https://doi.org/10.2352/ISSN.2470-1173.2017.18.COLOR-034

A metric for the evaluation of color perceptual smoothness

Cedric Marchessoux; Barco N.V.; Kortrijk, Belgium
Johan Rostang; Barco N.V.; Kortrijk, Belgium

Abstract

This paper presents a new metric for evaluating the color perceptual smoothness of color transformations. The metric estimates three dimensional smoothness to cover the full gamut of the transform. This metric predicts any artifacts like jumps in any gradient introduced by the transformation itself. From the state of the art, three works have been found and compared for evaluating their pros and cons. Based on these previous proposals, a new metric has been developed and tested with several applications. The metric is based on the perceptual distance: $CIEDE2000$. The defined metric is dependent on the number of ramps and the number of colors per ramp but these two parameters can be reduced to a single one called granularity. The proposed metric has been applied on the AdobeRBG and sRGB color spaces with and without the addition of artificial artifacts and tested for a large variety of granularity values. Several basic statistics have been proposed and the root mean square seems to be a good candidate for representing the global smoothness. The metric demonstrated robustness for evaluating the global smoothness of a transform and also or detecting small jumps.

Introduction

Aim

The term "color transform" hides a wide variety of different operations ranging from simple color filtering applied on a photo to gamut mapping algorithms through the effects of discretization inherent to any digital systems. For any of them, ensuring smooth color transition is very important to avoid.It is also important to ensure a smooth transition in colors to avoid creating disruptive elements on the resulting image. The aim of the work presented in this article was to find a metric for evaluating the smoothness of a display color transformation. This metric should predict any artifacts like jumps in any gradient introduced by the transformation. In this new framework, either the system should be in a native mode or a transform should be applied via the application of three-dimensional Look-Up-Tables (LUT) or ICC profiles. The metric should evaluate the color perceptual smoothness of different color calibration, or any gamut mapping transformation. The metric should be able to evaluate the presence of artifacts on an entire LUT in minimum time. The metric estimating three-dimensional smoothness should not be reduced and minimized to a "mono-numerosis" and rather extended to a whole set of statistics.

State of the art

From the state of the art, three works have been found, the first one from Green in 2008 [1], the second one from Kim in 2010 [2] and the third one by Aristova in 2011 [3].

Green proposed in 2008 [1] a methodology for estimating the smoothness of a color transform applied on a color gradient. The method is represented by the Figure 1 originally from the presentation of the paper [3]. For any input colored ramp with n pixels, the metric $CIEDE2000$ [4] defined by the CIE is computed between two L^*, a^*, b^* consecutive triplets of the ramp, resulting in a $(n-1)$ $CIEDE2000$ ramp. From this resulting ramp, a second derivative is calculated by simply computing the approximation of a derivative by subtracting two consecutive elements of the $CIEDE2000$ ramp resulting in a $(n-2)$ ramp for which summary statistics are used.

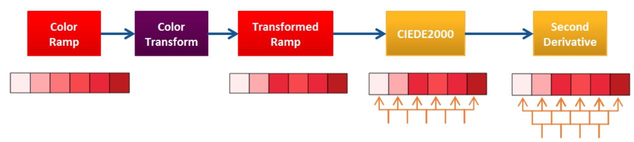

Figure 1. *Green's framework.*

In Aristova paper [3], the author assumes that the 95^{th} percentile is calculated representing the smoothness of the color transform of the input ramp.

Kim et al [2] proposed another framework represented on Figure 2. It is also applied on 1D ramps by using the first and second derivatives; but contrary to Green the color distance is estimated thanks to the original Euclidean $CIEDE$ metric instead of $CIEDE2000$. The derivative is approximated by subtracting two consecutive elements of the ramp. In addition to the estimation of rough transitions - or tone jumping - Kim proposes to evaluate also tone clipping. The first derivate was used for evaluating the tone clipping by using the 5^{th} percentile and the second derivative for tone jumping with the 95^{th} percentile in the evaluated ramp. The resulting tone jumping is weighted by applying a weighting factor calculated with the first derivative and the tone clipping.

