The Perceptual Study of the Tolerance of Spectral Images Based on Bootstrap Analysis

Qiao Chen[1, a], Lijie Wang[1], Stephen Westland[2]

[1] School of Media and Communication, Shenzhen Polytechnic, Shenzhen, China

[2] Centre for Colour Design Technology, School of Design, University of Leeds, UK

[a]chenqiao_cn@hotmail.com

Abstract. Spectral images contain a large volume of data and place considerable demands on computer hardware and software compared with standard trichromatic image storage and processing. Although the imformation of reflectance spectra may be represented efficiently using linear models, a key task is to answer how many basis functions of a linear model are necessary for a given accuracy of representation. Because the most common application of spectral images is the reproduction of color and images, it is highly important to study the tolerance of human perceptions over spectral images represented by low-dimensional linear models. In this study several data sets and psychophysical studies have been used to investigate how many basis functions are necessary to represent a spectral image to be indistinguishable from the original image. The raw objective experiment datas are analysed by a comprehensive probit analysis based on the bootstrap method, and results show that the perceptual tolerance for spectral images reconsctucted by variouse number of basis functions is highly image dependent.

Introduction

With the rapid development of computer softerware and hardware, spectral images which have been used widely in satellite remote sensing are stepping into our daily life such as printing, displaying, etc. But spectral images whose each pixel contains full information of spectral reflectance provide a huge volume of redundant data for the purposed of color reproduction, it is therefore highly desirable to find a way to represent spectral images efficiently by compressing them into a more compact form. It is known any spectral reflectance distribution $P(\lambda)$ can be approximated to a specified degree of accuracy as a weighted sum of basis functions $B(\lambda)$, thus

$$P(\lambda) = a_1 B_1(\lambda) + a_2 B_2(\lambda) + \cdots + a_n B_n(\lambda)_{.} \tag{1}$$

where $P(\lambda)$ is the reflectance spectrum, $B_i(\lambda)$ is the *ith* basis function and a_i is the coefficient or weight for the *ith* basis function [1]. Once the $B_i(\lambda)$ are known, a set of weights a_i is sufficient to specify any reflectance spectrum in the model. It has been stated that spectral reflectance of surfaces and spectral power distributions of illuminants are highly constrained [2,3,4], therefore they can be approximated by low-dimensional linear models of limited weighted sum of basis functions. A key question, however, is how many components are necessary for a given accuracy of representation. Many studies [5,6] have been done to study the number of basis functions required for low-dimensional linear models, which are all derived from the spectra data sets themselves, and few works are concerned with human perception tolerance. The present studies together with the considered each reflectance spectrum in isolation and therefore did not address the question of how many basis functions are required to accurately reproduce a spectral image where pixels are viewed in relation to each other. The question is also ill-posed, however, since it is understood that the number of basis functions required depends to a great extent on the intended application of the linear model [7]. While the most common application is the reproduction of color and images, which highly relate to human perception, it is important to understand the relationship between human perceptual tolerance of spectral images and the number of basis functions of the corresponding linear models.

This work is an extension of our preliminary psychophysical study [8], and the aim of this work was to investigate, psychophysically, the variation of the number of basis functions required to reproduce natural complex spectral images. The human perceptual threshold is then analyzed by probit analysis [9] using the bootstrap method [10].

A general psychometric function $P(x)$ which can be expressed as Equation 2 is considered as a method to relate the probability of stimulus responses p (usually correct response) and stimulus intensity x so that the cumulative normal distribution is approximated by a continuous function along the stimulus axis (Guilford, 1954).

