Perceptual Color Scales for Univariate and Bivariate Data Display

Hongqin Zhang and Ethan D. Montag; Munsell Color Science Laboratory; Chester F. Carlson Center for Imaging Science; Rochester Institute of Technology; Rochester, NY, USA

Abstract

Ten univariate and six bivariate color-encoding schemes were created within the perceptually uniform CIELAB color space. The effectiveness of these color scales was evaluated in three psychophysical experiments. Experiments I and II tested the ten univariate scales and Experiment III tested the six bivariate schemes. Experiments I and III were paired-comparison experiments in which observers judged the utility of the various renderings. Experiment II evaluated the scales by having observers judge the values of indicated points in the images. Experiments I and II demonstrated that the performance of Spectral L and the three diverging color scales were significantly better than the other six. Experiment III showed that the constant hue plane scheme had a better rendering performance than the double cone and cylinder schemes. In both the double cone and cylinder schemes, the narrow hue range performed better than the one with wide range. There was no strong image dependency for univariate scales, but there was for the bivariate schemes.*

Introduction

In scientific data visualization, how data is represented visually has a significant effect on the user's perception and interpretation of the data. It is recognized that the use of color is a powerful technique for enhancing the perception and interpretation of data [1-8]. For example, it is known that people can detect at most a few dozen different intensity levels using grayscale, while varying color in all three dimensions allows the discrimination of hundreds or thousands of different colors. The use of color, therefore, has the potential for providing more insight into the data.

The construction of color schemes is a subtle task and the design process is mostly ad hoc. Some attention has been given to the development of color scales based on perceptual properties since it is often easier and more intuitive for people to separate differences in perceptual variables such as Lightness, Hue, and Chroma than in display red, green, and blue [4-6].

Since the choice of color scales is crucial to the comprehension of the data represented, it is necessary to examine the effectiveness of different color scales. Most often, this evaluation has been subjective rather than based on psychophysical procedures. In this paper, a variety of univariate and bivariate color encoding schemes based on human perception and color appearance are developed and evaluated through quantifiable psychophysical procedures.

Color Scale Generation

Univariate color scales

Ten univariate color scales were designed within the CIELAB color space. The ten univariate color scales, as shown in Fig. 1, are: 1) a gray scale using digital RGB (RGB); 2) a gray scale based on CIELAB L* with the monitor white corresponding to an L* = 100 (L*); 3) an L* scale with constant hue and varying chroma with the maximum chroma as the midpoint (Magenta L*); 4) an L* scale with constant hue and increasing chroma (Yellow L*); 5) an L* scale with both changing hue and chroma (Spectral L*); 6) an L* scale with a red-green component (Red-Green); 7) an L* scale with a yellow-blue component (Yellow-Blue); 8) a divergent red-green scale with maximum lightness at the midpoint (Diverge RG); 9) a divergent yellow-blue scale with maximum lightness at the midpoint (Diverge YB); and 10) a spectrally divergent scale with lightest yellow as a midpoint (Diverge S).

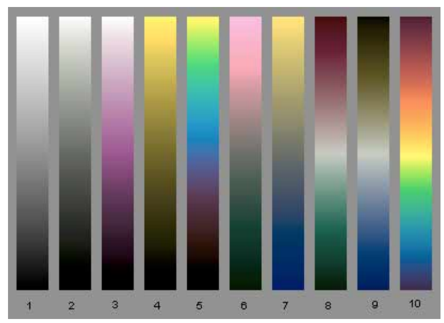

Figure 1. The ten univariate scales generated in CIELAB.

Bivariate color schemes

An important task in scientific data visualization is to present data from multiple sources simultaneously. Bivariate color schemes provide a method for combining two data sets in the hope that the resulting image allows the observer to easily and intuitively interpret these two sources of information simultaneously.

