Statistical Summaries of iCAM Image-Difference Maps

Scot R. Fernandez, Garrett M. Johnson, and Mark D. Fairchild
Munsell Color Science Laboratory, RIT
Rochester, New York

Abstract

Traditional color appearance models are capable of predicting the appearance of spatially simple color stimuli under a wide variety of viewing conditions and have typically been applied to images by treating each pixel as an independent stimulus. Current research has led to the development of a new image color appearance model and a modular framework for image difference metrics, known as iCAM, that combines attributes of traditional color appearance models with attributes of spatial vision models to predict the appearance of spatially complex stimuli within complex viewing environments. The model output is a spatial map of various appearance attributes such as lightness, brightness, colorfulness, chroma, saturation, and hue. With such correlates for two images it is possible to construct image-difference maps along each of these perceptual dimensions. From the image-difference maps, summary image-difference statistics can be computed (analogous to an average CIELAB ΔE_{ab}* across the pixels in an image). Such summary statistics are candidates for various types of image quality metrics. The advantage of computing differences on the output maps of an image appearance model is that the spatial filtering properties of the human visual system have been accounted for and any imperceptible differences are not represented in the image appearance maps. This paper presents an analysis of various summary statistics applied to iCAM image-difference maps for the interpretation of visual scaling results from a psychophysical data set. The data set includes analyses of complex imagery for color-preference reproduction. The results of the analysis conclude that the modular framework of iCAM, scales evenly across the separate image-difference maps, and behaves, as one would expect for the adjustment dimension manipulated in the psychophysical data set tested.

Introduction

Image quality models can be separated into two groups: Device-dependent and Device-independent. Device Dependent modeling relies on knowledge of imaging system parameters and correlates them with human perception generally through the use of psychophysical experiments to determine image quality. Unfortunately, the relationships developed are only good for the specific system tested and if the system is modified then the relationships need to be recalculated.[1]

Device-independent image quality models utilize information within the images themselves to understand image quality and perceivable differences. This paper adopts a modular framework of iCAM to better understand image appearance modeling. Figure 1 presents this framework, which has been described by Fairchild and Johnson.[1-3] Furthermore a description of the model along with example images and source code can be found at www.cis.rit.edu/mcsl/iCAM.

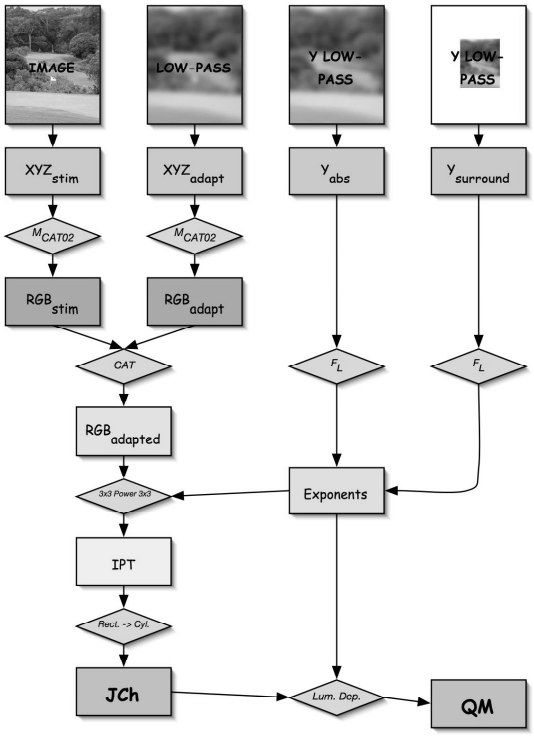

Figure 1. Flow chart of the iCAM image appearance model.[3]

This framework has many applications including image rendering and digital video rendering, but for the purpose of this paper we will focus on the iCAM's difference perceptibility capabilities to understand image quality. The difference perceptibility workflow we are utilizing in this analysis is presented in Figure 2.

