A Methodology for Analyzing and Improving Color Models

Timon Braun

Keywords: prepress, color model, color processing

Abstract

In the pre-press and reproduction industry, colorimetric color models play a big role in processing and manipulating colors. However, even the most accepted standards of such models still face issues with correctly representing colors with more consistent perceptual properties. Therefore, this paper presents a methodology for analyzing color models, to help develop a new color model with more consistent perceptual properties. Applications of such a model, like color processing, color management and color manipulation, should have improved accuracy. A crucial aspect of such a color model is consistent and defined conversions for common data exchange spaces, like CIELAB.

An important part of such a methodology is a way to evaluate models. This will be done by comparing set pairs of colors with given properties, in terms of lightness, chroma, and hue. This method can analyze how well a color space reflects the expected given properties.

1. Introduction

Image is everything. Most of our mental activity is devoted to our sight, processing everything our eyes catch. In today's society, this includes hundreds, if not thousands of ads, that we see every day. To make sure that an image stays in someone's mind, it has to appeal to the eye, and much of this appeal hinges of the appropriate use of color.

Studies were done to measure the responses of our eyes to different colors, mapping the stimula- tion of our eyes to the corresponding wavelengths. Based on this data, systems for color measurement and comparisons have been created since 1931. While the studies have been shown to be accurate since then, and some of the

New College of Florida

first color models are still frequently used to this day, the color systems have various problems with accurate representation of human color vision. Human color perception is highly nonlinear, as it changes based on the viewing conditions, like the brightness and hue of the illuminant, contrast surrounding the viewing location, and other factors. These many complex mechanisms of our color perception need to be accounted for by a color system, and all current color models fall short of the requirements of the printing industry.

2. Current Issues

No color model can truly match the exact human perception. Therefore, color models attempt different methods to reflect the human perception as accurately as possible. However, there are many common problems amongst all of them, which will be discussed in this section. The CIELAB color model will be the main focus, as it is the most established in the color management industry. While CIELAB has many desired properties, there are some notable flaws, mainly inconsistencies between the numerical representation and human perception of colors.

2.1 Hue Constancy

The biggest issue of the CIELAB color model is it's lack of preservation of perceived hue. This is most visible with blue colors, which tend to turn purple with reduced chroma [Mor99]. In an attempt to address this issue, [BFE98] provides two hue-linearized versions of the CIELAB color model using Lookup Tables. While these versions did perform better in the blue region, the CIELAB color model performed just as well, or better in other regions. The IPT color model managed to address this problem well, however prediction of colorfulness and lightness did not perform well enough to replace the CIELAB model [EF98]. IPT is mainly used for transformations since these most benefit from the improved hue constancy.

A new color model has been developed in 2020, called Oklab. It attempts to combine the hue calculation of the IPT color model with the lightness and chromaticity calculations of the CIECAM16 model, an appearance model based on the CIELAB color model. However, no research seems to have been done on this color model yet, so it's properties are still to be determined.

2.2 Color Appearance Models (CAM)

In an attempt to improve the aforementioned issues of color models, as well as encompassing more aspects of the human color vision, CAMs were developed. The requirements for a model made by Hunt, at the Vienna symposium, are 12 principles for consideration in establishing a model. These principles were used as the guiding rules in the formulation of CIECAM97s, which performs as well as, if not better, than any previously published CAM according to a wide range of experimental data in the testing conducted in [TC98]. However, these CAMs still have critical issues, as can be seen most obviously by CIECAM97s having the same hue constancy issue as CIELAB [Ebn98]. Colors represented in a CAM take the viewing conditions into consideration as well. This limits their ability to compare colors directly, due to the added constraints for differing viewing condition.[1]

Aside of the above mentioned issues, most importantly a color system needs to be usable. Due to the added inputs required and overall algorithmic complexity, the CIELAB space is still more established in the industry.

2.3 Color Difference Equations

Color difference formulas address the inaccuracies in a color space through formulas, which provide a more accurate value than the euclidean distance between two colors, as mentioned in ??. This already leads to the first issue, difference formulas only return the absolute distance between colors, not how they are different, therefore rendering them effectively useless when trying to modify a color. Additionally, as seen in the CIELAB space, small differences using δE^*ab do not correspond to small differences in human perception, and vice versa [Min98].[2] This is due to the color discrimination threshold of the human eye greatly differing from the range of color differences. Thus, the CIE 2000 Color Difference Formula defines a calculation so that the color difference calculated by color meters becomes close to the color discrimination threshold of the human eye on the solid color space of CIELAB. Difference formulas are only effective for small color differences, becoming inaccurate at medium distances already. The Optical Society of America's Committee on Uniform Color Scales has shown that no space with uniform scale for large color differences exists [Mac74].