$$P(x;\alpha,\beta,\gamma) = \gamma + (1-\gamma)F(x;\alpha,\beta), \qquad -\infty \leq x \leq \infty \tag{2}$$

where γ is the chance performance which gives the lower bound of $P(x)$, and in the case of 2AFC paradigm, γ is fixed at 0.5. F which is a two-parameter function is typically a sigmoid function, such as Weibull, logistic, cumulative Gaussian, etc. the parameter α and β of $F(x)$ will determine the displacement and slope of the psychometric function. Therefore, the threshold and the slope of a psychometric function can be derived from the inverse of F at a particular performance level (usually 75% for 2AFC) as a measure of displacement and the slope of F. Parameters are estimated by a fitting procedure (maximum-likelihood method), but it is also important to estimate the error for those parameters (variability) that is to compare response thresholds or slopes across experimental conditions, which will depend on the number of experimental trials taken and their placement along the stimulus axis. Probit analysis has been largely used to estimate variability based on asymptotic theory which requires a large number of data points (number of stimulus levels), and thus not suitable for realistic psychophysical settings. While a bootstrap method has been found to provide better and more reliable results for variability [11].

Experimental

Five observers who had normal color vision and normal or corrected-to-normal visual acuity participated in the experiment. For each spectral image, a linear model was constructed with up to 9 basis functions to reconstruct the spectra of the original image. Only 9 basis functions are used because 9 basis functions in total already give a very small reconstruction errors in the previous study [8]. The linear models with number of basis functions variant from 1 to 9 were used to reconstruct the original spectral of each pixel using least square best fit technique. Thus for each original image, 9 reproductions corresponding to 1 to 9 basis functions will be obtained. The reconstruction errors CIELAB color difference will be calculated. For the purpose of psychophysical experiment, A 17 inch CRT monitor (Sony Trinitron) was used for image rendering. Screen resolution was 1024×768, and refresh rate was 80 Hz. A GOG model had been derived using Minolta CS-1000 telespectroradiometer. The monitor was calibrated each time before experiment, and the white point was set to D65. All images were displayed in a darkened room. The performance of the GOG model has an average color difference of 0.72 ΔE^*_{ab} with a standard deviation of 0.39 ΔE^*_{ab} based on 64 test colors.

In this experiment, a set of 4 forest images and 4 coral images were prepared for the experiment and image size is 128×128 pixels. Some images are different with the images we were used in the previous preliminary study. The 4 forest images are: park2, rainfor5, cootth12 and bank, while 4 coral images are: horshe5, horshe10, hoshow12 and hoshoe30. Thus in this experiment, for each image, 9 reproduced images were generated for linear model of number of basis functions varing from 1 to 9. The reconstruction errors in CIELAB ΔE^*_{ab} can be theoretical computed from reflectance and weighting functions. However, as images are rendering on CRT, the real image color differences will be different with theoretical computations because of errors caused by quantisation and GOG model itself.

Figure 1: Image arrangements for psychophysical experiment.

In a two alternative forced choice (2AFC) paradigm observers viewed three horizontally displayed images as described in Figure 1. The centre image was always the original image; the left-hand and right-hand images were randomly selected so that one was the original image and the other was the reconstructed image (varying numbers of basis functions were used in the reconstruction). The observers were informed that the central image was the original and were forced to choose whether the left-hand or right-hand image was an identical match to the original. In each trial, the reproduced image was randomly located either on the left or right side of the original and the order of each image was in a random sequence. Each observer acted at least 15 repeated trials for each comparing process, as the number of trial is important and suppose to be large in such vision research psychophysical experiments [12]. Therefore at lest a total of 1080 trials (15 (repetitions) × 8 (number of images) × 9 (1 to 9 basis functions)) were assessed by each observer. The viewing distance was 80 cm, so that the visual angle is about 3 degree.
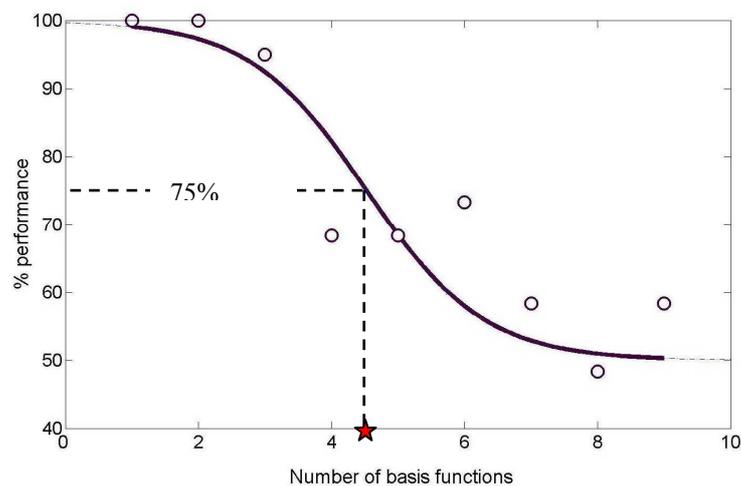

