P-03: New Metric for Display Color Gamut Evaluation

Moshe Ben Chorin, Dan Eliav, Shmuel Roth, Assaf Pagi and Ilan Ben-David
Genoa Color Technologies, Herzelia, Israel

Abstract

We suggest a metric for evaluating color gamut of displays based on summing the probability of using a certain part of the display color gamut multiplied by an impact parameter for that color region. We further present methods for evaluating the probability of use, and the impact of color.

1. Introduction

Until about a decade ago all color displays were the same from color point of view. All displays were based on CRT with well-known response and practically the same color gamut. The CRT color gamut, which is limited by the choice of phosphors, is standardized for example in Rec. 709. The emergence of other displays, such as projection or LCD displays created the possibility to alter the color gamut, however at least during the first years of these displays the color gamut was usually restricted with respect to that of Rec. 709. In recent years technological advancement in different fields provided the opportunity to expand the color gamut of new displays beyond that of Rec. 709 or even NTSC. These new displays are based on either RGB or multi-primary technologies [1].

Flexibility in the choice of primaries offered a possibility for a design of the color gamut according to the required application [2]. An application of major interest is that of TV and home entertainment. For this application, there is ample evidence that enhanced color saturation contributes to the viewing experience [3], and thus the expansion of the color gamut beyond that of Rec. 709 is of great interest. However, the question of what should be the design goal for the color gamut and how well the color gamut of a certain display suits its intended application is not yet answered.

In the absence of better metric, the display industry adopted the NTSC primaries as reference, and the metric for color gamut which rules the scene is known as NTSC percentage (%NTSC), the relative area of the triangle enclosed by the primaries of the examined display with respect to that enclosed by the NTSC primaries, measured in the two-dimensional x-y chromaticity space. While this metric may have been relevant for early non-CRT displays, whose color gamut was close that of Rec. 709 (in the above metric Rec. 709 represent 72% of NTSC), it is not relevant for wide gamut displays due to its failure in providing any qualitative information with regards to color reproduction issues. In particular, it does not provide any insight or design goal for the required performance of displays.

In this paper we would like to suggest guidelines for a new metric for color gamut evaluation, which goes beyond the simplistic approach of calculating relative area with respect to a certain reference area. We describe the logic behind this metric and bring some initial results of perception tests made in order to substantiate this concept.

2. Limitations of the %NTSC and other relative area metrics

Although the %NTSC metric is widely used it suffers from various limitations. First of all, the tendency to use any metric (including the %NTSC metric) as a marketing tool implies that higher numbers are set as design goal. The %NTSC metric is thus doomed to press color gamut area towards larger and larger numbers. Although at first sight this may be conceived as harmless one must remember that as long as data is provided in digitized form with a fixed number of bits per channel, the increase of color gamut implies that the same number of data points is used to cover a larger area in the color space. Thus, the average color distance between neighboring data points increases, which may result in non-smooth color transitions and increased image noise. Moreover, the extra area provided is not necessarily useful in the reproduction of colors, since it may represent colors, which are rarely available in real life situation. Combined with the data limitation, this further implies that a too large gamut would waste digital codes (and color gamut volume) on irrelevant colors.

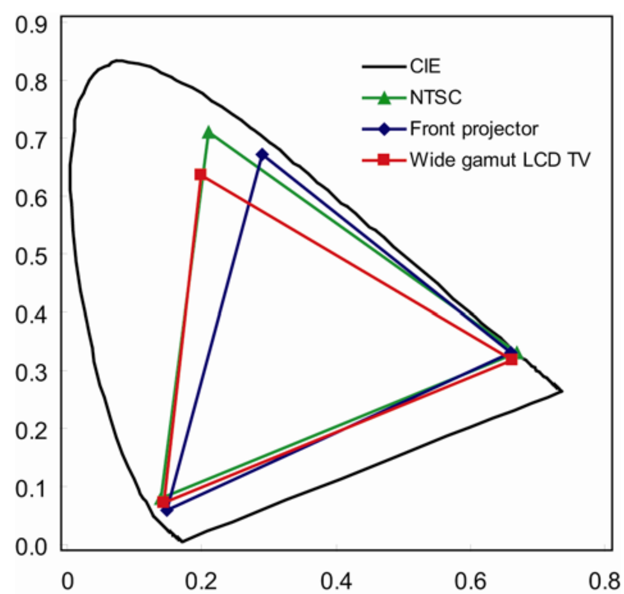

Figure 1: A front projector and a wide gamut LCD TV, both sold as "Home Cinema" displays compared with NTSC color triangle. Note that both displays cover 87% of NTSC.