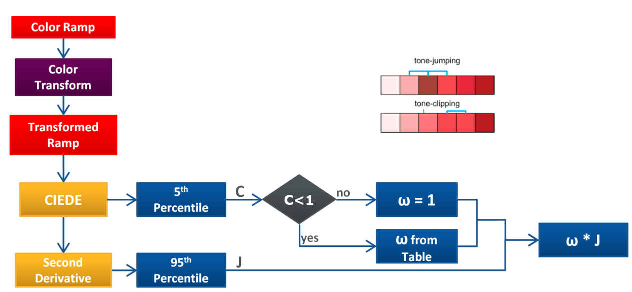

Figure 2. *Kim et al framework.*

Kim has also shown that the optimum percentile level was determined to be 95^{th} to best fit the subjective data from the measurement of the magnitude of tone jumps of 96 test gradations.

Aristova et al [3] proposed to evaluate not only one ramp but all vertical and horizontal ramps of one *RGB* plan converted to $L^*a^*b^*$. Aristova's framework computes a first derivate on each individual color component: L^*, a^* and b^* by subtracting two consecutive elements of the ramp as illustrated by Figure 3. Then a second derivative is applied. The 95^{th} percentile is computed for all ramps. The maximum of both vertical and horizontal ramps are extracted for the studied plan. These two maximums are multiplied and the result corresponds to the smoothness of the selected plan. Aristova finally summarizes the global smoothness to the average smoothness of the plans.

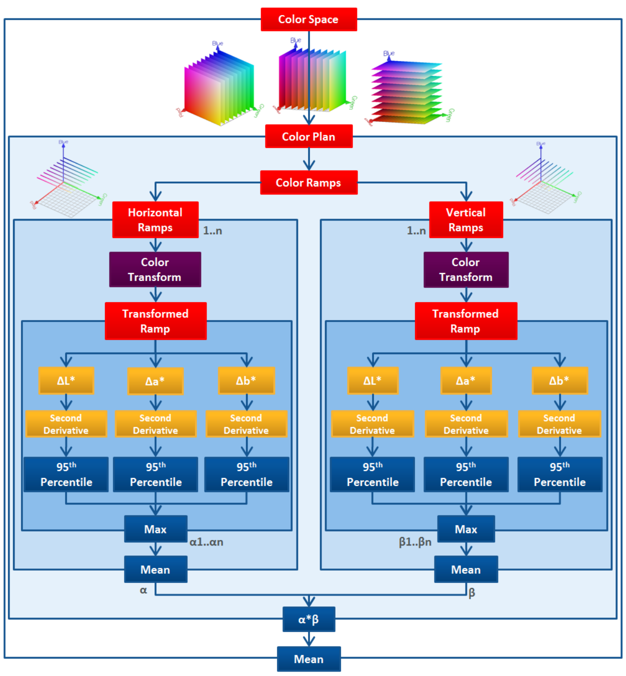

Figure 3. *Aristova et al framework.*

Method

The three metrics presented in the previous section do have pros and cons. We hereby propose a new metric that keeps only the pros of the metrics from the state of the art without their weaknesses. *CIEDE*2000 is the last standard proposed by the CIE and therefore this metric should be the core color perceptual differences metric for defining our smoothness metric. Aristova proposed to evaluate the smoothness component-by-component: L^*, a^*, b^* and by taking the maximum of the 95^{th} percentiles, it results in a very strict metric. Meaning that if one of the three components has a jump it will be detected. The proposal of considering any artifact is interesting and we propose not to take into account only the arithmetic mean but also other basic statistics such as root mean square, median, standard deviation, minimum and maximum but for all 95^{th} percentile of the ramps calculated from *CIEDE*2000. The metric is used to evaluate the perceived difference between two colors by taking into account the three components L^*, a^*, b^* at once. There is no need to separate them. Evaluating ramps is a good approach and it should be extended not only by plan like proposed by Aristova but in all directions in three dimensions. The idea of estimating the tone clipping introduced by Kim [2] is a good one but the estimated factor has a variable value depending on the content making difficult to have a generic framework.

Description of the proposed metric

Our metric as represented on Figure 4 can be considered as an extension of Green's framework with inspiration from Aristova's framework. Green's framework has been extended to three dimensions while keeping using *CIEDE*2000. The aim is to apply the metric and to get an estimation of the smoothness for any gradients.

