference formula is used to calculate the number of color JNDs. However, there are additional constraints that apply: the color calibration algorithm must not reduce the usable part of the gamut of the display, the primary and secondary colors remain on the edges of the RGB color cube, and the gray colors are DICOM GSDF compliant.

The effect of the CSDF for the primary and secondary colors can be seen in Figure 12 and Figure 13. The graphs show the color differences between successive points for six different lines in the color cube. When looking at the line from black to green, indicated with the green line and squares in Figure 12, you will observe that the color differences around point 5 are much smaller than at the end of the line. This means that color differences around this point 5 are much less visible than around, for instance, point 12. Comparable to the explanation about DICOM GSDF (Section 3.1), certain parts of the colors are compressed in terms of color JNDs, where other parts are stretched. This indicates an inefficient use of the available color resolution.
After applying CSDF, the color differences between successive points are much more uniform, as indicated in Figure 13. This optimizes the visibility of color details over the entire color space. The effect of the CSDF can be clearly seen in Figure 14. The figure shows a cube composed of different little cubes with small color steps between adjacent cubes. On the left side is the image before the application of the CSDF. On the right side, the image has been color calibrated. It is obvious that small color differences are much more visible after the color calibration.

The accuracy of the CSDF cannot be assessed by calculating the color differences between the color response of the display after calibration and the known state. However, this does not mean that the color calibration cannot be validated in the case of CSDF. It can be checked by calculating the color differences for every color step and by comparing them with the average color difference. The challenge for an accurate color calibration in the case of CSDF lies in having all these differences as close as possible to each other in order to obtain a truly perceptually uniform display.

Figure 14: Enhanced Visibility of Color Differences After Calibration (left native, right CSDF)
Medical relevance of color
5. Medical relevance of color

5.1 The Medical Relevance of a Perceptually Uniform Color Calibrated Display

At SIIM (The Society of Imaging Informatics in Medicine), there was much discussion surrounding the additional modalities that CIIPs (Certified Imaging Informatics Professionals) now need to manage, as well as how different the images and workflow are beyond Radiology and Cardiology. Specifically, people spoke at length about the growing amount of visible light images they need to manage dermatology, endoscopy, and microscopy, agreeing that the quantity of these images is growing and will eventually surpass the number of radiology images.

With few established standards, people are struggling to support these image types. The hospitals are asking Radiology to take on this challenge since they already have a strong process and set of standards. How does (or will) a calibration solution support these types of images? Is there a medical relevance to calibrating a display in such a way that a perceptually uniform color response is obtained as described in section 4.2?

Depending on the specific discipline of medicine, there may be other requirements for the representation of colors. For surgery and examinations that use endoscopes, for example, an exact representation of colors is an absolute requirement. The endoscope combined with the display can, in this case, be considered an extension of the doctor’s eyes. The same can be said for the interpretation of wound photographs, where the color provides an indication if a wound is healing. In all these cases, the medical display should be calibrated toward a known state, usually sRGB.

The situation is different for the emerging markets of digital pathology or quantitative imaging. For these types of images, it is critical, that the doctor can visualize relevant medical features in the image similar to the situation depicted for grayscale images. To facilitate this discovery, it is important to visualize the specific differences between the features and the background of the image. For these cases, the display must be appropriately calibrated to represent the colors in such a way that the visibility of the features is optimized over the complete range of colors. This situation is analogous to GSDF for grayscale images (see Section 3.1), which takes into account the actual gamut of the display and that which can be validated and adjusted automatically over the lifetime of the display.

In a conventional digital chain for pathology, until now, the display has not been considered an essential part of optimizing the detectability of the features in the scanned slides. The typical approach has been to represent the colors in exactly the same way that the pathologist perceives them when looking through the microscope. To obtain this, the scanned slide is saved in the sRGB color space, and the display is assumed to be sRGB-calibrated.

However, this approach contains some flaws. First of all, one must ask what the correct color is. The colors perceived when using a microscope depend upon the spectrum of the light source of the microscope. Thus, a slide will look different from microscope to microscope. In addition, every hospital has its own procedures for preparing and staining slides. Although more or less the same procedure is used in the different labs, the intensity of the staining can vary a great degree. To further complicate the situation, the colors can differ quite a bit after scanning the slides, depending on the scanner used.