Furthermore, the %NTSC metric quenches the available two-dimensional gamut information into a single number representing area. However, there may be many color gamuts that would yield the same %NTSC, but nevertheless would have completely different reproduction of color and impact on the viewer. For RGB displays with triangular gamut area, and under the assumption that the red and the blue primaries have fixed position in the chromaticity space, the position of the green point may be wildly shifted without affecting the gamut area, since the only requirement for equal area in such circumstances is that the height of the triangle should stay the same. The relative area approach thus does not provide any guideline for the preferred green chromaticity. As an example consider figure 1, in which the color gamut of an LCD display is compared with that of a projection display, both made by the same manufacturer. Although both displays have an 87% NTSC gamut, it is clear that many RGB inputs would translate to different colors in each. It is obvious that the color design of these displays derives from technological constraints rather than from color quality absolute merits.

The reason for the failure of the %NTSC metric in the above case is that it treats all colors as being of equal importance. However, this is not the case for all applications [2]. The color gamut should be tailored to meet required color performance of a specific application. In particular, for home cinema and TV applications a logical design target should be the cinema color gamut. Indeed, in a recent publication the cinema color gamut has been suggested as a reference for a relative area metric instead of the NTSC triangle [4]. Furthermore, it was also suggested that the values of relative area would range from 0 – 100%, and that colors outside the reference color gamut should not be counted. Thus color gamut area outside the reference gamut or above 100% is not beneficial. In addition, ref. 4 also advocates the use of the u'-v' chromaticity space rather than the x-y space for the area comparison, because the u'-v' space is perceptually more uniform than the x-y chromaticity space [4]. Although this suggestion still relies on the calculation of relative areas, nevertheless it overcomes some of the problems discussed above, in particular the tendency to push the color gamut further and further towards green. The choice of cinema gamut as reference is closely associated with the relevant application of home entertainment, where the "look and feel" of cinema should be reproduced. It also provides an upper limit for the required color gamut. However, the question of which color space is best for quality evaluation, and more importantly, why should a color gamut quality parameter depend on the choice of color space, remains.

3. Guidelines for a better metric and methods for its definition

The relative area metrics we have discussed above may be stated mathematically as:

$$Q = \frac{\int_{Display\ Gamut} du'\,dv'}{\int_{\mathrm{Re}f\ Gamut} du'\,dv'} \qquad (1)$$

This metric includes explicitly in its definition a color coordinate system, the u'-v' color space in the case of eq. 1.

We believe that a color quality metric should not depend on the choice of coordinate system, since it should represents human preference regardless of its underlying mathematical framework. To realize that consider that we have only N different colors of equal importance in the world, and that the quality of a reproduction system is determined by its ability to reproduce these colors, namely by counting how many of the N colors fall within the color gamut of the reproduction system. The logic behind this criterion is straightforward and we note that this measure is independent on the choice of coordinate system.

At first glance, it may appear that the metric Q, defined in eq. 1 meets the above criterion, since it represents the ratio between the gamut areas of the reproduction system to that of the reference. However, the ratio of areas does not necessary equal to the ratio of the number of points within the display gamut to the total number of points N, even if all N points are within the reference gamut. Indeed, the integral over areas would yield a different number if the coordinate system is changed from u'-v' to x-y, as demonstrated in ref. 4, which shows that the relative gamut areas with respect to the NTSC triangle of different displays calculated in u'-v' and x-y coordinate system differ.

Turning back to the "points counting" approach we note that the only case where the relative area metric would yield the same result as the "points" metric would be when the distribution of the points with the reference gamut is uniform. Indeed the rational behind choosing u'-v' as a chromaticity space instead of the x-y space is that the first has better perceptual uniformity. However, this uniformity relates to the ability of the human visual system to discriminate colors and not to the distribution of colors in nature or in cinema. Yet, there is no reason to believe that different colors appearing in nature (or on a TV display) are uniformly distributed over the respective gamut.

