

The Use of Color in Multidimensional Graphical Information Display

Ethan D. Montag
Munsell Color Science Laboratory
Chester F. Carlson Center for Imaging Science
Rochester Institute of Technology, Rochester, New York, USA

Abstract

The performance in judging values in a univariate map encoded using five different color scales was tested in eleven subjects. Digital elevation maps (DEMs) were encoded using: 1) an RGB gray scale (RGB), 2) a gray scale based on CIELAB L* (L*), 3) a L* scale with an added red hue component (Red L*), 4) an L* scale with continuous hue change (Spectral L*), and 5) a gray scale based on luminance (Luminance). Performance was tested using an Evaluation task and a Production task. For both tasks judgments were made both with and without legends for all five encoding schemes. The results show a significant effect of choice of encoding scheme, the presence or absence of a legend, and an interaction between these two factors. Performance with a legend was significantly better than without one. The Spectral L* scale led to the best performance while Luminance encoding was the worst. This experiment is a first step in using quantifiable psychophysical procedures to evaluate the effectiveness of different color encoding schemes on the interpretability of multidimensional graphical images.

Introduction

The last step, and arguably the most important step, of the imaging chain is the visual inspection of the collected image. In many applications of imaging, the goal of the imaging system is a faithful reproduction of the object that has been imaged. For example, in photography, the resultant photographic print should look like the originally captured object. The imaging system may also modify the original appearance in order to enhance the reproduced image for aesthetic purposes. For this type of imaging system, it is desirable to have detectors that match the sensitivity of the human eye.

However in many applications, the role of the imaging system is to transduce what is invisible or unavailable to the eye to a visual representation that can be interpreted by the human visual system. In these types of systems, important decisions must be made in how to represent the collected data in a form that is interpretable by the eye and brain of the end user. This type of information is necessarily multidimensional in nature. At least two dimensions are needed for spatial aspects of the collected data and additional dimensions are added to this framework. Color, defined as that aspect of visual perception that allows an observer to distinguish differences structure-free fields, is a natural tool for displaying this multidimensional information. In order to use color appropriately, we must understand how the visual system interprets images.

The goal of this research is to discover rules and techniques for the display of multidimensional graphical information. In order to achieve this goal we need to consider the psychophysics of color and spatial perception. This knowledge must then be applied using appropriate design techniques in order to produce images that allow the end user to easily and accurately extract the relevant information from an image.

The theoretical implications involved in the design of color scales has been considered in a number of articles (see for example refs. 1-3). However, in most cases the evaluation of these different schemes has been based on subjective judgments of effectiveness rather than by psychophysical evaluation. Although interactive exploration will most likely be necessary for implementing pseudocolor effectively, empirical results will be required in order to help guide the development of these user interfaces. This paper a presents some experimental results demonstrating how the ability of interpreting images can be affected by the encoding schemes used to generate the image.

Experimental

An experiment was performed in order to test how different encoding schemes affect performance in judging points on image maps. The images consisted of two digital elevation maps (DEMs) from the United States Geological Survey. The elevations were encoded using five different scales using color and intensity to represent elevation. The hypothesis was that an encoding scheme that mapped elevation in a perceptually linear manner would improve performance. In addition, the availability of the encoding scheme, by way of a legend, should also affect performance differentially with the different scales. Eleven subjects participated in the experiment.

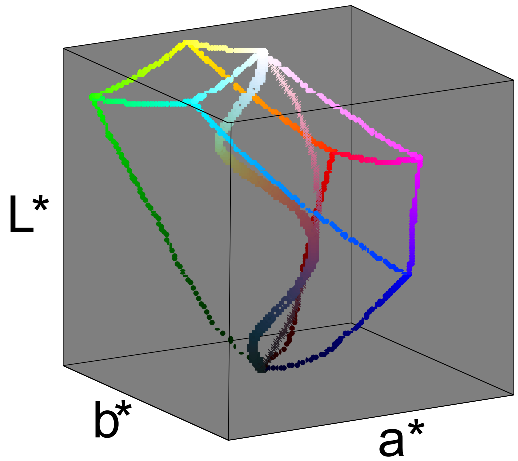

Figure 1. The Red L and Spectral L* color scales shown inside the outline of the CRT gamut in CIELAB space.*