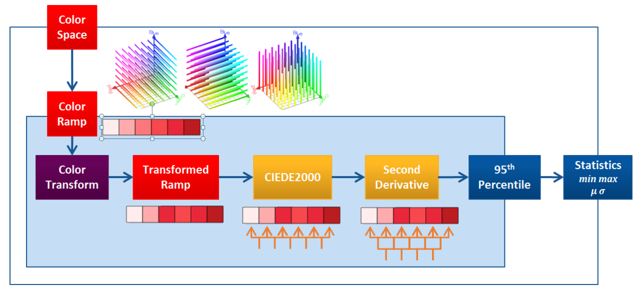

Figure 4. *Proposed framework.*

Therefore all gradients in any direction are taken into account. For that, based on the three dimensional representations of the color transform from *RGB* to $L^*a^*b^*$, in the three main directions of the Red, Green and Blue axes, $L^*a^*b^*$ ramps are computed.

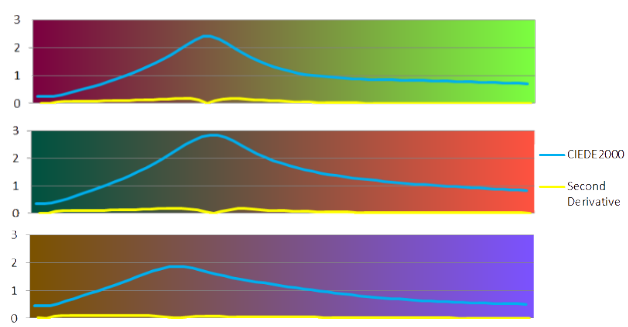

Figure 5. *Example of color gradients used to compute the global smoothness. The blue plots represent the* $CIEDE2000$ *difference color-to-color along the ramp for the ideal sRGB gamut and the yellow plot represents the variations of* $CIEDE2000$.

Figure 5 illustrates the process on three different color gradients. For each ramp, *CIEDE*2000 is computed between two consecutive $L^*a^*b^*$ triplets along the ramp, resulting in a $(n-1)$ ramp. *CIEDE*2000 values are represented by a blue plot for the three example gradients on Figure 5. From there, a second derivative (yellow plots on Figure 5) is approximated by the absolute difference of two consecutive elements of the $CIEDE2000$ ramp, resulting in a $(n-2)$ ramp of positive values. This value represents the smoothness of the color transform for the input ramp and is noted α_i where i is the index of the input ramp.

This is repeated in the three directions, and three 2D-tables (one for Red, one for Green and one for Blue directions) are obtained from which basic statistics can be calculated: arithmetic mean, root mean square, median, standard deviation, minimum and maximum of the combined tables can represent the global smoothness. Considering the fact that a perfectly smooth color transform would result in exactly null second derivatives along any color gradient, the sum (or the arithmetic mean) of α_i must be as low as possible. However, reducing the global smoothness to the average of the ramp smoothnesses may be inefficient to represent punctual discontinuities, especially for fine granularities in which a single discontinuity would be hidden. Standard deviation would clearly reveal such a unique discontinuity, but it would not reflect a uniform difference between two transforms. For this reason, the root mean square appears as a better candidate as it combines both aspects. Figure 13 and section 'Detection of Artifacts' illustrate the fact that some statistics are more sensitive than others to such punctual discontinuities. For this reason, it remains interesting to consider other statistics to have a better view of the situation.

The new presented metric is a three dimensional extension for covering the full gamut of the transform. Nevertheless the precision of the metric should not be sacrificed to improve the computation time. It depends on both the number of considered color gradients or ramps and on the number of colors composing the gradients. These two parameters can be combined into one single parameter called granularity 'g'. In other words, the number of color composing the gradients is constant to all ramps. The chosen approach is actually to select a sample of evenly spread RGB triplets and build the color gradients from this sample. This way, both the number of ramps and the number of colors per ramp are defined by the sample granularity. The number of colors per ramp is given by the granularity 'g'. The number of steps per ramp 's' is equal to $g-1$. The number of directions is given by the parameter d and its maximum value is 3. The total number of ramps 't' is equal to $d.n^2$. All the previous parameters are depicted on the Figure 6.