An important issue therefore, is the determination of the probability distribution. Ideally, we would like to have a representative collection of nature colors that could be analyzed to provide that information. Still, in our view, in the discussion of consumer electronics displays video material should be regarded as the reference of color source, rather than actual colors. As video source we regard material originated from both video and film capture. In order to study the color distribution in such material we have examined 250 RGB frames extracted from about 100 different video sequences. Figure 2 depicts the logarithm of the probability distribution (note that the distribution covers several orders of magnitude). As expected, grays appear more frequently then other colors. The probability decays in a slower rate along the blue-yellow axis, along the CIE standard daylight illuminants curve. The relatively high proportion of blue may be associated with either the use of blue in dark night scenes or the fact that sky and water bodies may occupy large areas of the frame. An important aspect is the low presence of magenta and greenish-cyan colors, while the presence of yellows, oranges and bluish–cyan is relatively high. This presents important information for the designer of a wide gamut display, since although the distribution is taken from recorded material, prepared to be consistent with Rec. 709 limited gamut, it nevertheless provides some clues for the more general distribution of colors.

A word of caution is necessary. We believe that the distribution of colors in video is likely to differ from that which the naked eye would observe in the real world, because of editorial and aesthetic considerations taken in its preparation. For example, dialogs in films are often shot in such way that faces occupy most of the frame, whereas in reality, at normal speaking distance the viewer will observe the face of their counterpart at a much narrower angle, with a lot of surround colors around it. In addition, choices made by directors and photographers are biased towards drawing attention, creating an impact and achieving certain aesthetic goals, resulting in a contemplated selection of colors in the scene, or at least not a random one. Finally, the preparation process of the material for screening or broadcasting imposes further constraints on the color distribution, e.g. gamut compression into a target standard.

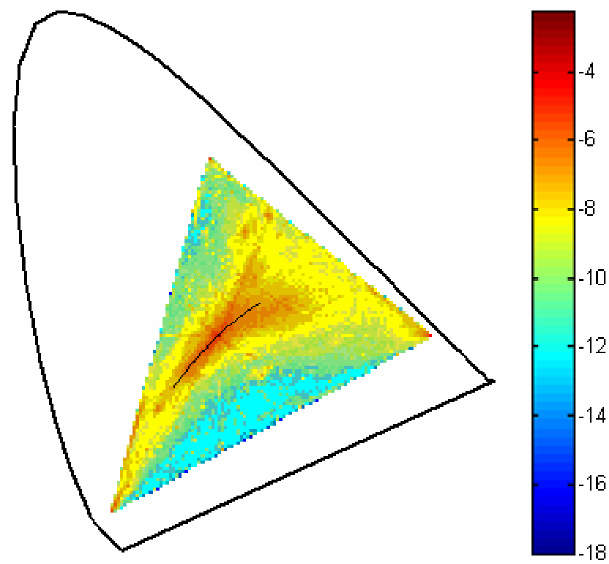

Figure 2: Log distribution of colors in video content. The black line indicates the day light illuminant locus.

The results in figure 2 are sufficient for showing that indeed the distribution of colors is far from uniform, and thus the relative area (or volume) approach should be discarded. Hence, we suggest that the probability distribution of colors in video and film would be used to weight the area (or volume), namely defining the quality metric as:

$$Q = \left| \log \left(\frac{\int_{Display\ Gamut} p(u',v')du'dv'}{\int_{Ref\ Gamut} p(u',v')du'dv'} \right) \right|$$

where p(u',v') is the probability distribution of having a color with coordinates u'-v'. The integral over display gamut of the probability distribution is by definition the number of points within this gamut, under the assumption that all N points are within the reference gamut. Therefore, it is clear that changing the coordinate system would not affect the new ratio we have defined. Note that the probability function is currently provided only within the Rec. 709 gamut. However, it can be expanded to cover the color gamut of cinema using some gamut reconstruction heuristics, as will be discussed elsewhere. The logarithm of the ratio is taken in order to compensate for the exponential nature of the distribution, and the absolute value set the scale to positive numbers. The quality scale as defined here would yield smaller numbers for more efficient gamut, closer to the reference gamut, and can be regarded as a measure of discrepancy from the ideal reference.