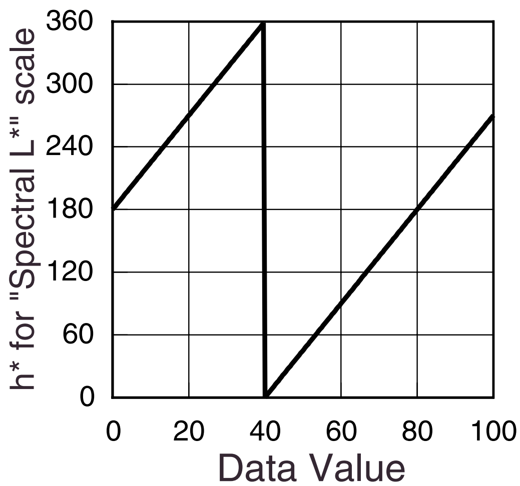

Figure 2. Hue variation in Spectral L color csale.*

Stimuli

The DEMs were presented on a calibrated CRT using the Mapping Toolbox in MATLAB. One DEM was of an area near Rochester, NY, and the other was a more mountainous area near Denver, CO. The gamma of the monitor was approximately 1.6. The elevations in the two DEMs were encoded using five different schemes: 1) a gray scale using digital RGB (RGB), 2) a gray scale based on L* with the monitor white corresponding to an L* = 100 (L*), 3) an L* scale with a constant hue component ((Red L*), see fig. 1), 4) an L* scale with changing hue ((Spectral L*, see fig. 1 and fig. 2), and 5) a gray scale based on luminance (Luminance). Figure 3 shows each of the scales.

Figure 4 shows the DEM for Colorado encoded by the Spectral L* scale. Each map was presented with a white border and a neutral 20% surround (D65).

Procedure

There were two tasks. In the first task (Evaluation) a point on the image was indicated by a cross-hair (which was open in the center so as not to obscure the indicated point). The subjects' task was to type in the value of the indicated point. In the second task (Production) the subjects were presented with a numerical value. The subjects' task was to click on a point corresponding to that value in the image using a mouse. A custom cursor was used that did not obscure the chosen point. The elevations were scaled from 0 (the lowest elevation) to 100 (the highest elevation) which were always encoded as black and white, respectively. If the subject evaluated a point on the image as being 1/3 the maximum value, they would indicate by typing 33. If the subject was to click on a point with a value of 75% they were to choose a point in the image that appeared to be 3/4 of the maximum value. The subjects were not told that these images were DEMs.

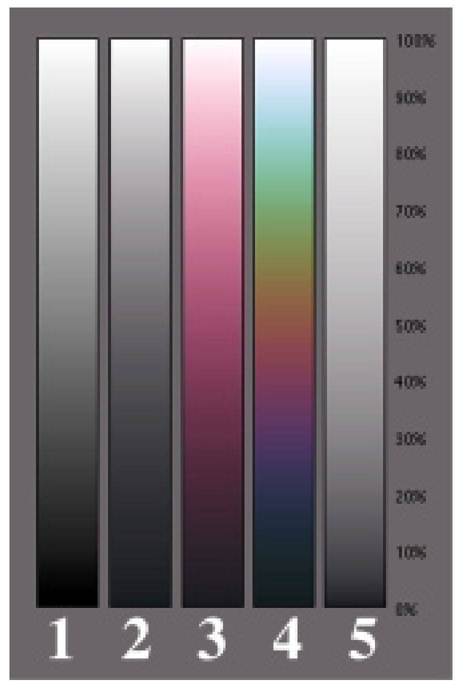

Figure 3. The five scales used to encode the DEMs.