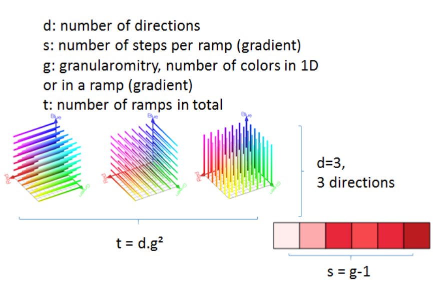

Figure 6. *Parameters of the proposed metric.*

Influence of granularity

As explained above, the defined metric is dependent on the number of ramps 't' and the number of colors per ramp but reduced to one single parameter called granularity 'g'. The number of colors composing the gradients must be as high as possible. Judging the smoothness on a too limited sample of color ramps would not be representative of the overall smoothness of a color transform. The number of colors on color ramps is also important. By having more colors on the gradients, chances to detect some local irregularities are increased. Also important to be noticed, the resulting smoothness amplitude depends on the number 'g'. With a high granularity value, the smoothness will have a tendency to have a smaller amplitude whereas with a small granularity or a limited number of colors composing the ramps (gradients), the amplitude of the smoothness will be larger. This is depicted on the Figure 7, where the y axis is the value of smoothness(based on root mean square) for two color spaces: *AdobeRGB* and *sRGB*, and the x axis is the granularity g' parameter with a value from 5 to 256. The number of steps or colors directly affects the $CIEDE2000$ values calculated along the ramps, and so their derivatives and therefore the resulting statistics and finally the smoothness.

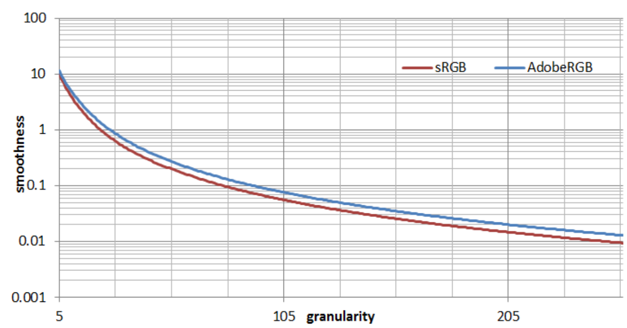

Figure 7. *Smoothness (root mean square) of sRGB and AdobeRGB color spaces as a function of the color sample granularity. The color space smoothness is expressed as the arithmetic mean of the gradient smoothnesses.*

To summarize, a finer granularity gives more accurate results but the computational cost is more important. A coarser granularity allows faster computation but may act as a low-pass filter and hide local artifacts. As mentioned above, the granularity has a direct impact on the absolute results of the proposed metric. Although, as the metric may be used to compare different color transforms or gamuts, it is important to figure out if the granularity can affect the conclusions of such comparison. Figure 8 presents the ratio of *AdobeRGB* smoothness versus *sRGB* smoothness for different granularity values from 5 to 256. Different statistics are given on the Figure: arithmetic mean, root mean square, median, standard deviation, minimum and maximum. One can notice the ratio is almost constant for the different statistics which could be used to define the smoothness as long as the granularity is superior to 50.

Results and validation

The method described above has been used to evaluate the smoothness of different well defined color gamuts such as *sRGB* [5] and *AdobeRGB* [6]. Table 1 presents a set of statistics resulting in the analysis of both color gamuts with a granularity of 52. These results are in line with Figure 8.

The data is also represented in boxplot in Figure 9. It is noticeable that *AdobeRGB* always present higher results than *sRGB*.

Color smoothness of *sRGB* and *Adobe RGB* color spaces as they are defined for a granularity of 52 with basic statistics: real minimum, real maximum, median, arithmetic mean, root mean square and standard deviation.

	sRGB	*AdobeRGB*
Minimum	0.0247	0.0294
Maximum	0.9788	1.2126
Median	0.1211	0.1457
Arithmetic mean	0.1709	0.2235
Root mean square	0.2206	0.3021
Standard deviation	0.1395	0.2032

This is mainly due to the fact that *AdobeRGB* gamut is wider than *sRGB*. Thus an equivalent granularity necessarily ends up in larger differences between colors.

