Multidimensional Analysis Reveals Importance of Color for Image Quality

Ethan D. Montag* and Hirokazu Kasahara†
****Munsell Color Science Laboratory,***
Center for Imaging Science, Rochester Institute of Technology
Rochester, New York
†Epson Research & Development, Inc.,
San Jose, California

Abstract

A psychophysical experiment was performed to determine the psychological dimensions involved in judging image quality. Seven different prints for each of two images, a portrait and a landscape, were produced using a combination of 5 printers and different paper types. The experiment consisted of two parts that were run concurrently. In the first part, paired-comparison was used to evaluate image preference. In the second part, judgments of similarity and dissimilarity were made using triad presentations. The paired-comparison data were analyzed using Thurstone's Law of Comparative Judgment and Dual Scaling, a multidimensional statistical technique that reveals the independent dimensions used in categorical judgments. The judgments of similarity and dissimilarity were analyzed using nonmetric multidimensional scaling. The results indicate that the psychological stimulus space can be characterized well in two dimensions. An ideal point model can be used to identify preference in this space. Variation in subjects' preferences can be characterized predominantly in one dimension and the subjects are fairly consistent in their response along this dimension. The psychological stimulus space correlated highly with color variation in the images. We conclude that multidimensional techniques can be used to analyze image preference and find relationships between psychological and physical variables relating to image quality. Specifically, our results indicate that color is of primary importance for judging image quality in our particular situation.

Introduction

What are the factors involved in judging image quality? Traditionally to study this question, one might choose a parameter or a set of parameters based on a priori factors, vary these parameters on a set of images, and run a psychophysical experiment to evaluate the effect of these parameters on image quality by asking subjects to judge image preference. Based on these results, one could then try to optimize these parameters. Another typical approach is to use a model of human vision, perhaps incorporating the results from the type of experiment just described into the model, and use the model to predict the change in the appearance of an image from an original when certain parameters are varied.

In this experiment, we take a different approach. In this experiment we try to identify the factors involved in image quality by using multidimensional analysis of preference judgments and judgments of similarity and dissimilarity from a group of subjects who viewed prints made using printers "off-the-shelf" without any prior manipulation of the image characteristics. A simple analysis of the preference data tells us which printer does the best job, however, using multidimensional scaling and a statistical procedure called dual scaling, we attempt to identify the characteristics of the image that are involved in the subjects' decisions. These factors are identified by correlating the results of the psychophysical experiments with extensive physical measurements of the characteristics of the printers and prints used in the experiment.

Given the current state of the art, these results, in the short term, will identify those aspects of desktop printing which can stand improvement or are at least most critical in judging image quality. In the long term, we hope that this type of analysis will lead to image quality metrics, which can be used to predict and/or optimize image quality.

Experimental

Five different makes of printers, four common desktop ink-jet printers, identified as P1 through P4, and one continuous tone thermal dye diffusion printer, P5, were used in the experiment. The continuous tone printer was chosen to compare with the ink-jet printers because the quality of this printer is supposed to approach "photographic quality." These printers were used "as is" without any special calibration or characterization.

Two images were used in the analysis: a portrait (9.5 cm × 12.0 cm with a 0.4 cm white border) and a landscape image (12.5 cm × 10.0 cm with a 0.4 cm white border). These images were specified in RGB and these values were used in printing without any special color management. The images were printed on the glossy paper provided by each manufacturer for their printers. For printer P1, three different papers were used: a thick glossy paper, a matte paper, and a thin glossy paper. There were therefore a total of seven different prints for each of the two images. For the actual experiment, four copies of each print were made. During the experiment, the computer would randomly select one of the four copies for use in a trial in order to account for any possible variation in the printing process.

The images were viewed in a controlled viewing room with fluorescent D50 simulators. Twenty-one subjects participated in the experiment. The subjects were faculty, staff, students, and visitors of the Center with various experience in judging images. Many of the visitors were from Japan so that a comparison of the results from different groups of subjects could be made.