Each task was performed with and without an accompanying legend that indicated how the values were encoded in the image. When a legend was present it appeared as one of the bars in figure 3 with labeled percentages in steps of 20%. In the no-legend conditions, only a white and black patch was displayed showing the maximum and minimum values. The image maps were presented in random orientations in order to minimize memorization. The trials were presented in random order. For each of the 2 maps, 5 scales, 2 presentation conditions, and 2 tasks, there were three trials presented in a factorial design. Therefore each subject completed a total of 120 trials in about one hour.

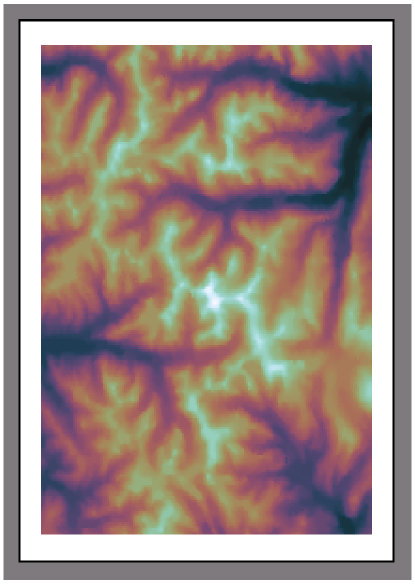

Figure 4. DEM of Colorado encoded by Spectral L.*

Table I. ANOVA of error. Asterisk indicates significance at .05/15 level.

Source	DF	P-value
B (task, Evaluation or Production)	1	0.1002
C (location)	1	0.15387
D (presence/absence of legend)	1	0*
E (encoding scheme)	4	0*
B*C	1	0.00838
B*D	1	0.34445
B*E	4	0.02008
C*D	1	0.02327
C*E	4	.0014*
D*E	4	0*
B*C*D	1	0*
B*C*E	4	0.00303*
B*D*E	4	0.0515
C*D*E	4	0.01083
B*C*D*E	4	0*
Error	1280	

Results and Discussion

The performance is measured by the absolute value of the difference between the target value and the chosen value as shown in equations 1 and 2.

$$\text{Evaluation}: \ error = \frac{|true\ value - response\ value|}{true\ value}$$

$$\text{Evaluation}: \ error = \frac{|true\ value - selected\ value|}{target\ value}$$

An ANOVA was performed on the error scores using task, map location, presence/absence of legend, and encoding scheme as the four factors for analysis. The analysis revealed significant effects of choice of encoding scheme, presence or absence of a legend and an interaction between these two factors. There was also an interaction between choice of encoding scheme and map location as well as some three way interactions and a 4-way interaction. Comparisons between conditions were made controlling for error due to multiple comparisons. The results are shown in Table I.

A summary of the results is plotted in figure 5. The error is plotted in reverse order on the ordinate so that higher points on the graph indicate better performance. Non-overlapping error bars indicate significant differences.

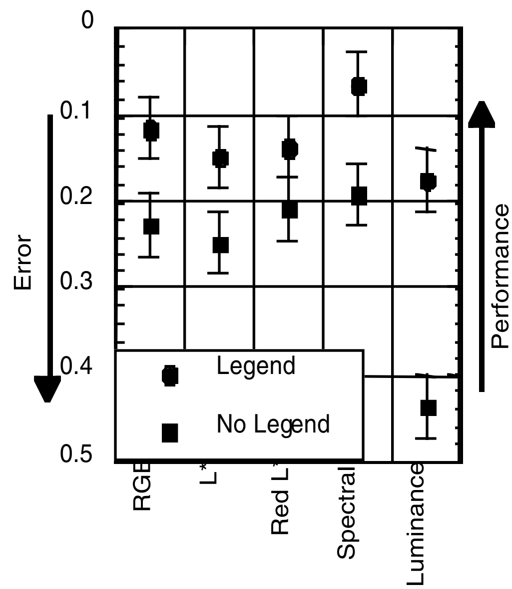

Figure 5. Errors for the five encoding schemes with legend present or absent. Errors are averaged over the two tasks.