Detection of artifacts

The capacity of the metric to detect artifacts has been studied more in detail. This is done by introducing some punctual discontinuities in an *sRGB* color space and comparing the resulting smoothness with the default *sRGB*. Only one artifact is introduced at a time. Artifacts of different sizes and colors can be

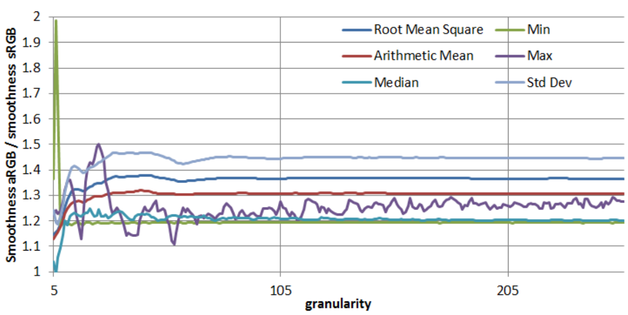

Figure 8. *Ratio of Adobe RGB smoothness to sRGB smoothness for different statistic methods versus the color sample granularity.*

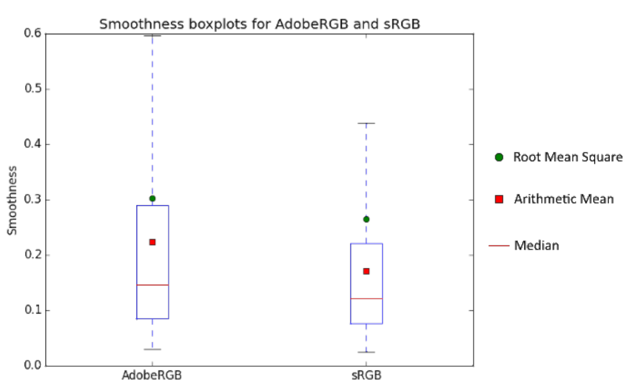

Figure 9. *Box plots of the smoothness data of Table 1 for sRGB and Adobe RGB color spaces. The minimums and maximums are re-calculated excluding the outliers. The outliers are not displayed.*

introduced. In the present example the artifact consists in a modification of the hue (from *HSV* representation of the *RGB* cube) of a cube of size $8*8*8$ centered on $(123;82;65)_{RGB_{256}}$ within a *RGB* cube of resolution $256*256*256$.

Figure 10 shows an example of an artificial artifact introduced at the intersection of the color ramps already presented by Figure 5. The artifact is shown on three different plans of the *RGB* cube.These plans are orthogonal and intersect at the center of the artifact. On Figure 10, each plan is composed of $256*256$ colors and the artifact is a cube of $8*8*8$. Of course, on the plan the size of the artifact is reduced to $8*8$.

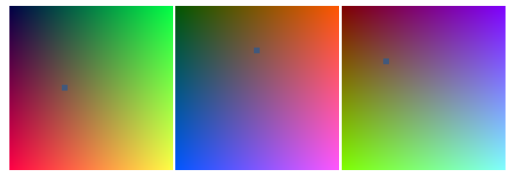

Figure 10. *Three plans extracted from a* RGB *cube used to evaluate the effect of an artifact on the calibration. The plans are orthogonal and intersect at the position of the artifact.*

Figure 11 presents the three color gradients already shown on Figure 5 after the introduction an artifact. The plots still correspond to a *sRGB* gamut, and the granularity is $g = 64$. A peak is clearly visible on both $CIEDE2000$ and its derivative at the artifact's position.

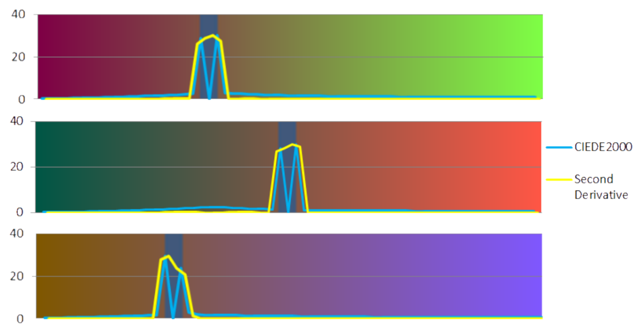

Figure 11. *Three plans extracted from a* RGB *cube used to evaluate the effect of an artifact on the calibration. The plans are orthogonal and intersect at the position of the artifact.*