Figure 2: An example of the perceptual response of discrimination of original and linear-model-reconstructed images by one observer. Percent-correct response based on 15 trials per level is plotted (circles) as a function of the number of basis functions. The smooth curve is a logistic regression. For the criteria of 75%, threshold as numbers of basis functions is indicated by the red star.

It was assumed that when the reconstructed image was very different from the original then the observer would correctly select the original image with a high probability. Conversely, when the quality of the reconstruction was high it would be difficult for the observer to identify the original and discrimination performance would tend towards 50% correct responses. Psychophysical data were fitted by psychometric functions using *psignifit* version 2.5.6 [13], a software package which implements the maximum-likelihood method. In the experiment, logistic regression had been

applied to fulfil the fitting. For each psychometric function the criterion level of discrimination performance was set to 75% correct responses, which has often been used as JND, and this threshold was therefore determined as illustrated in Figure 2 which shows an example of the fitting process of one observer's results.

Results

Thresholds for each observer shown in Table 1 were obtained for data pooled over each image, and 95% confidence intervals were calculated by a bootstrap method based on 1999 replications for each image.

Table 1: Threshold in term of number of basis functions of each 8 images for 5 observers.

| Observer | Test images | | | | | | | |
	rainfor5	cootth12	bank	park2	5horshe	10horshe	12hoshoe	30hoshoe
VC	3.04	2.00	3.82	1.85	4.26	5.23	5.69	4.20
QC	2.63	2.14	2.78	3.02	4.16	7.71	5.78	3.57
WL	3.57	2.44	2.00	2.82	4.24	4.72	5.62	3.59
FC	4.70	2.94	2.48	3.10	3.88	3.88	6.06	1.91
AG	2.00	2.39	2.48	2.00	2.96	4.00	5.69	2.84

It can be seen the sensitivity for each observers are variant a lot, with also large variations across different individual image. For example, In Table 1, for the forest image rainfor5, the variation of observers is from 2.00 to 3.04, and for image 30hoshoe, a much larger variation of observers occurs from (1.91 to 4.2)., it can be found that variations over images are much larger than the variations over observers. This may be due to two distinction set of images. Taken over subjects, the mean thresholds for each image are shown in Figure 3.
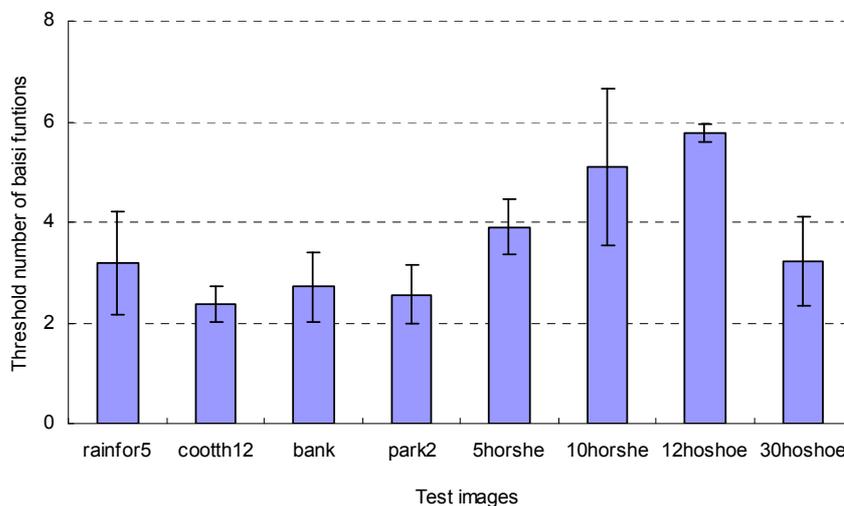