3. Methodology

A color model should perform well in different aspects, which might be competing. Therefore, one perfect color model may not be possible, instead it may be more suitable to provide modified color models for different purposes. As an example, in process control, a color model for reproduction may require different properties than a metric to measure the difference between two given colors.

This paper focuses on the application of color management for color conversion between different color spaces. In this context the main application is replacing colors that cannot be reproduced by a best matching alternative.

In the color management industry, CIELAB is of major importance. It's also the only colorimetric color model used for archiving image data, like TIFF format. In general, it does a reliable job as an exchange format. It is also accepted as a metric for color differences: dE-76 is the Euclidean distance in CIELAB, for general usage. Major problems with CIELAB are hue shifts, non-linear chroma perception, and the data range allowing L,a,b coordinates to not reflect real, existing colors.

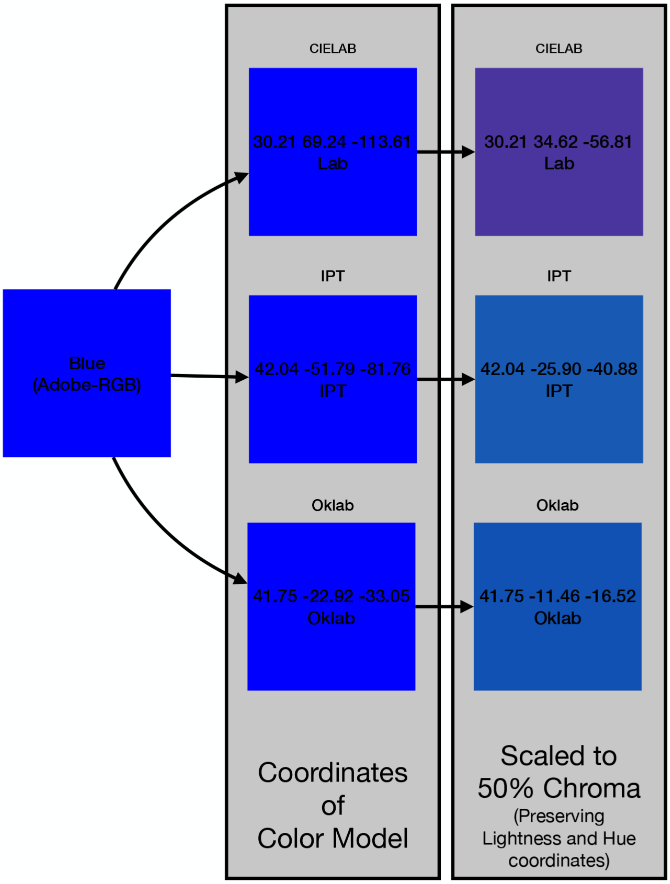


Figure 1: *Demonstrating the color tone change of certain color models. The blue tones of boxes visualize their color values.*

3.1 Color pairs with given/expected properties

To quantify this hue shift, color sets defined to have the perceptually same color hue can be used. We can compare the colors in such a set to calculate the difference between them in the color model we would like to test. If we now match the lightness and chroma of the colors in that set, this difference will correspond to the perceived difference in hue according to the color model being tested. Since all colors in the set should have the same perceptual hue, this reflects the difference between perceptual hue and the hue according to the color model.

An Algorithm takes the values of two colors as an input and uses a given color model to calculate the error between the color model and the perceived colors, according to the data set. This can be done by converting the colors using a chosen color model, matching different properties (in this case Lightness and Chroma), and converting back into a known color system, like CIELAB, for evaluation. This is shown in Figure 2. Assuming the color model has consistent properties, the colors should be the same because the unknown properties (Lightness and Chroma) were adjusted and the matching property (Hue) was kept unchanged. The resulting error from such a color model would be small.