The simplest bivariate schemes use a planes or surfaces that vary along perceptual dimensions in a suitable color pace. Other schemes relax these constraints to achieve a balance between using a wider hue range for conveying separation and increasing lightness for conveying order [7]. Trumbo suggested two types of bivariate schemes [6,7]. The first is a square lying in a plane or curved surface with its principal diagonal along the gray axis as shown in Fig 2a. In this type of scheme, the two variables are not represented by a single perceptual variable. The second type is part of a cylinder surface as shown in Fig 2b. Here, one variable is represented by varying hue with constant lightness and chroma and the other is represented by varying lightness with constant hue and chroma. Univariate information is thus represented by a single perceptual variable, but positive and negative associations between variables are not preserved. In addition to these two types of schemes, a third one was proposed in [6], which consists of a section from a double cone (one converted) as shown in Fig 2c. In this scheme, univariate information is represented by a progression in hue, lightness and chroma combined. Diagonals of positive correlations are represented by constant hues while diagonals of negative correlations are represented by constant lightness and chroma. This is actually a modification of Trumbo's first scheme to include a wider range of hues in order to increase element separation.

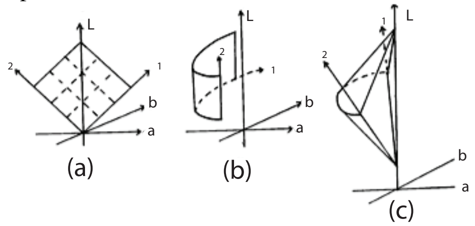

Figure 2. The three types of bivariate schemes.

For this research, six bivariate color scales were designed based on the three types of schemes with some modifications and generated within CIELAB color space. Specifically, the six bivariate encoding schemes, as shown in Fig. 3, are: 1) a square in a constant hue plane with one variable represented by red, the other variable represented by green, and principal diagonal along gray axis (conHue_RG), 2) same as 1), but with one variable represented by yellow, the other variable represented by blue (conHue_YB), 3) a section from the surface of a double cone with univariate axes represented by a hue with changing lightness and chroma (doubleCone_w), 4) same as 3), but a smaller section with narrow hue range (doubleCone_n), 5) a portion of a cylindrical surface with one variable represented by hue and the other variable represented by lightness, here, instead of using a constant chroma, this scheme was slightly modified by making chroma maximum in the middle lightness, (cylinder_w), 6) same as 5), but with a narrow hue range (cylinder_n).

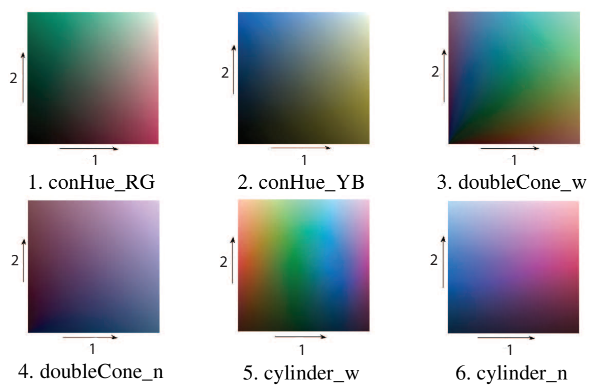

Figure 3. The six bivariate color encoding schemes generated in CIELAB.

Experimental

Experiments I and II evaluated the ten univariate color scales and Experiment III tested the six bivariate schemes. Experiments I and III used paired-comparison in which observers judged the utility of the various renderings. Experiment II evaluated scales by having observers report the values of indicated points in the images. In all of the experiments the color scales were presented to the observers along side the corresponding images. In all experiments the stimuli were displayed on an Apple Cinema HD LCD display with a 20% gray background in a darkened room. The LCD display was carefully characterized [9]; this characterization was used when rendering the images to get the required colorimetric values.

Experiment I

Experiment I was designed to compare the ten univariate color scales using four scientific images which show changes in magnitude or intensity. The four images consist of one digital elevation map (DEM) with 256 levels, one CAT scan medical image (Spine) with 64 levels, one spectral band from remotely sensed satellite imagery (SanFran) with 256 levels, and one material abundance map derived from the analysis of the same scene (AbunMap) with 100 levels, as shown in Fig 4.

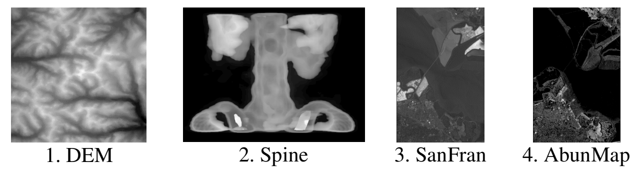

Figure 4. The four images used in Experiment I (shown in RGB Gray Scale).