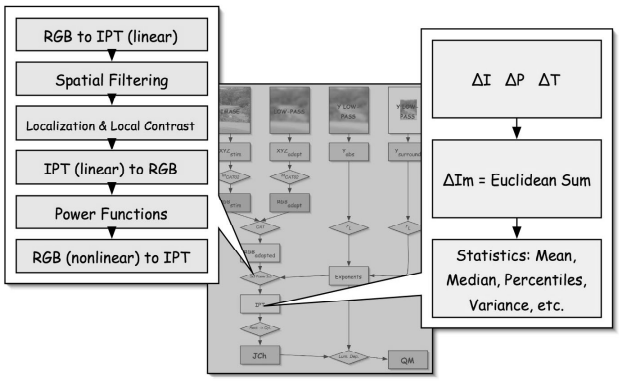

Figure 2. Implementation of iCAM for image difference and image quality metrics.[3]

To summarize Figures 1 and 2, the model requires colorimetrically characterized image data from a reference image and the reproduction. Both the images are processed to account for chromatic adaptation, and then converted to RGB signals (cone signals), which are then transformed into opponent signals, where spatial filtering is applied. Then a nonlinear compressive function is applied along with a transform into IPT space where we can perform statistics to understand image perceptibility describing the image differences.[3,4]

Psychophysical Database

The goal of this research is to better understand the capabilities and considerations needed to utilize iCAM to evaluate difference in images manipulated along colorimetric dimensions. The psychophysical data set utilized in this analysis, described by Fernandez and Fairchild, evaluates color-preference reproduction along colorimetric dimensions.[5]

Image Characteristic Ranked Order Experiment

This psychophysical experiment was a rank order design, in which observers were presented sets of manipulated images with the task to order the images from most preferred to least preferred. Each set of images represented a ramp of a single global colorimetric manipulation applied to all the images in Figure 3.

To create the sets of manipulated images, the images were adjusted along eight different CIELAB dimensions. The colorimetric dimensions chosen were a logical extension of experience from adjusting manipulating images, and later correlated to the analysis of previous research. Four of the dimensions affected color balance (additive shifts of a* and b*); the other four manipulations were lightness (a gamma adjustment of L*), contrast (a sigmoid adjustment to L*, with an threshold at 50.0 L*), Chroma (multiplicative adjustment to C_{ab}* at a constant h_{ab}), and Hue rotation (h_{ab} rotation at a constant C_{ab}*). The direct and indirect dimensions of adjustment are two of the color balance dimensions that manipulated the image along the 45° axes of the a* and b* coordinate system.

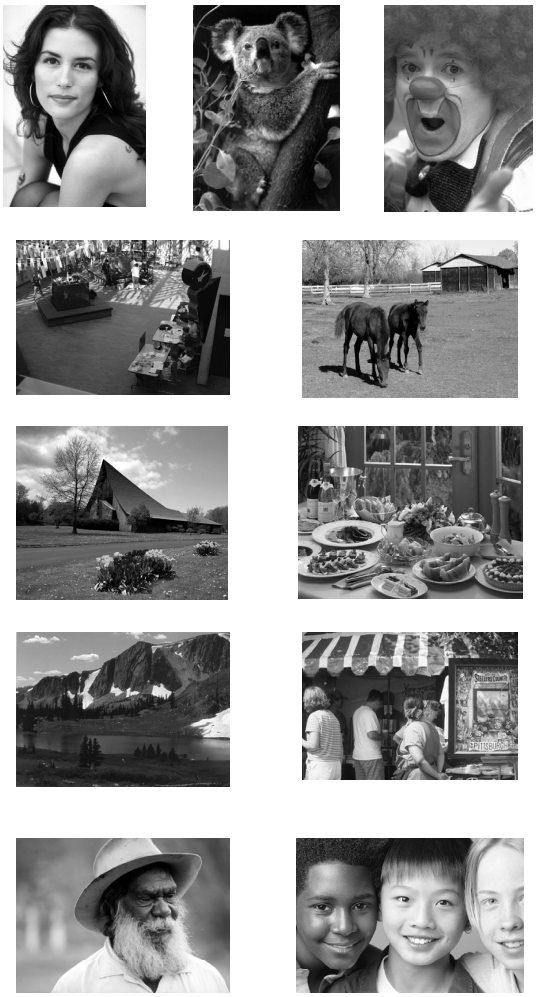

Figure 3. Image set for Experiment– (From left to right, top to bottom) 1. Model, 2. Koala, 3. Clown, 4. Indoor Scene, 5. Horses, 6. Church, 7. Dinner, 8. Mountains, 9. Art-fair, 10. Bearded Man 11. Harmony