The experiment consisted of two tasks, paired-comparison and judgments of similarity and dissimilarity. For the paired-comparison, every pair of images was presented to the subject in a unique random order. The subject's task was to indicate which of the two images he preferred based on overall image quality. There were 21 pairs per image for a total of 42 trials per subject. For the similarity/dissimilarity judgments, every possible combination of three images (triads) was presented to the subject in a unique random order. For each triad, the subject's task was to indicate which two images were most similar and which two images were most dissimilar based on the overall quality of the images. There were 35 triads per image for a total of 70 triad trials (with 140 judgments of similarity and dissimilarity). The trials for the two tasks were randomly intermixed during a single session that lasted approximately one hour. A computer was used to indicate to the experimenter which prints (designated by a code) to use in each trial. The experimenter recorded the responses in the computer.

In addition to the psychophysics, physical measurements of many of the color and spatial aspects of the images were made. These measurements included CIELAB values and distributions of these values for various features of the images, the paper and primary colors. Spatial measurements included the tone reproduction curves of the prints, gloss, graininess, and values of the MTF at different points weighted by the transfer function of the visual system. These measurements were used in the analysis to correlate the psychological dimensions with physical parameters of the prints.

Results and Discussion

The data from the paired-comparison trials were analyzed using Thurstone's Law of Comparitive Judgment (Thurstone, 1927) to create an interval scale of image preference. These same data were analyzed using dual scaling (Nishisato, 1994), which is described in more detail below. The data from the triads were analyzed using Kruskal Nonmetric Mutidimensional Scaling (MDS) (Schiffman, 1981).

The results from the Thurstonian analysis of the paired-comparison are shown in Figure 1 for the portrait image. The rank order of the printers was different for the two images indicating image dependence. In addition, the error-bars indicate that there is a large overlap in performance among the printers. This may be due to lack of power (more subjects needed), a true parity among certain printers, or different patterns of response among subjects that average out in the results.

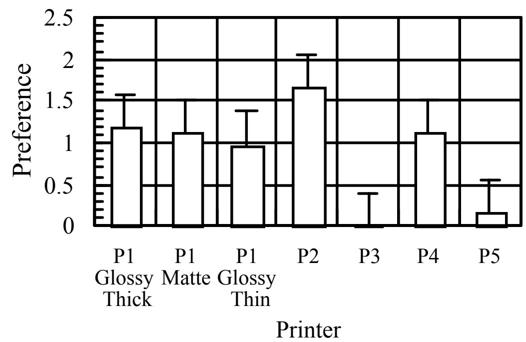

Figure 1.Thurstone Analysis: Portrait Image

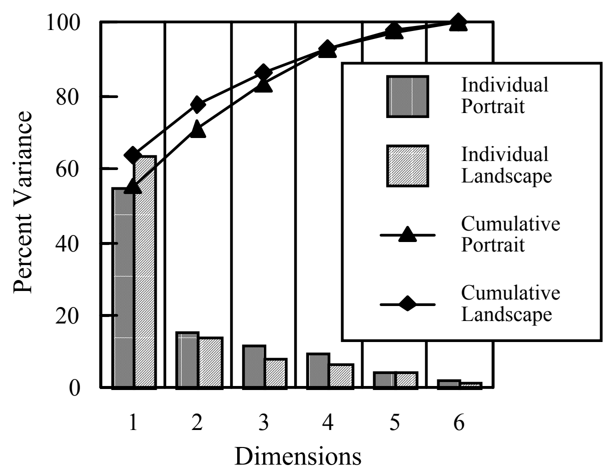

Figure 2.Percent and cumulative percent variance from Dual Scaling.

Dual scaling is a multidimensional technique performed on categorical data that reveals the independent dimensions used in the preference judgments scaled for both stimuli and individual subjects. It can be thought of as nonlinear principle component analysis for use with categorical data. Fig. 2 shows the percent variance and the cumulative percent variance explained by the dimensions for the dual scaling analysis. The first dimension accounts for the majority of the variation while the remaining dimensions are of less importance. This result is of importance because it supports the underlying theoretical assumptions used in constructing a one-dimensional interval scale using Thurstone's Law. It also demonstrates that there is a single primary dimension involved in subjects' judgments of preference.