Each image was rendered using the ten univariate color scales and a paired comparison experiment was then conducted. In each trial, a pair of images rendered by using two different scales was displayed. The observers chose the one image in the pair that provided more useful information. A more useful image was defined as one that allowed easier and more meaningful discrimination of objects and variations within the scene. The order of presentation was randomized. For each observer, there were 10 scales, 4 images, and 3 repetitions resulting in a total of 540 trials.

Experiment II

Experiment II tested how the different univariate encoding schemes affected performance in evaluating the magnitude of points in the images and how the results correlated to those from Experiment I. The ten univariate color scales were tested using the DEM and Spine images. In each trial, a point on the image was indicated by a cross-hair with an open center. The observers' task was to type in the data value of the indicated point. The color scales were displayed alongside the images with indicated intensity values. In order to minimize memorization, the images were presented in random orientations. The order of presentation was randomized. For each observer, there were 10 scales, 2 images, and 3 locations resulting in a total of 60 trials.

Experiment III

Experiment III was a paired comparison experiment designed to compare the six bivariate color encoding schemes using six pairs of simple synthetic images with different surface features, one pair of complex synthetic images, and one pair of material abundance maps as in Experiment I. The four univariate synthetic patterns were constructed using simple mathematical functions. They were chosen to represent different surface properties [5]. The eight univariate images used in this experiment are shown in Fig. 5.

In each trial, a pair of images rendered by using two different bivariate schemes was displayed along with the two univariate images (in RGB gray scale) and the corresponding encoding schemes. The observers chose from the two colored images the one that better represented the information in the two univariate images. The order of presentation was randomized. For each observer, there were 6 scales, 8 images, and 2 repetitions resulting in a total of 240 trials.

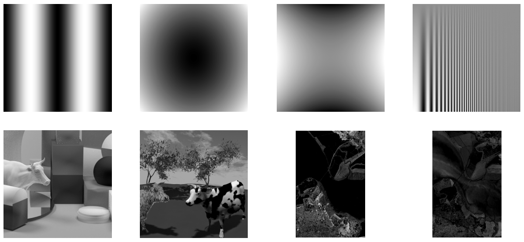

Figure 5. The univariate images used in Experiment III.

Results and Discussion

Experiment I

24 observers with normal color vision participated in Experiment I. The paired-comparison results were analyzed using Thurston's Law of Comparative Judgments, Case V to produce an interval scale of effectiveness in terms of providing more useful information. The interval scale with 95% confidence limits for each of the four images is shown in Fig. 6.

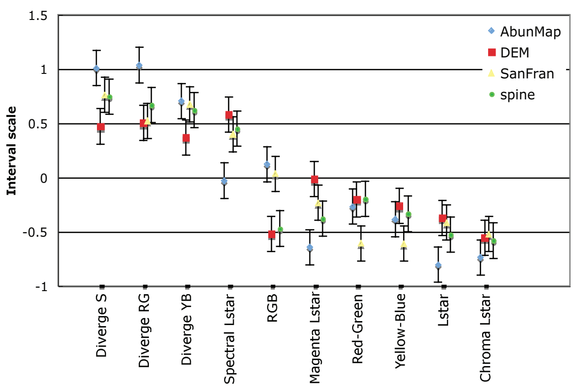

Figure 6. Interval scale of the ten univariate scales for each of the four images

The three diverging scales had the best overall performance but there were no significant differences among them. Spectral L* had comparable performance to the diverging scales on three of the four images except the abundance map. It is known that to achieve an informative representation, it is helpful to maximize the number of distinct perceived colors along the scale while avoiding artificial boundaries. The reasons that these four scales were significantly better than the rest may be attributed to their higher perceived dynamic range (PDR) and contrast. In both the spectral diverging and the opponent channel diverging schemes the chromatic contrast was enhanced. Though lightness was not increasing monotonically in these schemes, the sense of order can be conveyed by the color bar alongside the image. For spectral L*, the monotonic increasing in lightness may give a sense of order and convey surface shape information while achieving wider separation in data values by cycling through a range of hues. RGB had significantly better performance than L* perhaps due to its greater contrast variation.

Experiment II

Experiment II was conducted with 23 observers having normal color vision. The performance was measured by the relative error, which was the absolute value of the difference between the target value and the response value divided by the number of discrete levels of the image.