The eight manipulations were applied to the eleven images to generate eighty-eight sheets of randomly ordered six-image sheets that varied around the nominal image. Each sheet demonstrated the effect of a single adjustment applied globally, and consisted of three steps above and below the original image. The increments were clearly

perceivable, but not objectionably large as to describe the variability in preference of good images. The increments used to generate the image sets are in Table 1.

Table 1. Adjustment Ranges and Increment Values for Experiment

	Starting Value	Ending Value	Increment
Gamma adjustment	0.55	1.30	0.15
Sigmoidal adjustment	0.55	1.55	0.20
Chroma adjustment	0.75	1.30	0.11
Hue Angle adjustment	-0.07	0.11	0.035
a* adjustment	-7.50	7.50	3.00
b* adjustment	-7.50	7.50	3.00
Direct adjustment	-7.50	7.50	3.00
	-7.50	7.50	3.00
Indirect adjustment	-7.50	7.50	3.00
	7.50	-7.50	-3.00

The sheets were printed on a Fujix Pictrography 3000, at a resolution of 300 dots per inch. The printing system was characterized using a 10 x 10 x 10 LUT, and a tetrahedral interpolation technique. The printer's forward characterization was utilized to convert the RGB images into CIELAB space, were all manipulations where applied and then the inverse characterization was utilized to convert the CIELAB images back to RGB. This workflow minimized gamut issues. A pictorial representation of a print sheet from the experiment is presented in Figure 4. This sheet represents an example of an adjustment of lightness. In addition to the placement of the manipulated image sets being randomized within each sheet, the order of image and applied manipulations were randomized throughout the entire book of image sets.

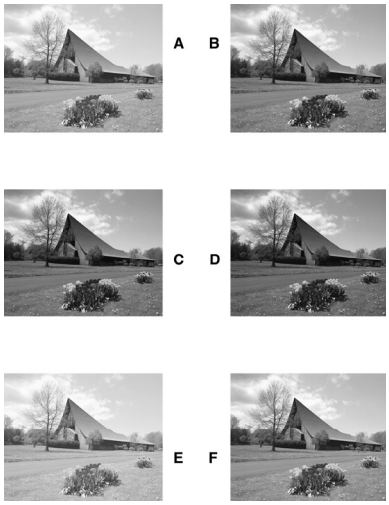

Figure 4. Sample sheet of image set from the psychophysical experiment

The experimental analysis, performed using Thurstone's Law of Comparative Judgment, resulted in magnitude scales along each colorimetric dimension. The results of the experiment are based on 77 observers each making 88 observations. The observer population consisted mostly of student, staff, and faculty of Imaging Laboratories. The breakdown of the population is 17 women, 60 men from the age range of 17 – 44.

Analysis and Discussion

Analysis I – Understanding the Image-map Metrics

The first exercise in this evaluation is one of an academic nature; the goal was to get acquainted with using the iCAM framework and metrics used to calculate image differences. The results of this analysis are presented in Tables 2-5. This evaluation calculates, for each image in Figure 3, the difference from image to black, image to white, and image to inverse image. The black image is simply the tristimulus value 0 for all pixels; the white image equated to the white point of D50 for all pixels, and the inverse image was the image minus the white point of D50 multiplied by -1.0. These calculations provided insight into the range of values one should expect resulting from the model.