Figure 3 shows the sample image configuration and the subject configuration for the portrait image for the first two dimensions from the dual scaling analysis. The line connecting the stimuli (filled diamonds) shows the rank order of the preference determined by Thurstonian scaling. There is good agreement in the first (x) dimension with the Thurstonian scaling. The configuration of the subjects (open circles) in these two dimensions indicates good consistency in the way the subjects judged preference: the points fall in a narrow band along the first dimension.

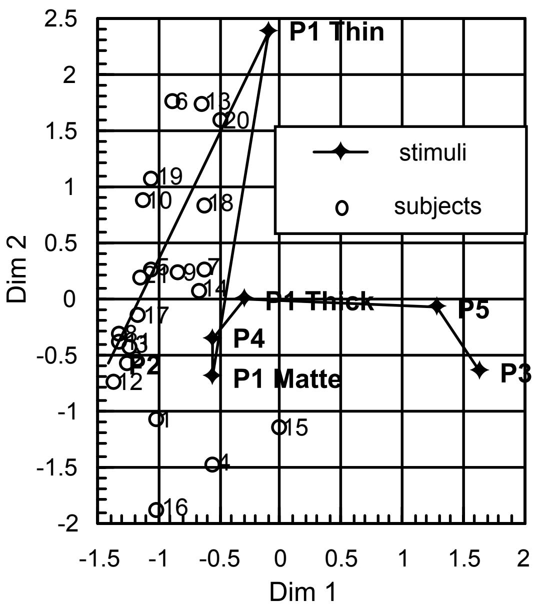

Figure 3. Dual Scaling configurations for Portrait image.

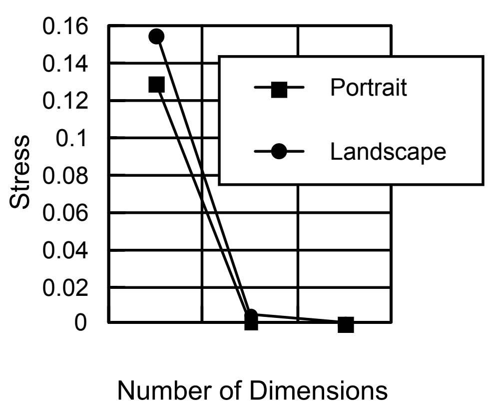

Figure 4. Stress of MDS configurations.

Approximately 1/3 of the subjects were Japanese. The pattern of responses of these subjects did not differ from the rest of the subjects. This is relevant in the later analysis that shows that the skin color in the portrait image was an important factor in judging image quality. Similar conclusions were drawn from the analysis of the landscape image.

The MDS analysis of the judgments of similarity and dissimilarity were analyzed in one, two and three dimensions. The final stress, stress being an index of goodness-of-fit that is minimized in the MDS algorithm to find the configurations, is plotted in Figure 4.

Based on the stress and the configurations themselves, further analysis was done using the 2D configurations. The 2D configurations for the portrait image can be seen plotted by the circles in Figure 5. The straight lines connect the stimuli in rank order according to the Thurstonian scaling.

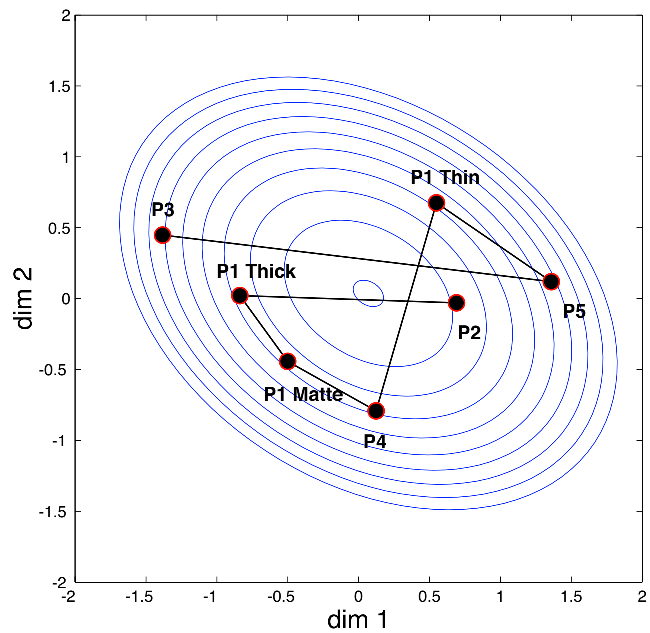

Figure 5. Ideal point model for the 2D MDS configuration of Portrait image.