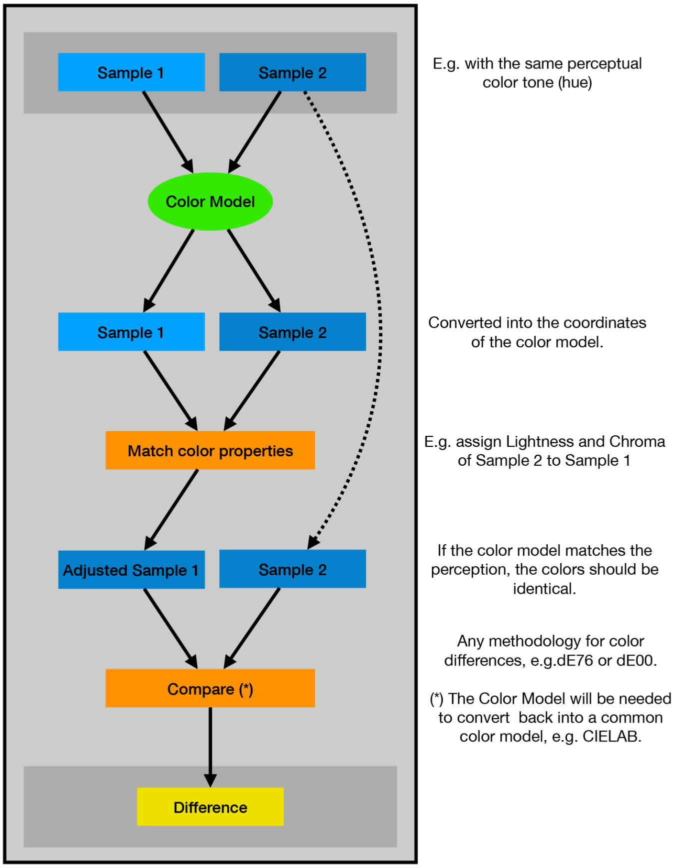


Figure 2: *The method of testing hue consistency of a color model. The blue tones of boxes visualize changes.*

Color models may evaluate color properties slightly different, for example, Lightness in CIELAB may be slightly different than in other perceptual color spaces. This may have technical reasons, like limitations in formulas, or following a different methodology. Typically these differences are rather small compared to the Hue. Therefore it may be worth considering eliminating the Lightness and Chroma from the difference calculation. This is possible because most difference calculations separate these properties anyway, which is visible in the calculations for the CIE 2000 Color Difference Formula. For the same reason, this may be ignored. Indeed adding or removing this from the difference calculations did not have a noticeable impact on the results.

Now we can use common color difference formulas to check for the difference. It should also be noted that color difference formulas are typically only valid for smaller color differences and real colors. Simple metrics like deltaE76, the Euclidian distance in CIELAB, may be most reliable.

$C'_i = f(C_i),\ i = \{1, 2\}$: Converted using the color model
$C''_1, C''_2 = \mathrm{MapLC}(C'_1, C'_2)$: Mapped Lightness and Chroma of the colors, maintain hue!
$C'''_1 = f^{-1}(C''_1),\ C'''_2 = f^{-1}(C''_2)$
δ = Difference Formula (C'''_1, C'''_2) : Color Difference of the converted colors

, where C1 and C2 are two different colors with the same hue.

This approach is straightforward and can be used for both, training and evaluation. Furthermore, one may consider using a portion of the data for training and use the full data set for evaluation. This is a common approach for optimizations.

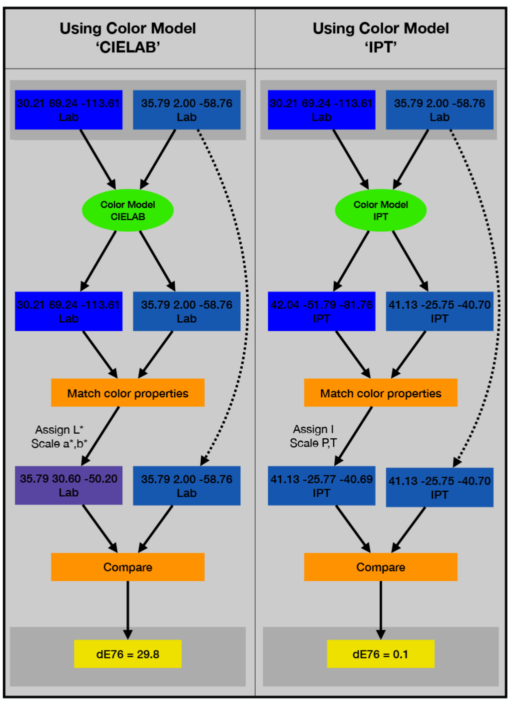


Figure 3: *Example hue consistency test of CIELAB (left) and IPT (right). The blue tones of boxes visualize their color values.*

The bigger problem is to find such data sets. Even-more as a significant amount of samples is required to get stable results. Luckily color scientists provide some data sets. Most common is the Munsell color system and some test data from Ebner&Fairchild. A big advantage of the Munsell colors is that they span the complete space of real colors. Whereas the Ebner&Fairchild data set is using a smaller color space, only.

We use 4 different data sets:

- Ebner&Fairchild[3]: This is the data set with the smallest gamut consisting of printable colors.

- Munsell-real data[4]: This data set is based on Munsell renotation data using the real colors.

- Munsell-all data[5]: This data set is based on Munsell renotation data using all colors within the typical CIELAB boundaries (L in 0..100, a/b in -128..128).

- CIELAB grid: A grid of 11x11x11CIELAB samples in the typical CIELAB boundaries (L in 0..100, a/b in -128..128).