Table 2. Analysis I - ΔIPT

	Mean ΔIPT	StDev ΔIPT	Max ΔIPT	P50 ΔIPT	P95 ΔIPT	P99 ΔIPT
Img vs. Black	1.47	0.19	1.99	1.46	1.79	1.87
Img vs. White	0.60	0.18	1.36	0.59	0.89	0.92
Img vs. Inv.	0.57	0.20	1.41	0.57	0.89	0.93

Table 3. Analysis I - ΔI

	Mean ΔI	StDev ΔI	Max ΔI	P50 ΔI	P95 ΔI	P99 ΔI
Img vs. Black	1.45	0.19	1.93	1.45	1.78	1.86
Img vs. White	0.52	0.19	1.10	0.52	0.81	0.84
Img vs. Inv.	0.45	0.23	1.12	0.51	0.81	0.84

Table 4. Analysis I - ΔP

	Mean ΔP	StDev ΔP	Max ΔP	P50 ΔP	P95 ΔP	P99 ΔP
Img vs. Black	0.09	0.08	0.87	0.06	0.21	0.28
Img vs. White	0.12	0.06	0.78	0.09	0.22	0.26
Img vs. Inv.	0.15	0.10	0.85	0.10	0.33	0.44

Table 5. Analysis I - ΔT

	Mean ΔT	StDev ΔT	Max ΔT	P50 ΔT	P95 ΔT	P99 ΔT
Img vs. Black	0.12	0.09	0.95	0.09	0.26	0.36
Img vs. White	0.24	0.10	0.75	0.24	0.39	0.43
Img vs. Inv.	0.24	0.10	0.75	0.24	0.39	0.43

The results in the tables above are averages across the entire set of images. Furthermore the image statistics are based on the absolute value of each image difference map, this step hides the direction of the difference but provides an accurate representation of the magnitude of the differences, which is important particularly for the percentile evaluations.

The selection of the statistics utilized here was to demonstrate how they behave and provide examples of when they might be used. The P50, P95, and P99 percentile evaluations returns the value at which that corresponding percentage of differences would fall below that value, meaning that for P50, 50% of the values are below the value indicated. P50 is most easily interpreted as the median value of a data set. We included P95, P99, and Max as a demonstration that this model could be used to determine tolerances verses the analysis of mean, which yields a magnitude scalar value which would not work as well if you were trying to optimize a system to minimize variability. The reason the analysis was repeated on the individuals' image difference maps was to understand how the components behave in relation to the overall differences computed.

To understand the tables above the ΔIPT analysis is simply a Euclidean combination of the ΔI, ΔP, and ΔT image difference maps produced by two images in the IPT opponent space, as seen in Equatation 1. In IPT opponent space I represents light-dark opponents, P represents red-green opponents, and T represents yellow-blue opponents.

$$\Delta IPT = \sqrt{\Delta I^2 + \Delta P^2 + \Delta T^2} \qquad (1)$$

The interesting note about the results of this exercise is that the layers behave how one would expect. The differences between these images are obviously lightness differences, and the calculated difference was primarily accounted for by the ΔI image difference map. Furthermore, all of the layers appear to behave on similar scales, with similar magnitudes for mean and variability. This is because of a weighting factor described by Ebner and Fairchild.[4] The P and T layers are multiplied by a factor of 1.5 to achieve similar values as the ΔI image difference map. Typically, I values range from 0 to 1 and the P and T layers have values that range from –1.5 to 1.5.

Analysis II – iCAM Image Metrics Applied to Data Set

This exercise repeated the above analysis on the psychophysical data outlined in the Psychophysical Database section. This analysis was completed without segmenting the data by image content or by adjustment dimension. The results are presented in Tables 6-9.

Each table represents the results of the specified statistical summary for 528 image pairs. So for each group of six images set, for example see Figure 4, the reference image was selected, which for this analysis the most preferred image of the set, and then the difference for the reference image and each image in the entire set were calculated. From the differences return from iCAM the summary statistics were then applied.