Figure 12 corresponds to the relative variation of the smoothness with the introduced artifact compared to the original smoothness of *sRGB* as a function of the granularity. It shows that the artifact induces a rise of the smoothness value obtained thanks to the proposed method. it also emphasizes the importance of the granularity. Indeed, a peak is visible on the plot for a granularity $g = 64$, corresponding to twice the spatial frequency of the artifact: $\frac{256}{8} = 32$. This corresponds to the Nyquist theorem [7], indicating that a signal can be correctly detected only if the sampling frequency is at least twice as high as the signal's frequency.

Studying the effect of an isolated artifact also allows to emphasize the importance of the different statistics obtained from the list of ramp smoothnesses α_i.

On Figure 13, one can observe how the artifact studied above affects different basic statistics. It is clearly noticeable that the

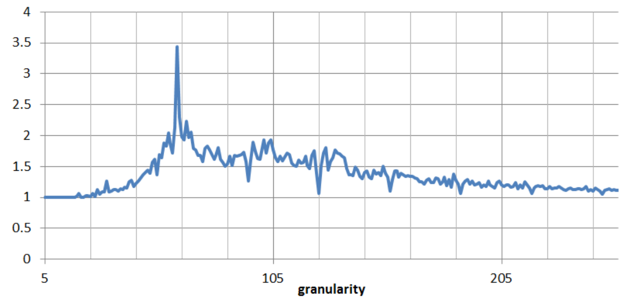

Figure 12. *Relative variation of the smoothness (root mean square) with the artifact introduced compared to the original smoothness of* $sRGB$ *as a function of the granularity.*

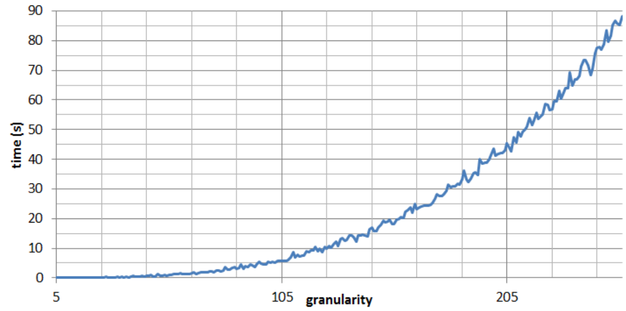

Figure 14. *Computation time of the smoothness of sRGB gamut as a function of the granularity.*

arithmetic mean and the median are unable to detect punctual variations, while the standard deviation is much more fitted to do so. Root mean square presents variations similar to standard deviation although they are less pronounced. Of course the maximum is clearly the most impacted by the artifact, but it would not be affected by a second artifact of lower amplitude, while it would be the case of root mean square or standard deviation. Therefore for keeping one statistic, we recommend to use the root mean square with a sufficiently high level of granularity.

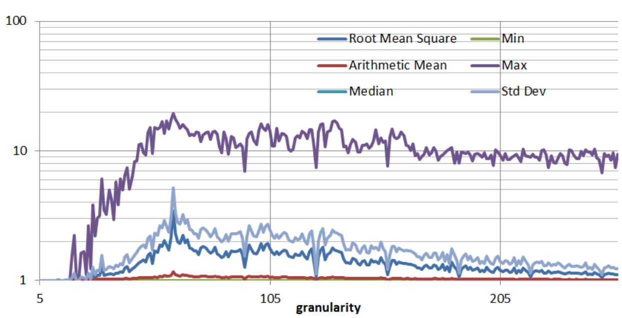

Figure 13. *Relative variation of the smoothness (represented by different statistics) with the artifact introduced compared to the original smoothness of* $sRGB$ *as a function of the granularity.*

Computation time

Figure 14 presents the computation time as a function of the granularity for the proposed method. The metric has been implemented in C++. This estimation of time takes into account the sampling of the color space, the color conversion from RGB to $L^*a^*b^*$, the computation of the different ramp smoothnesses and the statistical analysis.

This bench test was performed on a Quad-Core AMD Opteron Processor 2384 (2.70GHz). According to the method description given previously, the complexity of the algorithm is $\Theta(g^3)$, although it could be improved and optimized by the use of parallel programing techniques.