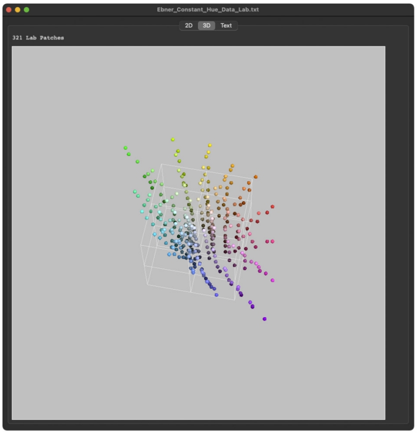


Figure 4: *The Ebner&Fairchild data set. (from ColorLogic Software)*

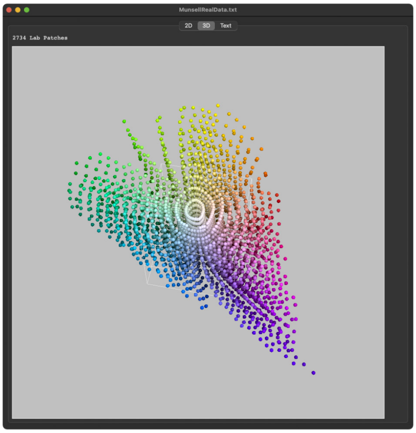


Figure 5: *The Munsell real data set. (from ColorLogic Software)*

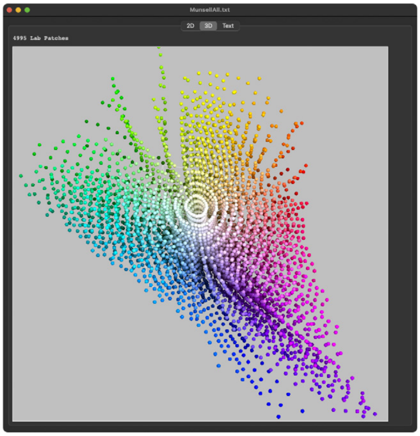


Figure 6: *The Munsell all data set. (from ColorLogic Software)*

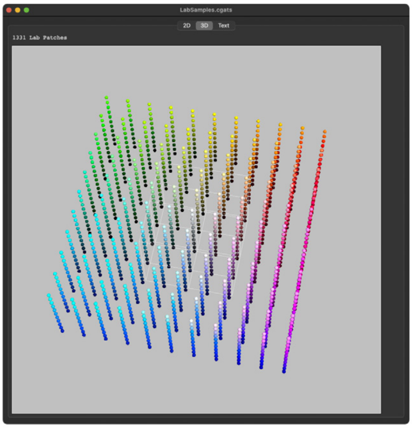


Figure 7: *The CIELAB grid data set. (from ColorLogic Software)*

3.2 The Complete Methodology

A training method for color models can be defined using a set of color values with given properties. Color sets with a perceptually same color hue, can be used to test or train a color model. A program was written according to this training method. The program implements three different functions: The generation of sample data to test the hue consistency of a color model, the test routines to analyze the model properties and the optimization of new color models demonstrating that better color models can be generated.