Even though the approach suggested above solves the problem of coordinate system, we believe it lacks an important ingredient. It is clear that most of the relevant information in an image lies in its luminance signal, and it is also known that on the average the world can be approximated as gray with a relative luminance of 20%, an approximation known as the gray world assumption. This is consistent with our findings that the distribution of colors peaks around the gray scale. Nevertheless, we know that colors, and in particular saturated colors, are important to our perception of images. Research on color naming across cultures indicates that colors of high saturation and unique hues are frequently named, red being the color named by most cultures (after black and white), followed by yellow and green and then blue [5]. Most cultures do not have unique color names for pastel colors. There is also plenty of work devoted to the influence of colors on our feelings, again indicating that saturated colors have more effect than pastel ones. This may be due to the relative scarcity of saturated colors in nature, which call for attention when viewed against the less saturated environment. Therefore, we suggest that the weight function would be a multiplication of the probability distribution and a color impact factor that represents the importance of various colors to the viewing experience.

The definition of specific color impact is subtler. In order to determine the impact of different color regions on the preference of displayed images, we have conducted a series of perception tests. In all tests, two displays, one having a color gamut close to Rec 709 and the other a wide color gamut, were placed side by side and set in such way that colors within their common color gamut would be as close as possible. Images that were displayed in this setup differed mainly in the highly saturated portions of their content since grays and pastel colors matched very closely on both displays, thus, the isolation of the contributing factor for a declared preference was enabled. The wide gamut display featured gamut extension throughout the whole color space, with

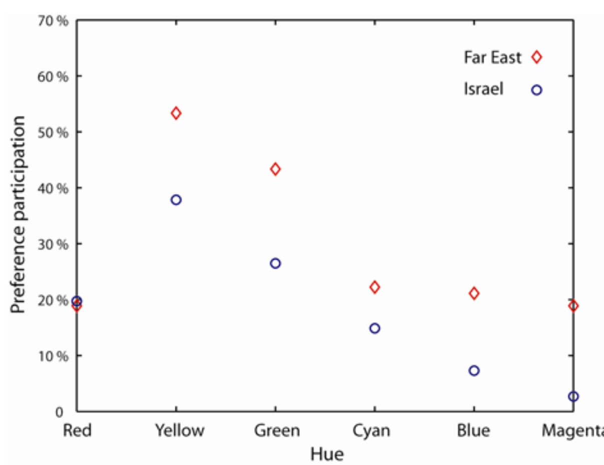

Figure 3: Contributions of the various (saturated) hues in the image to the overall preference of one display over the other.

the main extension in the yellow and cyan regions. In the experiments, different sets of images were shown on two displays set as described above, and the participants were asked to state their preference and associate it with six major hues (red, green, blue, cyan, yellow and magenta). Images in which the distribution of colors was inside the gamut of both displays were used for control. In both tests, the overall preference was significantly towards the wide gamut displays. The tests however, provided also an indication for the reason for this preference. Figure 3 depicts two sets of preferences measured in two separate experiments that took place in Israel and in the Far East. It can be seen that the gamut extension in the yellow portion had more impact than the extension in the cyan portion. In addition, a clear agreement in preference was found between the subjects of the two experiments, although they were performed in different locations and with different sets of images. This finding supports existing evidence for the low significance of cultural differences in the determination of viewing preferences [6]. The relatively low impact manifested for red and magenta has surprised us, as it stood in contrast to our intuition. We believe that this test should be repeated for a display with a larger gamut extension in the red – magenta portions. Preliminary results of work we have performed with such display indicate that this portion may have higher impact in deeper saturations, suggesting that color impact does not augment linearly.