Figure 3: Average discrimination thresholds in terms of the number of basis functions of 8 spectral images over all observers. Error bars represent standard errors through observers.

It can be found the overall responses of all observers are similar to those of individual observer that all forest images (plotted on the left-side of the graph) indicates smaller threshold values than the coral images (plotted on the right-side of the graph) for observers. The average thresholds for criterion level of 75% for the two classes of images were plotted in Figure 4 by pooling each observer's results of each class of images together.
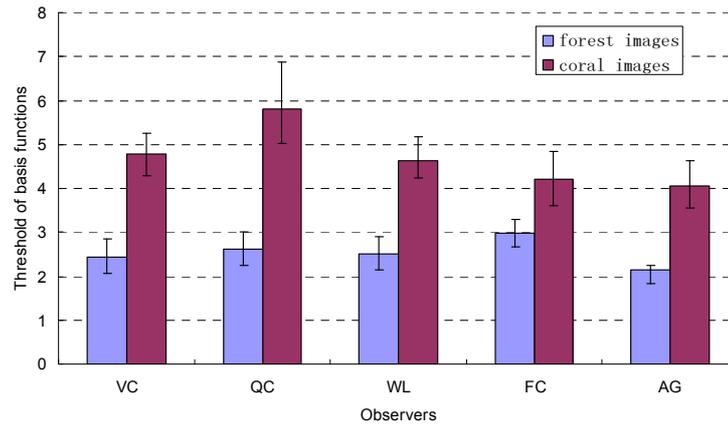
Figure 4: Discrimination thresholds in terms of the number of basis functions of coral images and forest images for 5 observers. Error bars of 95% confidence intervals were estimated from a bootstrap based on 1999 replications with resampling for each class of image.

It is evident that the average threshold for forest images is smaller than for coral images, which is consistent with the reconstruction errors in CIELAB ΔE^*_{ab} for forest images are always smaller than coral images with various number of basis functions. Table 2 shows the performance of thresholds in terms of CIELAB ΔE^*_{ab} color-difference. The results also indicate that they are all image dependent. It is noticed the visual responses for all 5 observers are generally consistent, but there are a few exceptional results. For example, in Table 2 the result of AG for the image "rainfor5" is much higher than other observers. This is also illustrated in Figure 5 that the red circle which stands for the percent-correct response of two basis functions has caused great uncertainties for the determination of threshold based on the criteria of 75%.

Table 2: Threshold in terms of CIELAB ΔE^*_{ab} of each 8 images for 5 observers

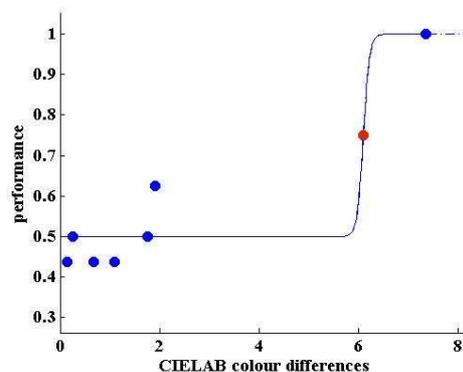| Observer	Test images							
	rainfor5	cootth12	bank	park2	5horshe	10horshe	12hoshoe	30hoshoe
VC	1.89	3.00	1.26	5.4	2.31	1.73	1.92	3.42
QC	2.44	2.78	1.68	3.1	2.88	1.33	1.89	4.02
WL	1.82	1.32	6.99	2.71	2.99	1.8	1.92	3.75
FC	1.12	2.07	1.44	2.46	2.55	2.31	1.86	4.62
AG	6.1	1.32	1.44	4.22	4.59	2.22	1.92	3.91
Total	1.84	2.84	6.40	3.67	3.37	2.10	1.91	3.95