The analysis demonstrates that the P and T layers are more variable. The probable reason for the increase in variability is that the data set includes six colorimetric manipulation dimensions that directly affect the P and T layers, and only two of the adjustments dimension that have localized affects on the I layer of the IPT color space. Furthermore, this demonstrates that the maximum Max value for the ΔI and ΔT images difference maps is very high considering the previous exercise. Fortunately, using the values from the P99 calculation yields that these spike values are most likely only one noisy pixel out of the 528 image difference calculations that this data analysis is based on. This noisy pixel could be the result of an image artifact as a result of possible clipping, or round-off error. Other than the discussed abnormal responses the results of the calculations appear to be normal and consistent with ones intuition of the model, in that the each of the summary statistics are similarly scaled and evenly contribute to the overall difference computation.

Table 6. Summary Statistics for ΔIPT Calculation

	Mean ΔIPT	StDev ΔIPT	MIN ΔIPT	P50 ΔIPT	P95 ΔIPT	P99 ΔIPT	MAX ΔIPT
Mean	0.04	0.02	0.01	0.04	0.05	0.08	0.55
StDev	0.03	0.02	0.02	0.03	0.04	0.09	0.43
MAX	0.16	0.09	0.11	0.20	0.26	0.55	2.11

Table 7. Summary Statistics for ΔI Difference Image

	Mean ΔI	StDev ΔI	P50 ΔI	P95 ΔI	P99 ΔI	MAX ΔI
Mean	0.011	0.004	0.010	0.016	0.019	0.075
StDEV	0.019	0.007	0.018	0.027	0.031	0.093
MAX	0.104	0.042	0.104	0.147	0.154	0.513

Table 8. Summary Statistics for ΔP Difference Image

	Mean ΔP	StDev ΔP	P50 ΔP	P95 ΔP	P99 ΔP	MAX ΔP
Mean	0.021	0.009	0.018	0.031	0.041	0.491
StDEV	0.025	0.010	0.024	0.033	0.041	0.428
MAX	0.123	0.087	0.127	0.231	0.238	2.093

Table 9. Summary Statistics for ΔT Difference Image

	Mean ΔT	StDev ΔT	P50 ΔT	P95 ΔT	P99 ΔT	MAX ΔT
Mean	0.021	0.013	0.021	0.031	0.060	0.321
StDEV	0.023	0.013	0.027	0.031	0.083	0.237
MAX	0.106	0.071	0.181	0.156	0.491	1.023

Analysis III – iCAM Image Metric Applied to the Segmented Data Set and Correlated with Preference Analysis

To introduce this next analysis, Figures 5, 6, and Table 10 have been generated to visually represent what is being discussed in Figures 7 and 8, and Tables 11 – 14. So for each ramp of a manipulation dimension/image combination the most preferred image was chosen as the reference image and the difference between the reference and each image in the ramp was calculated. In Figure 5, Image D was the most preferred image. Table 10 presents the actual results of the analysis, and Figure 6 is a plot of the data in Table 10.

Figure 5. Sample Ramp- Gamma Adjustment

Table 10. Corresponding Mean ΔIPT and Preference Data for Figure 5

Step	A	B	C	D	E	F
MEAN ΔIPT	0.11	0.06	0.03	0.00	0.02	0.04
Preference	1.67	0.70	0.06	0.00	0.53	0.98

The final analysis segments the second analysis into the component adjustment dimensions tested in the psychophysical experiment, and then correlates the ΔIPT computations to the preference results of the experiment. The results are presented in Figures 7 and 8.

The graphs in Figures 7 and 8 represent average mean ΔIPT verse mean preference for the individual adjustment dimensions. Each point on the plot represents the center of mass of eleven mean ΔIPT and correlated preference results, this would equate to averaging eleven of Figure 5. Since Figures 7 & 8 are averaged across an image set, one would expect to get a linear relationship for each adjustment dimension, meaning a relationship independent of image content or system. Theoretically, we could now use image difference and generate the preference difference without doing more psychophysics, thus being a device-independent image quality scale. So the slope of the line for a given dimension is now an image quality scale, furthermore the straighter the line the more defined the relationship is between the image dimension and preferences.