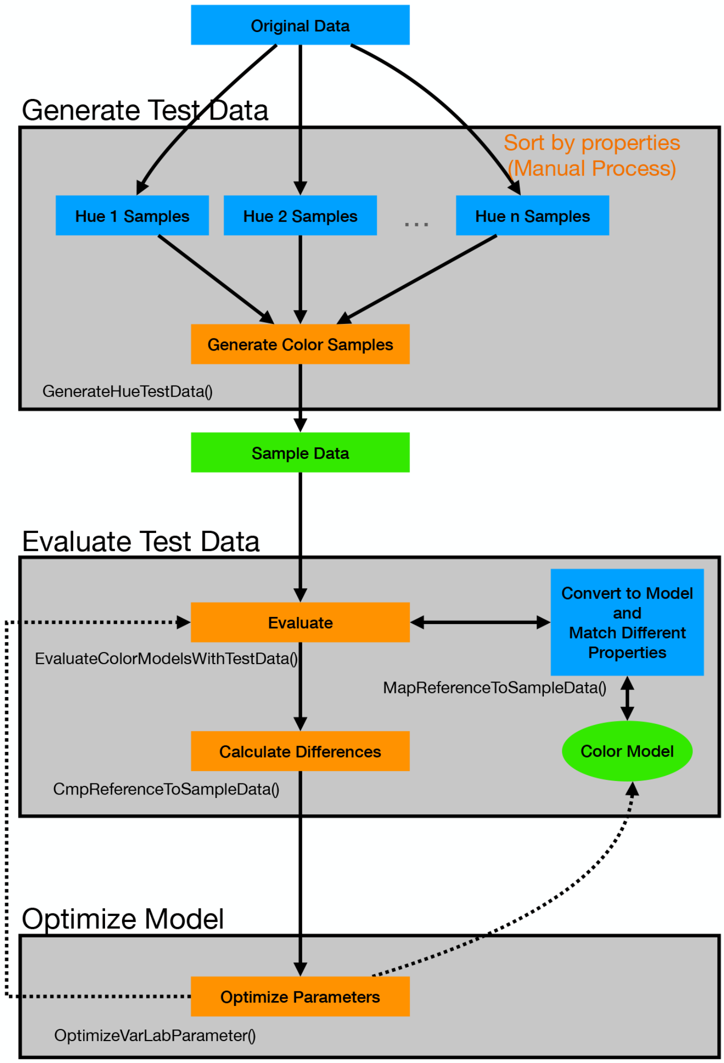


Figure 8: *An overview of the program workflow and components.*

To feed the evaluation with test data, it must be generated from existing data sets. The known published data sets are not formatted for automatic processing. Hence, the data must be separated and prepared so that the program can process it. The used data sets are lists of colors of the visually same color tone. The program will generate color pairs of the most colorful color (in Lab) and the remaining samples. A list of n colors having the same perceptual hue will create (n-1) pairs of samples. The generation of this test data must be calculated once. Then, the test data can be reused for later evaluation, constructing other color models or optimization of new color models.

The main focus of the program is the evaluation of a given color model. The program will process certain test routines for a given color model and will report the results. Such results are the statistical analysis of the accuracy compared to the sample data, and in addition, the stability compared to the input data. Also, it will report the statistical analysis of the stability when varying coefficients of variable color models. This is important for the design of new color models with not yet fixed coefficients. In addition, the program may generate test colors for visual evaluation of a color model or to compare the properties of different color models visually.

The evaluation of existing color models is the key to compare different color models or to identify the best model. The next step would be to optimize existing or new color models to perform better. Obviously, it is necessary to define what ’best’ means. The evaluation (Part 2 of the program) reports numerous statistical attributes which can be used to define the quality/error of a color model. For simplicity, the implementation uses the RMSE of the color tone samples as the measure for quality. Then, it defines three alternative models of slightly different concepts and complexity and uses the evaluation to optimize its coefficients. It can be seen from the results, as demonstrated in the results chapter, that these sample color models already outperform known/published color models in most criteria, also. This can be seen as a proof of concept. The fine-tuning of the models and the clear definition of quality shall be left for separate work.

3.3 Stability

A color model should be consistent, meaning small variations in the sample data shouldn’t result in large differences. Thus, stable functions are very important to have, however the question is how to test stability. It would be preferred to examine the gradients for a given color range. Functions and finding their derivatives may get very complex. A direct method to test for stability is using sample data, varying it and comparing the variation after conversion. CIELAB is a highly accepted color model and the general idea of a better color model is to find something similar without the ’hue- flaws’. Also, CIELAB is used as an exchange format. This is one more reason CIELAB samples are used to test the stability color model. The problem of ’virtual’ colors, CIELAB values not representing existing colors in nature, may be considered as a wanted side effect of this test.

In addition, modern colorimetry also works with scaled data. This is necessary to adapt to different white and black points of different color gamut (see color background). Often this scaling is done in XYZ, using idealized values for white and black. Hence, they may produce extreme values which may exceed natural color boundaries.

It may still make sense to separately test the stability of real colors, like the Munsell samples and of valid CIELAB combinations. Therefore the test was applied to both, a CIELAB grid in the common range and the parameters and real colors, which are well represented by the Munsell data set.

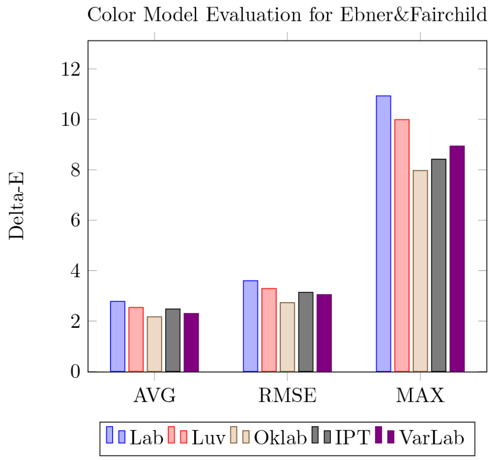
4 Results

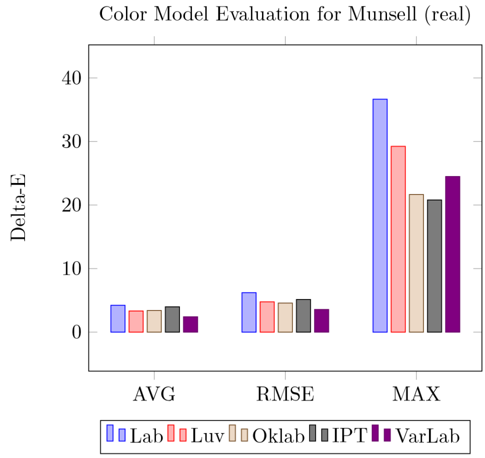
The methodology showed a lot of potential, both in analysing a color model, and finding potential improvements. To demonstrate the effectiveness of the methodology, an example color model, called VarLab, is used. It implements variations into a color model, according to the analysis of the