We further intend to examine a hypothesis that color impact is content dependent, and that the impact of a certain color is associated with the amount of information it carries within the image, for example a yellow object on yellowish background would carry less impact than a yellow object on bluish background. In their study of color temperature preference, Vogles and Heynderickx [7] found the preference to be content dependent as well; for example high color temperature was preferred when yellowish content was displayed, and the contrary for bluish content. We therefore predict that as long as consumer displays will use high color temperatures (>8000K), colors in the red to green regions will have greater impact than those in the blue to green region, and the opposite when low color temperatures are used. It should be however realized, that if content dependency exists, the impact of different colors may change from one image another. Thus, for the gamut quality metric an average impact should be considered.

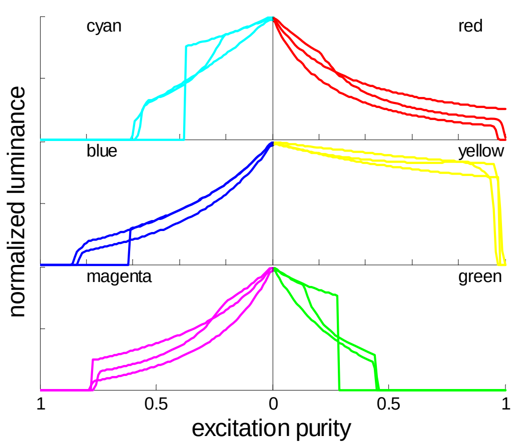

Figure 4: Luminance vs. excitation purity for film gamut (solid lines), DCI compliant display (dotted lines) and multi-primary display designed for home entertainment application (dashed lines)

So far, we have limited ourselves in the discussion to the 2D chromaticity plane. However, the color gamut is three-dimensional, a fact that should be considered in any display design. In figure 4 we compare the three dimensional color gamut of film, DCI compliant RGB projector and a multi-primary display, by presenting luminance – excitation purity [8] curves representing cross sections along different hue lines (straight lines connecting two saturated colors with opposite hues through the white point on an x-y chromaticity diagram). It can be seen that a good match is achieved between the multi-primary performance and cinema envelope, while the DCI compliant display falls short in certain regions, as well as providing non-required extra gamut in others. Nevertheless, the multi-primary display and the DCI projector yield similar scores in the relative area metric, showing again its shortcomings.

4. Conclusions

Relative area metric is attractive in its simplicity; however it fails to provide meaningful information regarding the quality of color reproduction of a given display. Our proposed metric is an attempt to relate the evaluation of the technical attributes of displays to the viewing quality domain, where the video content is expected to be reproduced optimally. The metric is based on examining the coverage of the display gamut with relation to the probability distribution and the impact of colors.

5. References

[1] S. Roth, I. Ben-David, M. Ben-Chorin, D. Eliav, and O. Ben-David, “10.2 Wide Gamut, High Brightness Multiple Primaries Single Panel Projection Displays”, SID Symposium Digest of Technical Papers -- May 2003 -- Volume 34, Issue 1, pp. 118-121.

[2] D. Eliav, S. Roth and M. Ben Chorin, “Application driven design of multi-primary displays”, Proc. 14th Color Imaging Conference, p. 280 (2006).

[3] H. de Ritter and S. Endrikhovski , “33.1 Image Quality is FUN: Reflections on Fidelity, Usefulness and Naturalness”, SID Symposium Digest of Technical Papers – May 2002 – Volume 32, p. 986-989 and references therein.

[4] M. S. Brennesholtz, “Expanded color gamut displays – Part 1”, Information Display September 2006, “Expanded color gamut displays – Part 2”, Information display, October 2006.

[5] T. Regier, P. Kay and P. S. Cook, "Foacl colors are universal after all", Proc. Nat. Acd. Sci. 102, p.8386 (2005) and references therein.

[6] S. R. Fernandez and M.D. Fairchild, "Observer preferences and cultural differences in color reproduction of scenic images", Proc. 10th Color Imaging Conference, p. 66 (2002)

[7] Vogels, I. and Heynderickx, I., “Optimal and acceptable white-point settings of a display”, IS&T/SID 12th Color Imaging Conf 233 (2004)

[8] G. Wyszecki and W. S. Stiles, Color Science: concepts and Methods, Quantitative Data and Formulae, 2nd Ed., Wiley 1982, p. 175