An ANOVA was performed on the error scores using color scales as the factor for analysis. The analysis revealed significant effect of choice of color scales. In order to determine which color scales were significantly different, multiple comparisons were performed with the error rate controlled conservatively by Tukey's honestly significant difference (hsd) criterion [10].

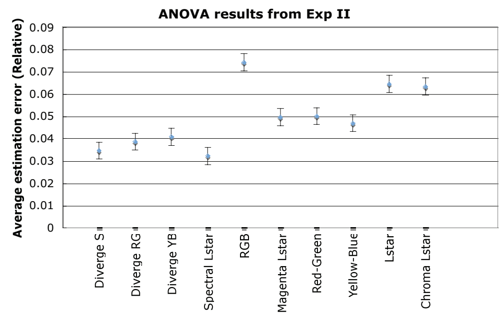

Figure 7. Average relative errors for the ten color scales

Figure 7 shows that Spectral L* had the lowest estimation error on average. This is consistent with the results in [2] and [5]. The three diverging schemes had slightly larger error than Spectral L*, but there are no significant differences among them. The performance of Magenta L*, Red-Green, and Yellow-Blue were comparable to each other, with larger errors than Spectral L* and the three diverging schemes and significantly smaller errors than RGB, L*, and Chroma L*. There was no significant difference between L* and Chroma L*, while RGB scale is the worst.

The main sources are thought to be: 1) The PDR of a scale, that is, a scale with more number of distinct levels would result in smaller errors, otherwise larger errors. 2) Spatial induction effect (simultaneous contrast), that is, the apparent lightness, hue, or saturation of a color can be altered by its surrounding colors quite substantially. This error was expected to be higher for scales using opponent channels. 3) The non-uniformity of a scale, that is, for interval or ratio data, equal steps of data value are not represented by equal perceptual steps along a scale.

Spectral L* has a large PDR since it includes both a large lightness range and plenty of hue variations. It is also less likely to suffer from simultaneous contrast than the scales based on opponent colors. The poor RGB scale performance may be attributed to its small color difference steps along the scale.

The correlation between the interval scale values from Exp. I and the estimation error was calculated. A negative correlation was found indicating that, in general, the scales that were rated as presenting more information had lower errors and a low rating in Expt. I corresponded with higher errors. The correlation between the interval scale values and the estimation error for the DEM image was quite high (R^2 = 0.84) while for the Spine image, the correlation was lower (R^2 = 0.61).

Experiment III

Experiment III was conducted with 15 observers having normal color vision. The paired-comparison results were analyzed as before. The interval scale values with 95% confidence limits for each of the eight images is shown in Fig. 8.

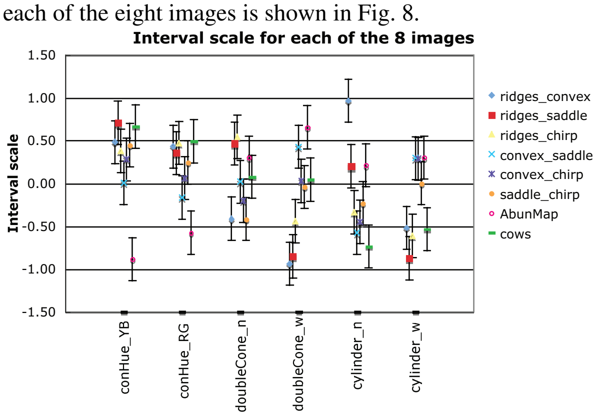

Figure 8. Interval scale of the six bivariate schemes for each of the eight images

Generally, the constant hue plane schemes performed better than the double cone schemes while the double cone schemes were rated better than the cylinder schemes. In the constant hue plane schemes, a good sense of order and surface features or objects shape information can be conveyed, and the data is visually separated into two classes on either side of the principal diagonal which represents the associations between the two univariate variables. Due to the narrower hue range, information may be more clearly presented rather than causing artifacts or confusions from false boundaries caused by rapid hue shifts. In the double cone schemes, the two variables may obscure one another since they are not represented by a single perceptual variable. Though associations may be depicted between variables, with diagonals of positive associations represented by constant hues and diagonals of negative associations represented by constant lightness and chroma, the observers might not intuitively notice these when being asked to choose the colored image that better represents information in both univariate images. The separation benefit of a wide hue range seen in Expts. I and II may be offset by the artifacts or confusions it may cause. With a narrow hue range, this scheme will reduce to a constant hue plane scheme, which improves the overall balance between the univariate information and their associations being conveyed. In the cylinder schemes, each of the two variables is represented by a single perceptual variable, but the univariate information may also obscure each other since the three perceptual dimensions are not independent from each other. This scheme also results in poor associations between variables. Likewise, attention should be paid to hue range utilization.