This is illustrated in Figure 15 (origin: Dr Juan Ruiz, Complejo Hospitalario De Toledo), where you can see the same slide scanned by three different scanners with the same illumination. The two left images have been scanned by a device from the same manufacturer, yet even then, reflect variations. Therefore, it is not a wise strategy to count on the exact representation of colors in the case of digital pathology.

Figure 15 Example of the Same Slide Scanned by 3 Different Devices
For quantitative medical imaging, the above-mentioned objections are invalid. Quantitative medical imaging is typically the visualization of calculated values as pseudo colors on top of other medical images or as images by themselves. Because these colors are calculated, it is possible to define a color space in which the image is rendered, for instance sRGB, and by using a correctly calibrated display, the calculated colors can be accurately visualized. An example of such an image is shown in Figure 16.

When looking at this image in more detail, one observes that only a very small amount of the scale is represented by reddish colors and that the greenish colors represent the largest range of the quantitative values. Unfortunately, it is very difficult to distinguish between many of the colors on this scale. A perceptually linear color scale could help to optimize the visualization of the quantitative colors and potentially reveal hidden details in the image. However, this can only be realized when taking into account the entire color gamut of the display used for the visualization of the image.

Both digital pathology and quantitative imaging need to visualize the differences between the features and the background of the image. For the reasons explained in this section, one can conclude that at least both quantitative imaging and digital pathology will benefit from a perceptually linear color display.

5.2 Clear-base or Blue-base

A legacy we inherited from film-based radiology is the use of clear- and blue-base film. When moving to digital medical imaging, certain modalities have preferred the use of blue-based imaging, while general radiology moved toward clear-base.

As more and more mammography images moved into the PACS systems, this created a challenge for those radiologists who preferred to read images on blue-based systems. With the introduction of the more advanced color displays, on-the-fly switching between clear- and blue-base is simplified, allowing radiologists to read images with precisely the type of illumination they desire.
6. Conclusion

As the use of color in medical imaging continues to evolve, going beyond simple annotation to depicting more complex diagnostic information, it is essential that medical color displays meet a higher standard in line with the standards for grayscale. We have briefly examined color and its interaction with human vision as well as some of the challenges and considerations with meeting this standard for medical color imaging. Recent technological breakthroughs with respect to color calibration have equipped diagnostic displays to achieve color accuracy and consistency. These medical displays provide major benefits over their grayscale predecessors, which can be summarized as follows:

1. A medical color display allows the use of a much richer GUI, enabling faster diagnosis while supporting improved workflow.
2. The continuous trend toward viewing medical images from multiple modalities has experienced staggering growth. Although first and second generation multi-modality displays, such as the Barco Coronis Fusion 6MP and Coronis Fusion 6MP LED, addressed many of the requirements, new imaging challenges demand more advanced calibration technologies such as those found in the new Barco Coronis Units.
3. Just as with a monochrome display, the default characteristics of the monitor will not calibrate a color display in a way to optimally distinguish between different shades of color. In order to do so, more advanced color calibration methods are needed.
4. Advanced color calibration algorithms do not make any sense if they are implemented on displays that are not consistent. Special precautions are needed to guarantee that the color remains constant over time.
5. Consistency is also important even in the case whereby a radiologist or physician uses quantitative medical color imaging. The consistent representation of color on displays deployed throughout the hospital makes it much easier for the doctor to interpret the images.
References
7. References

1. Perceptual uniformity of commonly used color spaces Ali Avanaki2, Kathryn Espig1, Tom Kimpe1, Albert Xthona2, Cedric Marchessoux1, Johan Rostang1, Bastian Piepers1
   1 BARCO N.V., Healthcare Division, Kortrijk, Belgium; 2 Barco Healthcare, Beaverton, OR, USA

2. Does the choice of display system influence perception and visibility of clinically relevant features in digital pathology images?
   Tom Kimpe1, Johan Rostang1, Ali Avanaki1, Kathryn Espig1, Albert Xthona1, Ioan Cocuranu2, Anil V. Parwani2, Liron Pantanowitz2
   1 Barco Healthcare; 2 Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, PA, USA

3. Color difference evaluation for digital images using a categorical judgment method
   Journal of Optical Society Am. 30, 616-626 2013 Haoxue Liu, Min Huang, Guihua Cui, R. Luo, Manuel Melgosa

4. CIE Publi 142-2001, CIE S 014-6/E:2013 (CIE TC 1-57 Chair Alan Robertson)