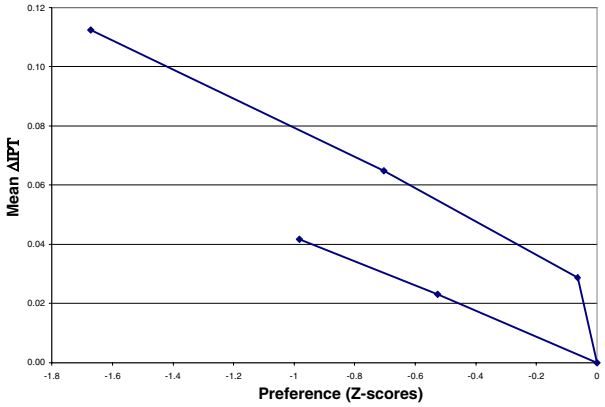

Figure 6. ΔIPT verses Preference for Figure 5.

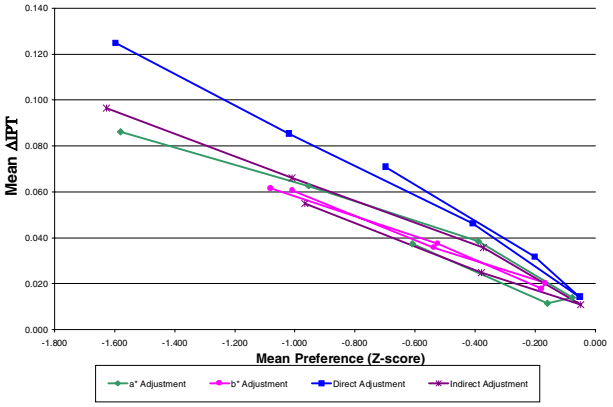

*Figure 7. ΔIPT verses Preference for a*b* Dimensions*

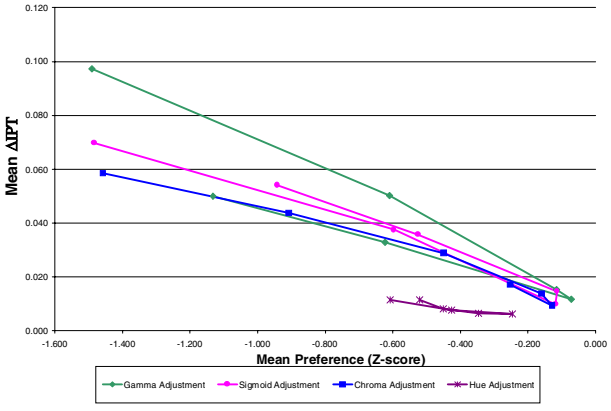

*Figure 8. ΔIPT verses Preference for non- a*b* Dimensions*

It is interesting to note that in Figure 8 Hue Angle Rotation manipulation separates it's self from the other dimensions. This is indeed representative of the experiment in that the discernability of this dimension was significantly more difficult than the other seven, and it is encouraging to see that the model also concluded that the calculated image difference is very small in comparison to the other dimensions. Furthermore, the Direct and Indirect dimension were a Euclidean distances of a*b* adjustments which were visually larger and were calculated to have the largest image differences. These plots are very representative of the order of discernability between adjustment dimensions, and for the most part support our goal in the psychophysical design phase of generating scales that were visually uniform across dimensions without being largely objectionable.

To further this analysis some of the prior mentioned summary statistics were calculated for each of the image difference maps in order to better understand how the individual layer behaved when segmented by adjustment dimension. These results are presented in Tables 11– 14.