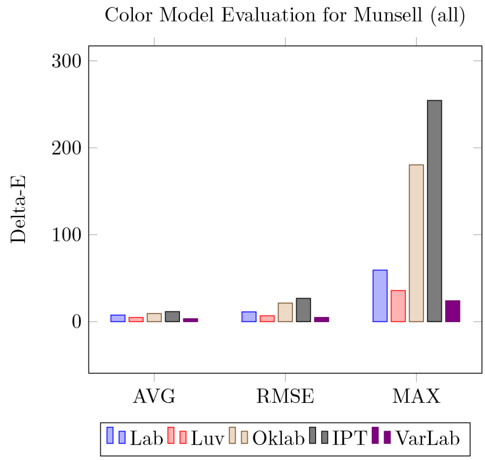
methodology. There are a few different versions of VarLab. However, since the goal of this project is not to create a new color model, but to develop a methodology of improving color models, they are all grouped together.

4.1 Hue Consistency

To compare color models, the average (AVG), root mean square error (RMSE), and maximum (MAX) values were used. Using data samples may lead to the problem of weighting the differences, since a high number of uncritical areas may obscure more critical areas. Other measures, like the 95th percentile, may be considered as well. Different color areas may have different priorities. Natural colors seem more important than extreme, virtual values, however, virtual colors are important to stabilize the color model. For smaller gamuts and real colors, Oklab and IPT perform better than CIELAB and CIELuv. This is not really surprising because they were explicitly designed to improve known hue issues in CIELAB. The bigger the gamut of the samples is, the worse OkLab and IPT perform. Comparing with the Munsell data shows big deviations. It seems that these models were optimized to match certain criteria of natural color data. Overall test models trained with the Munsell-all data performed significantly better. A visual comparison showed that Oklab and IPT seem to be optimized for the blue area, which is known to be most critical in CIELAB.


Color Model Evaluation for Ebner&Fairchild
Delta-E
12
10
8
6
4
2
0
AVG
RMSE
MAX
Lab
Luv
Oklab
IPT
VarLab


Color Model Evaluation for Munsell (real)
Delta-E
40
30
20
10
0
AVG
RMSE
MAX
Lab
Luv
Oklab
IPT
VarLab


Color Model Evaluation for Munsell (all)
Delta-E
300
200
100
0
AVG
RMSE
MAX
Lab
Luv
Oklab
IPT
VarLab

4.2 Stability

As a measure for stability, this paper evaluates the nearby area of real and virtual sample data when converting from CIELAB into a different color model. Most important is how much the gradient varies. Uniform low or high values just indicate a constant relation of CIELAB differences compared to the tested color space. Consequently, the quotient of the maximum and minimum slope is critical.

Therefore, we look at the spread by taking the quotient $\left(\frac{SlopeMax}{SlopeMin}\right)$.

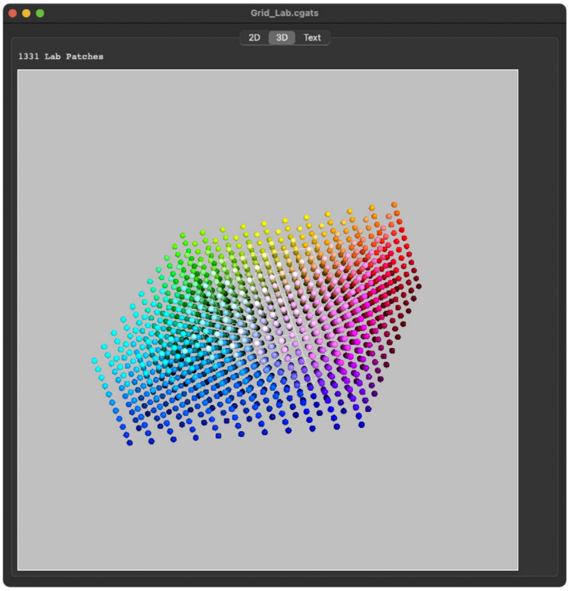


Figure 9: *A grid of CIELAB samples.*

For the typical CIELAB-Range, this gets more extreme, especially Oklab seems to 'explode'. This may lead to major problems when converting data into such a color space, applying some corrections and converting it back into CIELAB data.

It is interesting, that the variations of VarLab do not show these extremes of IPT and OkLab. This is surprising as the concept of the models are similar. The main variation is the exponent of the power functions. In IPT, the exponent is about 2.3, OkLab uses an exponent 3.0, and the VarLab models in the range [1.1, 1.3]. The closer the exponent is to 1, the more linear the function is and flat or steep areas near the origin are reduced.