Figure 5: An example of the perceptual response of discrimination of original and linear-model-reconstructed images by observer AG on image rainfor5. Percent-correct response based on 16 trials per level is plotted (circles) as a function of the CIELAB color difference of corresponding number of basis functions. A logistic regression is shown to fit the data.

Conclusions

A comprehensive psychophysical experiment has been carried out in this work to determine human visual perceptual threshold (based on a criterion of 75%) in terms of number of basis functions required over a set of spectral images. Though the individual observer's sensitivity varies greatly, for all observers, similar trends are presented, that the threshold of coral image 12horshe are the highest for nearly all observers except it is the second highest for observer QC, while the number of basis functions needed for forest images are relatively smaller than coral images which is also illustrated in Figure 4. Obviously, the psychophysical experiment results indicate that the discrimination threshold as the number of basis functions is highly image dependent. The results shows the number of basis functions required for coral images are relatively larger than for forest images. In general, the results of the psychophysical experiment are quite consistent with the average results of using traditional CIELAB color difference formula, but the testing results of performances of different CIE color difference formulae suggest that there are also discrepancies existing which means for complex images, besides color, spatial details also play an important role. This is possible and very often because complex images can not just be judged by using simple traditional color difference formulas which have been successfully applied to specify color accuracy for large-size-uniform surface patches. Instead rather more complex image difference metrics should be used to simulate the response of human visual system which judges the difference colorimetrically and spatially.

References

[1] Maloney, L. T. (1986). Evaluation of linear models of surface spectral reflectances with small numbers of parameters. *Journal of the Optical Society of America. A*, 3, 1673-1683.

[2] Maloney, L. T. (1999). *Physics-based approaches to modeling surface color perception*, Cambridge University Press.

[3] Romero, J., Garcia-Beltran, A. and Hermandez-Andrez, J. (1997). Linear bases for representation of natural and artificial illuminants. *Journal of the Optical Society of America. A*, **14**, 1007-1014.

[4] Owens, H. C. (2002). Color and spatiochromatic processing in the human visual system. *PHD Thesis, University of Derby*.

[5] Parkkinen, J., Hallikainen, J. and Jääskeläinen, T. (1989). Charateristic spectra of Munsell colors. *Journal of the Optical Society of America. A*, **6**, 378-322.

[6] Jaaskelainen, T., Parkkinen, J. and Toyooka, S. (1990). Vector-subspace model for color representation. *Journal of the Optical Society of America. A*, **7**(4), 725-730.

[7] Hardeberg, J. Y. (2002). On the spectral dimensionality of object colours. In *Proceedings of the First European Conference on Colour in Graphics, Imaging, and Vision*, 480-485.

[8] Q. Chen,, L. Wang, , S. Westland., A Perceptual Study of Linear Models of Spectral Reflectance.ppt. Proc. Of the 2nd International Conference on Graphic Communications,IEEE Computer Society, 2008, 309-312.

[9] Finney, D. J. (1971). *Probit analysis, 3rd edition*. Cambridge, Cambridge University Press.

[10] Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, **7**, 1-26.

[11] Foster, D. H. and Bischof, W. F. (1991). Thresholds from psychometric functions: Superiority of bootstrap to incremental and probit variance estimators. *Psychological Bulletin*, **109**, 152-159.

[12] Leek, M. R. (2001). Adaptive procedures in psychophysical research. *Perception & Psychophysics*, **63**(8), 1279-1292.

[13] Wichmann, F. A. and Hill, N. J. (2001). The psychometric function: I. Fitting, sampling and goodness-of-fit. *Perception & Psychophysics*, **63**(8), 1293-1313