Table 11. Mean of ΔIPT Image for Each Dimension

Adjustment Dimension	Mean ΔIPT	StDev ΔIPT	Max ΔIPT
Gamma	0.043	0.032	0.120
Sigmoid	0.037	0.024	0.079
Chroma	0.029	0.025	0.113
Hue	0.009	0.008	0.044
a*	0.042	0.030	0.109
b*	0.039	0.027	0.103
Direct	0.062	0.043	0.163
Indirect	0.048	0.033	0.127

Table 12. Mean of ΔI Image for Each Dimension

Adjustment Dimension	Mean ΔI	StDev ΔI	Max ΔI
Gamma	0.037	0.028	0.104
Sigmoid	0.034	0.022	0.078
Chroma	0.002	0.002	0.011
Hue	0.001	0.001	0.003
a*	0.001	0.001	0.005
b*	0.005	0.004	0.018
Direct	0.004	0.003	0.012
Indirect	0.005	0.004	0.020

Table 13. Mean of ΔP Image for Each Dimension

Adjustment Dimension	Mean ΔP	StDev ΔP	Max ΔP
Gamma	0.008	0.008	0.039
Sigmoid	0.005	0.004	0.015
Chroma	0.018	0.018	0.093
Hue	0.007	0.006	0.035
a*	0.041	0.029	0.109
b*	0.008	0.006	0.027
Direct	0.046	0.032	0.123
Indirect	0.032	0.022	0.085

Table 14. Mean of ΔT Image for Each Dimension

Adjustment Dimension	Mean ΔT	StDev ΔT	Max ΔT
Gamma	0.017	0.012	0.049
Sigmoid	0.009	0.006	0.026
Chroma	0.019	0.017	0.077
Hue	0.005	0.004	0.025
a*	0.003	0.002	0.008
b*	0.038	0.026	0.099
Direct	0.041	0.028	0.106
Indirect	0.034	0.024	0.092

This final analysis reveals that the iCAM framework for calculating image differences behaves, as one would expect. Each adjustment dimension yields max difference in their corresponding difference layer. For example, both Sigmoid and Gamma demonstrate a majority of their difference and variability in the ΔI image difference map, the a* manipulation corresponds well to the ΔP image difference map, and the b* manipulation corresponds well to the ΔT image difference map. The one interesting observation is that the IPT space appears to be slightly more sensitive along the a* dimension, and significantly more sensitive along the Direct dimension in comparison to the results of the Indirect dimension. These two non-linear sensitivity issue are probably a function of the color space chosen. If these discrepancies are inherent to the color-space then weighting factors could be used to correct the metric, which is analogous of what has been applied to the ΔE^*_{94} in comparison to it's predecessor ΔE^*_{ab}. [6]

Conclusion

The next generation of appearance models, iCAM, has been proposed as a modular framework to accurately represent images across many different viewing conditions. This paper has been an exercise to evaluate iCAM's capability to serve as an image perceptibility model along colorimetric manipulation dimensions. In the process several statistical calculation were applied to both the individual layer of the ΔIPT image as well as the ΔIPT components. This analysis has yielded that this image appearance model behaves, as one would expect for the adjustment dimension manipulated in this psychophysical data set. It also yields results on a reasonably uniform scale that agreed with both preference data and visual observations.

References

1. G.M. Johnson and M.D. Fairchild, Measuring images: Differences, Quality, and Appearances. SPIE/IS&T Electronic Imaging Conference, Santa Clara, in press (2003).
2. M.D. Fairchild and G.M. Johnson, Meet iCAM: A Next Generation color appearance model. IS&T/SID 10th Color Imaging Conference, Scottsdale, 33-38 (2002).
3. M.D. Fairchild and G.M. Johnson, Image Appearance Modeling. SPIE/IS&T Electronic Imaging Conference, Santa Clara, in press (2003).
4. F. Ebner and M.D. Fairchild, Development and Testing of a color Space (IPT) with improved Hue Uniformity. IS&T/SID 6th Color Imaging Conference, Scottsdale, 8-13 (1998).
5. S.R. Fernandez and MD Fairchild, Observer Preferences and Cultural Differences in Color Reproduction of Scenic Images. IS&T/SID 10th Color Imaging Conference, Scottsdale, 66-72 (2002).
6. R.S. Berns, Billmeyer and Saltzman's Principle of Color Technology. NY: John Wiley & Sons; (2000).