4.3 Visual Tests

The theoretical color space for real color is extremely large. This makes it very difficult to evaluate color properties because most colors can't be displayed on normal devices. In addition, typical color spaces are of a higher practical relevance

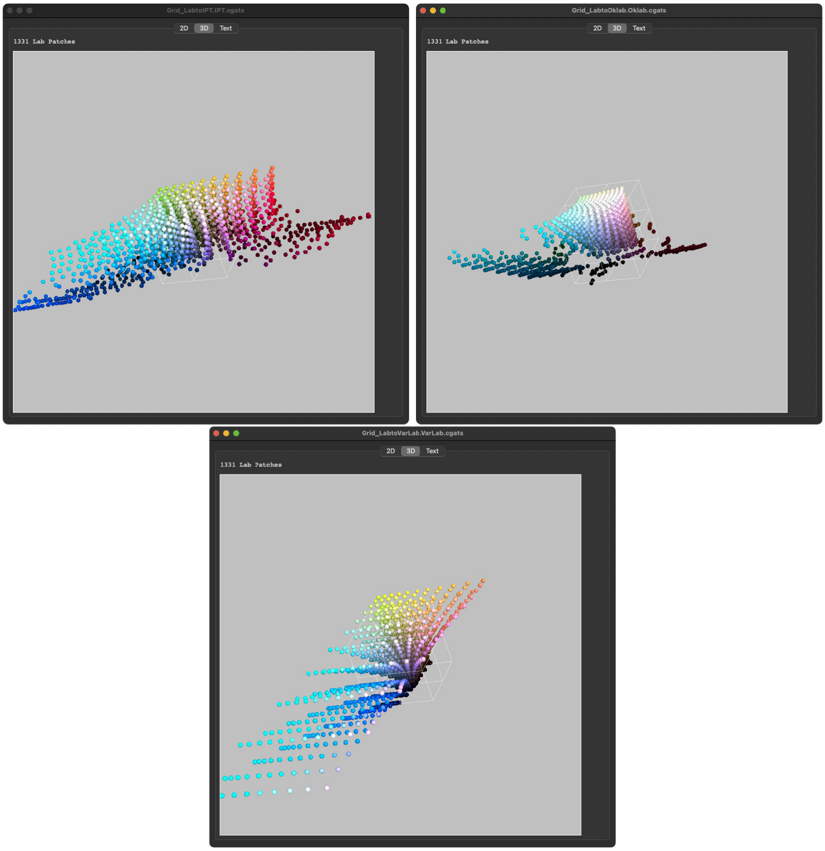


Figure 10: *Converted CIELAB samples to IPT (left), OkLab (right), and VarLab (bottom).*

than the theoretical boundaries. For a basic visual test, a set of colors was used based on a large but common color space, AdobeRGB. The different color models were used to scale the chroma to 50% for each color sample. Then, one may compare the original color samples and the converted color samples on a calibrated monitor. A good color model should maintain the perceptual Lightness and Hue of the colors. Of course, this would require some more scientific evaluation but it should be an indicator how the color models perform on more 'natural' colors.

The data was displayed on a calibrated monitor. It was easily seen that Oklab and IPT perform best in the blue area. However, they also show some questionable hue and lightness shifts in other areas.

Even though the tested color models perform best, they still show severe problems in some (blue) color area.