The MDS configurations represent the locations of the stimuli in psychological space. In order to interpret this space, there are a number of different models that can be employed including vector, ideal-point and neighborhood models. We found that we could relate the preference data to the MDS configuration utilizing an ideal-point model. The preferences from the Thurstonian analysis correspond with the distances from an ideal-point in the 2D MDS configuration. This model is a non-parallel elliptical model for the portrait image and the preferences from the Thurstonian scaling correlates with the distances from the ideal point with an $R^2 = 1.0$. Ellipses drawn based on the model equation are shown in Figure 5. A circular model fit the data with an $R^2 = 0.89$ (not shown) for the landscape image.

The MDS dimensions have no intrinsic meaning and the orientation of the configuration is arbitrary. In order to interpret the configuration, each of the 237 physical measurements were linearly regressed with the 2D configurations to see if any of these parameters were related to the psychological dimensions. Because of the large number of parameters, we chose a strict criterion for choosing possible candidates. Only those parameters with R^2 values greater than 0.90 were considered. For the portrait image, 7 of the 237 parameters met this criterion. The R^2 values, the slope and orientation of these vectors are shown in Table I and the orientations are plotted in Figure 6.

Table I. Physical measurements highly correlated with 2D MDS configuration for the Portrait image.

Item Name	R^2	Orientation
C^*_{ab} background	0.97	39°
b* background	0.97	141°
L* red primary	0.95	129°
C^*_{ab} skin	0.94	64°
a* background	0.93	54°
a* skin	0.92	42°
L* sweater 5 (red patch)	0.91	151°

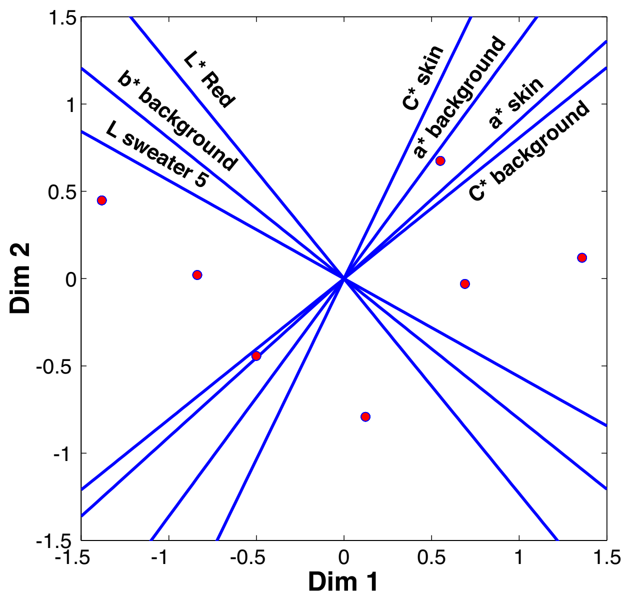

Figure 6. Vector directions of the 7 highly correlated physical parameters superimposed on the MDS configuration from Fig. 5.

Table II shows the 3 physical parameters that had R^2 values greater than 0.9 when regressed against the 2D MDS configuration for the portrait image.

Table II. Physical measurements highly correlated with 2D MDS configuration for the Landscape image.

Item Name	R^2	Orientation
a* cyan primary	0.97	3°
a* water	0.97	63°
a* mountain	0.94	68°

For the portrait image, as seen in Figure 6, the 7 vectors fall into two clusters that are at approximately right angles to each other. These groups also align well with the major and minor axes of the preference ellipses. Similarly, the 3 correlated vectors for the landscape image fall into two groups.

These parameters are not necessarily independent. Variation along one parameter may correlate with variation in another. For example, skin tone in the portrait image may vary from image to image along a line in color space in which hue, chroma, and lightness covary. Therefore, the underlying psychological dimensions may correspond to factors that correlate with collections of variables.

To explore what these underlying factors are, we computed the correlations between all pairs of these parameters. These results are shown in Tables III and IV.

Table III. R^2 values of pair-wise correlations between parameters from the Portrait image.