The availability of a legend improves performance. Although this result may seem trivial, the implications are quite important. Tufte[4], for example, lists as one of his principles of graphical excellence that complex ideas should be communicated with clarity, precision and efficiency. The reliance on a legend may indicate a lack of efficiency. That is, the process of interpretation of the graphic may be verbal rather than visual. The viewer may be scanning back and forth between the legend and the map using verbal decoders and mental phrases to match colors to values. What we would like is an intuitive, visual interpretation in which relative map values can be gleaned without recourse to a legend.

With no legend, Luminance encoding is far worse than the other methods. Luminance encoding would correspond to a monitor gamma of 1.0. An RGB gray scale on a monitor with a gamma of approximately 1.6 improves performance over the Luminance gray scale without a legend. A monitor gamma of 2.2 corresponds closely to the L* encoding. Performance in judging image values is therefore dependent upon the system's gamma curve when using gray scales based on RGB.

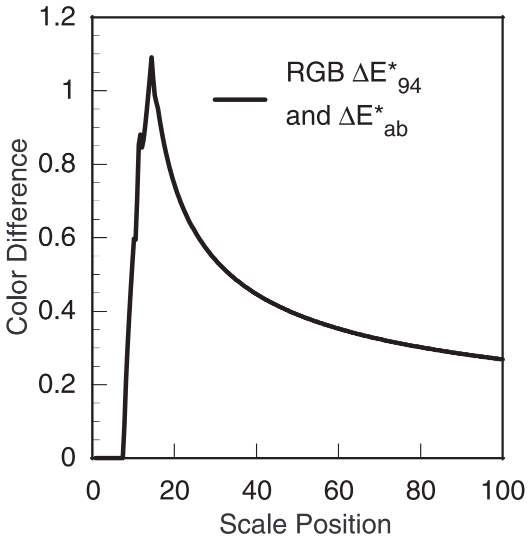

Figure 6. Color differences along RGB scale.