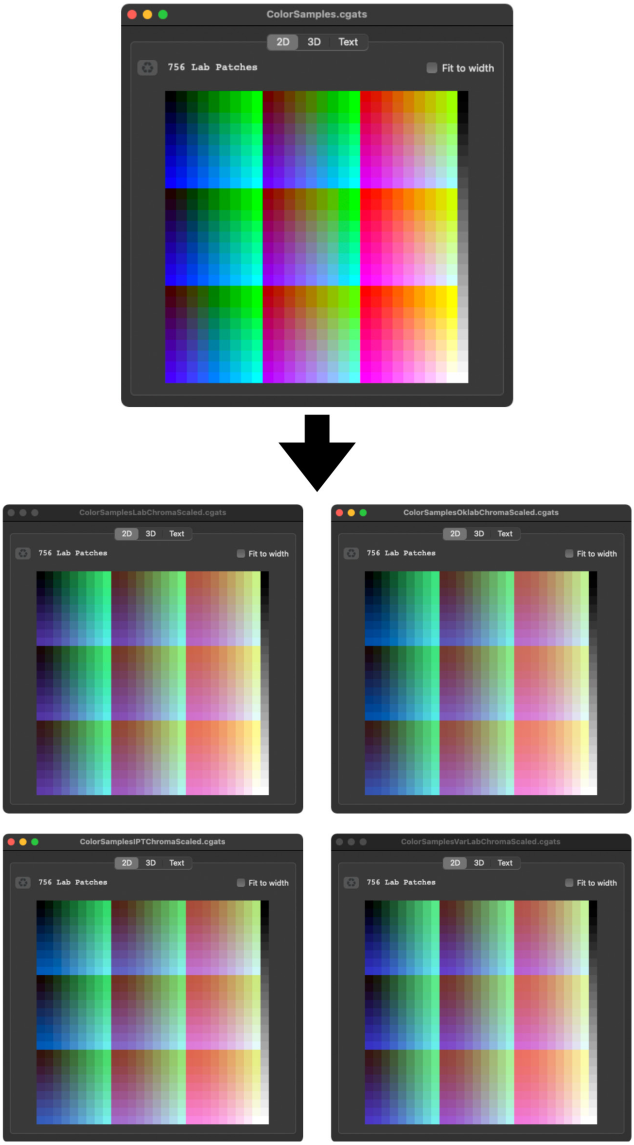


Figure 11: *Sample Colors (top) are scaled to 50% chroma, using the color models CIELAB (top left), IPT (top right), Oklab (bottom left), and VarLab (bottom right). (from ColorLogic Software)*

5 Next Steps

This paper focuses on the main problem of hue consistency and stability of the color models for common color values. It shows that simple models may already improve both. Using mathemati- cal tools, enabling more complex models, improved the hue consistency further, without changing stability significantly.

A problem of instability in all models resulted from the use of a power function with high exponents. They lead to extreme flat or steep curves near the origin. In addition, they spread the data for high values, showing up when using the virtual colors of CIELAB. This may be improved by using more complex curves than a power function.

Optimizing hue consistency is possible, and using data samples is a straightforward approach, which can be used for evaluation or training. The methodology could be transferred to other color properties, however mostly just makes sense for chroma. It would be applicable to lightness as well, but the lightness calculation of CIELAB is already effective. The methodology of IPT and OkLab, using LMS as an intermediate metric, required to alter the lightness. It was shown in the advanced VarLab model that it is possible to maintain the Y/L* methodology. This approach had multiple advantages: It reduces the number of variables of the model, it uses a common methodology for lightness, and it matches CIELAB-Lightness so that a chroma/hue correction maintains the same lightness.
Progressing on this approach will require the following steps:

- Define and agree on data samples. Possibly some weighting can be added, e.g. to reinforce certain color areas.
- Add samples for chroma consistency and uniformity, if required.
- Expand the model, VarLab demonstrates how results can be improved, it may be improved further e.g. by other curves or adding another matrix curve step.
- VarLab demonstrates how to use different curves for Lightness calculation and the other prop- erties. Exponents closer to '1' increases the stability significantly. It has to be evaluated if this also matches human perception better. This is related to chroma samples!
- Usage of more advanced optimization algorithms for the VarLab parameters. The simple implementation used in this paper often finds local minima, only.

References

[Mac74] David L MacAdam. “Uniform color scales”. In: JOSA 64.12 (1974), pp. 1691–1702.

[BFE98] Gustav J Braun, Mark D Fairchild, and Fritz Ebner. “Color gamut mapping in a hue-linearized CIELAB color space”. In: Color and imaging conference. Vol. 1998. 1. Society for Imaging Science and Technology. 1998, pp. 163–168.

[Ebn98] Fritz Ebner. “Derivation and modelling hue uniformity and development of the IPT color space”. In: (1998).

[EF98] Fritz Ebner and Mark D Fairchild. “Development and testing of a color space (IPT) with improved hue uniformity”. In: Color and imaging conference. Vol. 1998. 1. Society for Imaging Science and Technology. 1998, pp. 8–13.

[Min98] Konica Minolta. “Precise color communication”. In: Avaliable online: http://www. koni- caminoltaeurope. com.(accessed on: 21 Sept 2020) (1998).

[TC98] CIE TC. “The CIE 1997 interim colour appearance model (simple version), CIECAM97s”.In: (1998).

[Mor99] Nathan Moroney. “Assessing hue constancy using gradients”. In: Color Imaging: Device- Independent Color, Color Hardcopy, and Graphic Arts V. Vol. 3963. International Society for Optics and Photonics. 1999, pp. 294–300.

1 While a CAM could tell you that two shades of green presented in different viewing conditions will look the same to the human eye, once they are viewed in the same viewing conditions their difference will be noticeable. Many optical illusions invert this effect, presenting the same color in different conditions, making them seem different to the human eye. This can most easily be seen when looking at a white piece of paper under a red light: it will seem to be red. This also applies to much less noticeable differences in viewing conditions, however can still have a major impact.

2 A difference of 1 δE^*ab would be noticeable in pastel colors, however would go unnoticed with more saturated colors.

3 from www.markfairchild.org/files/Ebner_Constant_Hue_Data.txt

4 from www.rit.edu/science/munsell-color-science-lab-educational-resources #munsell-renotation-data

5 from www.rit.edu/science/munsell-color-science-lab-educational-resources #munsell-renotation-data