	C^*_{ab} bckgnd	b* bckgnd	a* bckgnd	L* red	C^*_{ab} skin	a* skin
b* backgrnd	-1.0					
a* backgrnd	0.95	-0.95				
L* red	-0.53	0.53	-0.29			
C^*_{ab} skin	0.31	-0.31	0.12	-0.91		
a* skin	0.64	-0.64	0.64	-0.92	0.90	
L* sweater 5	-0.73	0.73	-0.55	0.91	-0.83	-0.91

Table IV. R^2 values of pair-wise correlations between parameters from the Landscape image.

	a* water	a* mountain
a* mountain	0.96	
a* cyan primary	0.70	0.61

Looking at the correlations between the parameters that correspond with the MDS configuration for the portrait image, we find that these parameters fall into two groups. The C* background, b* background and a* background parameters are all highly inter-correlated forming the first group. The same is true for the L* red, C* skin, a* skin and L* sweater 5 (the red patch in the sweater). From this we can conclude that the dimensions in the psychological space correspond to one dimension that is concerned with the apparent color of the background in the image and another dimension that is concerned with the color of the portrait subject's skin tone. (The color of skin is likely heavily influenced by the contribution of the red primary and the chroma of the skin.) These results indicate that the colors of these two important features were the primary factors used in judging the relative quality of these images.

The primary influence in judging the relative quality of the landscape image was also color. The image had a blue cast and the red/green balance in the image would affect the appearance of the image. Therefore the cyan primary and the color of the mountains and water were the physical parameters that corresponded with the color balance in the image.

Conclusions

From the Thurstonian analysis of the paired-comparison data we found that there was image dependence in the judgment of image preference based on overall image quality. We could not conclude that any one printer was better than another. With dual scaling of the same data, we found that the subjects' decisions of preference varied mainly in one dimension and that there was a good deal of consistency among the subjects. Based on this idea of a predominant "image quality" dimension we could attempt to identify the factors that contributed to this dimension.

We found that the psychological space associated with each set of images could be characterized in two dimensions. Image preference could be mapped onto this space using an ideal point model. Further, by correlating the measured physical parameters with this psychological space, we found that color was the dominant determinant of image preference.

Given the variation in printers and media available at this point in time, we find that color is still the principal factor for judging print quality. This is not to say that other factors such as glossiness, graininess, sharpness and all the other possible "-nesses" (see Engledrum, 1995) are unimportant. Rather, in the case of the printers used here "off the shelf" there are at least 3 alternative explanations: either the variation in color overwhelms the variation in the spatial attributes of the images used to determine image quality, the spatial properties have approached optimal levels, or there is a certain level of parity in the spatial properties of the images produced by the different printers.

In future experiments we plan to utilize these multidimensional techniques to explore the comparative contribution of color and spatial aspects of hardcopy by manipulating these factors at different levels. In this way we hope to develop metrics that can be used to measure and improve image quality.

Acknowledgements

The authors thank the Seiko Epson Corporation, Japan, for the financial support of this research.

References

1. Engeldrum, P. (1995). A framework for image quality models. *J. Imag. Sci. Tech*. **39**, 312.
2. Nishisato, S. (1994). *Elements of Dual Scaling*, Hillsdale, NJ: Lawrence Erlbaum Assoc., Pub.
3. Shiffman, S. (1981). *Introduction to Multidimensional Scaling*. NY: Academic Press.
4. Thurstone, L. L. (1927). A law of comparative judgement. *Psychological Review*, **34**, 273-287.

Biography

Ethan D. Montag received his Ph.D. in Experimental Psychology in 1991 from the University of California, San Diego working in color vision. In 2000, he was appointed Assistant Professor at the Center for Imaging Science at RIT where he pursues work in color science in the Munsell Color Science Laboratory. His current interests include image quality, color gamut mapping, color vision, color tolerance measurement and visualization. He is a member of ARVO, OSA, ICVS, ISCC, and IS&T.

Hirokazu Kasahara received his M.E. in communication engineering from Osaka University, Japan in 1995. Since 1995 he has worked at SEIKO EPSON Corp. Currently he is a visiting scientist at the Munsell Color Science Laboratory.