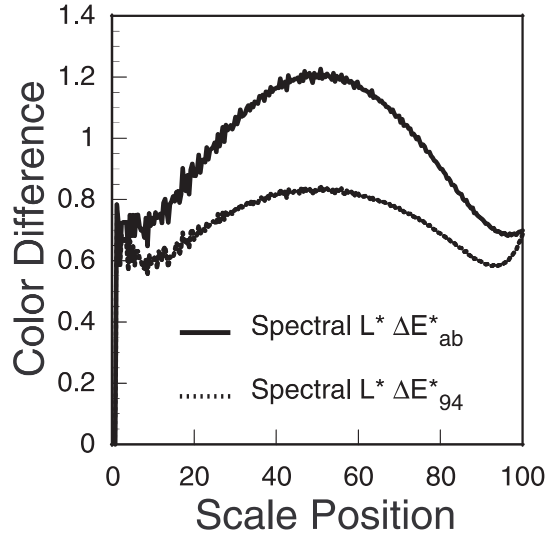

Figure 8. Color differences along Spectral L scale.*

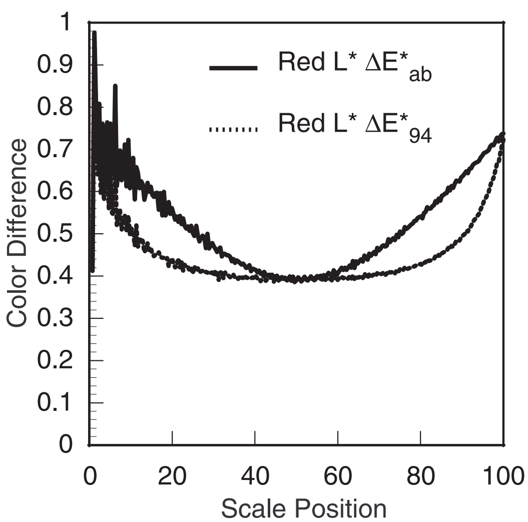

Figure 7. Color differences along Red L scale.*

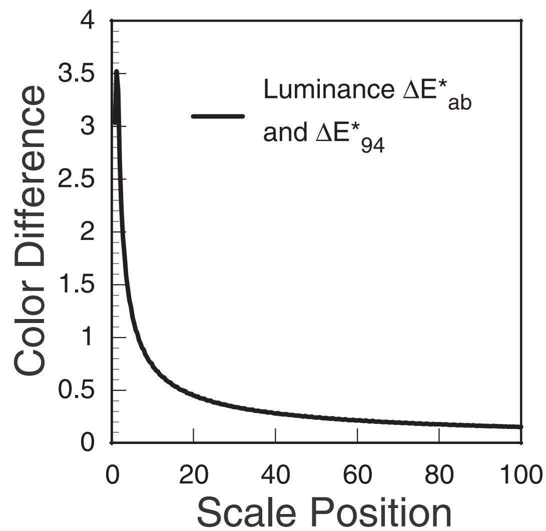

Figure 9. Color differences along Luminance scale.

By reanalyzing the data from the no-legend condition and the three different gray scales, we can find the best fitting exponential which matches the judgements of the evaluated or prompted values in the map. This assumes the subjects were not basing their judgments on their memory of any of the five encoding schemes. This analysis produced a best fitting curve with an exponent of 1.78. Further research is necessary in order to establish the shape of the optimum lightness function and how much improvement this function provides over other curves.

The addition of the hue information allows more accurate interpretation of the image values. Ware[5] found similar results. Further research is needed to determine how much additional hue information is needed to increase performance. There are many ways in which hue information could be used to supplement the gray scale encoding. This experiment only tests one possible scheme.

It is assumed that perceptual linearity in a particular color scheme will lead to better performance than color schemes based on device RGB or nonuniform color spaces. For this reason CIELAB was used in this experiment as a first approximation to a perceptually uniform space.

However, an artifact of the design of the Red L* and Spectral L* used in this study is that the step sizes are not uniform, in CIELAB units, throughout the extent of the scale. This is also true of the RGB scale and the Luminance scale. In figures 6-9, the color difference in ΔE^*_{ab} and ΔE^*_{94} are shown for the RGB, Red L*, Spectral L*, and Luminance scales. Because the contribution of ΔL* is the same in both ΔE^*_{ab} and ΔE^*_{94} the curves overlap for the RGB and Luminance schemes.

Conclusion

The current experiment demonstrates that the interpretation of multidimensional graphical information can be significantly affected by the way it is presented. The data maps used in this experiment are reasonable examples of those produced by imaging systems. With a modest amount of effort, proper presentation can improve performance by well over 30%.

This experiment is only a first step in understanding how to optimize the presentation of images. In this experiment only one additional dimension (elevation) is presented. More research is needed to discover methods to display multiple dimensions simultaneously. By applying our knowledge of visual processing to image display we can improve the ability of imaging to convey information clearly and accurately.

References

1. Philip K. Robertson and John F. O'Callaghan, The Generation of Color Sequences for Univariate and Bivariate Mapping. *IEEE Computer Graphics and Applications*, **6**, pp. 24-32, (1986).
2. Haim Levkowitz. *Color Theory and Modeling for Computer Graphics, Visualization, and Multimedia Application*. Kluwer:Boston, 1997.
3. B. Rogowitz and L. Treinish, Data Visualization: The End of the Rainbow. *IEEE Spectrum*, **35**, 12, pp. 52-59, (1998).
4. Edward R. Tufte. *The Visual Display of Quantitative Information.* Graphics Press: CT, 1983.
5. Colin Ware, Color Sequences for Univariate Maps: Theory, Experiments, and Principles. *IEEE Computer Graphics and Applications*, **8**, pp. 41-49, (1988).