The performance of the bivariate schemes as judged in this experiment showed a great deal of image dependence. Specifically, for the abundance map, the constant hue plane schemes are judges to be poor while the double cone scheme is judged as more informative. The spatial characteristics of the abundance maps and the synthetic images are very different. It may be the case that the more colorful appearance is eliciting a preference based on aesthetics rather than utility in these complicated images.

Conclusion

The results for Experiment I and Experiment II are consistent and correlated well to each other. Both experiments demonstrated that the performance of Spectral L* and the three diverging color scales were significantly better than the other six. This indicated that in order to achieve an informative representation of data, it is important to have a large PDR, that is, more distinct levels along a scale. It also seems that good contrast may improve the performance of a scale. There was no strong image dependency for univariate color scales indicating some general guidelines may be derived for more effective color scales design.

Experiment III demonstrated that the constant hue plane scheme had a better rendering performance than the double cone and cylinder schemes. It is expected that a good bivariate scheme would be able to convey both sets of univariate information without obscuring each other and also convey the associations between them clearly. The good performance of the constant hue plane scheme should be attributed to the fact that it satisfied these two requirements to some degree and achieved a good balance between them. In both the double cone and cylinder schemes, the scheme with wide hue range was judged to be worse than the one with narrow hue range indicating the utilization of hue should not be abused so as to avoid artifacts or confusions even though hue variations may increase separation between elements as in Expts. I and II. The strong image dependency for bivariate color schemes indicated that there may be no best scheme for all types of data.

References

[1] B Fortner, TE Meyer, Number by Colors: A Guide to Using Color to Understand Technical Data. (TELOS, The Electronic Library of Science, Spinger-Verlag New York, 1997) pg.275-293.
[2] ED Montag, The Use of Color in Multidimensional Graphical Information Display. IS&T/SID 7th Color Imaging Conference, Scottsdale. (1999).
[3] LD Bergman, BE Rogowitz, LA Treinish, A Rule-based Tool for Assisting Colormap Selection. IEEE Visualization, Proceedings of the 6th Conference on Visualization ’95, pg.118. (1995).
[4] H Levkowitz, GT Herman, Color scales for image data. IEEE Computer Graphics and Applications, Vol.12, No.1 pg.72-80. (1992).
[5] C Ware, Color sequences for univariate maps: theory, experiments, and principles. IEEE Computer Graphics and Applications, Vol.8, pg.41-49. (1988).
[6] PK Robertson & JF O’Callaghan, The generation of color sequences for univariate and bivariate mapping. IEEE Computer Graphics and Applications, Vol.6, No.2, pg.24-32. (1986).
[7] B Pham, Spline-based color sequences for univariate, bivariate and trivariate mapping. IEEE Visualization, Proceedings of the 1st Conference on Visualization ’90, pg.202-208. (1990).
[8] CA Brewer, Color use guidelines for data representation, proceedings of the Section on Statistical Graphics, American Statistical Association, Alexandria VA. pp. 55-60. (1999).
[9] EA Day, LA Taplin, RS Berns, Colorimetric Characterization of a Computer-Controlled Liquid Crystal Display, Color Res. Appl., 29:365-373. (2004).
[10] G Keppel, Design and Analysis: A Researcher’s Handbook, 2nd Ed., Englewood Cliffs, NY: Prentice-Hall, Inc. (1973).

Author Biography

Hongqin Zhang is a Ph. D. candidate in the Munsell Color Science Laboratory at RIT. She received a B.S degree in Optics from Shandong University, China in 1998 and a M.S degree in Optics from Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Science in 2001. Her research interests include data visualization and information display, image color appearance, psychophysics, device characterization and color management solutions.