Application and discussion

The metric described in this document has already been used successfully several times. The first one is also the origin of the metric development [8]. It is a method for making a monitor compliant with the required standard for medical applications (DICOM)[9]. This method makes use of ICC profiles, and metrics are used to assess the quality of the final calibration, including color interpolation from the different LUTs.

Moreover, the smoothness metric has been reused in [10] as a Quality Assurance method for the Color Standard Display Function (CSDF).

The present manuscript describes a metric to measure the smoothness of a color transform or color gamut by studying the perceptual color differences along a set of color ramps. These ramps are defined as orthogonal lines along the RGB representation of the color gamut, following one of the three axes. This configuration has been chosen as the easiest way to build ramps having all the same number of steps equal to the sample granularity. However one could imagine adding some ramps ringing around the achromatic diagonal, or parallel to it.

It is also possible to focus on a sub-part of the entire color space for some applications which do not make use of all available colors. For instance in the case of quantitative imaging applications which use color scales to spatially represent some numerical data. Such applications often uses color scales containing a limited number of colors covering a more or less well defined portion of the color gamut. For such application, the smoothness of colors outside this portion of the color space can be considered as irrelevant.

Conclusion

A new metric for evaluating the perceptual smoothness of any color transform has been described and implemented and illustrated thanks to several applications. The next step for the entire validation of the metric would be a psychophysical study with human observers for proving the correlation between humans and the metric.

References

[1] Phil J. Green, A smoothness metric for colour transforms, London College of Communication (United Kingdom) SPIE Proceedings Volume 6807: Color Imaging XIII: Processing, Hardcopy, and Applications, (2008).

[2] Youn Jin Kim, Yousun Bang, Heui-Keun Choh, Gradient approach to quantify the gradation smoothness for output media, Journal of Electronic Imaging, Volume 19, Issue 01, (2010).

[3] Anna Aristova, Zhaohui Wang, Jon Y. Hardeberg,Evaluating the smoothness of color transformations Gjovik Univ. College (Norway) SPIE Proceedings Volume 7866: Color Imaging XVI: Displaying Processing, Hardcopy, and Applications, (2011).
[4] Ming Ronnier Luo, Guihua Cui, B. Rigg, The development of the CIE 2000 colour-difference formula: CIEDE2000, Color Research & Application, Volume 26, Issue 05, (2001).
[5] IEC61966-2-1:1999, Multimedia systems and equipment - Colour measurement and management - Part 2-1: Colour management - Default RGB colour space - sRGB., IEC, (1999)
[6] Adobe Systems Incorporated, Adobe RGB (1998) Color Image Encoding, Version 2005-05, (2005).
[7] Harry Nyquist, Certain topics in telegraph transmission theory, Transactions of the American Institute of Electrical Engineers, Volume 47, Pages 617–644, (1928). Reprinted in Proceedings of the IEEE, Volume 90, Issue 2, pp 280-305, (2002)
[8] International Color Consortium, White paper #44: Visualization of medical content on color display systems (2016).
[9] NEMA, Digital Imaging and Communications in Medicine (DICOM), part 14: Grayscale Standard Display Function, PS 3.14-2004, standard, (2004).
[10] Tom Kimpe, Johan Rostang, Gert Van Hoey, Albert Xthona, Color Standard Display Function (CSDF): a proposed extension of DICOM GSDF, Medical Physics, Volume 43, Number 9, pages 5009-5019 (2016).

Author Biography

Cedric Marchessoux received the Ph.D. degree in electronics engineering from the University of Poitiers, France, in 2003. After a postdoctorate at Agfa-Gevaert, Belgium, he joined in 2006 the Barco healthcare division, Kortrijk, Belgium. He is co-authors of more than 60 publications in journals and conferences. He is co-authors of several patents and reviewers of different scientific journals. Since 2004, he has been involved in several international and European projects.

Johan Rostang received his engineer degree of digital imaging and computer vision from Télécom Saint-Étienne, France, in 2011. Since then he has joined the R&D department of Barco healthcare division, Belgium, as a member of the Technology and Innovation Group. His work was focused on optimizing the use of color information in different medical imaging modalities.