

Visual sensitivity to color errors in images of natural scenes

MIKEL A. ALDABA,¹ JOÃO M.M. LINHARES,¹ PAULO D. PINTO,¹ SÉRGIO M.C. NASCIMENTO,¹
KINJIRO AMANO,² AND DAVID H. FOSTER²

¹Department of Physics, Minho University, Campus de Gualtar, Braga, Portugal

²Sensing, Imaging, and Signal Processing Group, School of Electrical and Electronic Engineering, University of Manchester, United Kingdom

(RECEIVED March 8, 2006; ACCEPTED March 9, 2006)

Abstract

Simple color-difference formulae and pictorial images have traditionally been used to estimate the visual impact of color errors introduced by image-reproduction processes. But the limited gamut of RGB cameras constrains such analyses, particularly of natural scenes. The purpose of this work was to estimate visual sensitivity to color errors introduced deliberately into pictures synthesized from hyperspectral images of natural scenes without gamut constraints and to compare discrimination thresholds expressed in CIELAB and S-CIELAB color spaces. From each original image, a set of approximate images with variable color errors were generated and displayed on a calibrated RGB color monitor. The threshold for perceptibility of the errors was determined in a paired-comparison experiment. In agreement with previous studies, it was found that discrimination between original and approximate images needed on average a CIELAB color difference ΔE_{ab}^* of about 2.2. Although a large variation of performance across the nine images tested was found when errors were expressed in CIELAB units, little variation was obtained when they were expressed in S-CIELAB units.

Keywords: Color science, Color differences, Color reproduction, Natural scenes, Complex images

Introduction

Color errors occur in all image-reproduction processes and their visual significance is an important factor influencing perceived image quality. In traditional colorimetry, the perceived color differences between uniform stimuli observed on uniform gray backgrounds may be quantified with reference to the color spaces CIELAB, CIELUV, CIE94, and other approximately uniform spaces (Fairchild, 2005). With complex images, the application of color-difference formulae expressed as Euclidian distances in these spaces is not straightforward because the effects of spatial structure need to be taken into account (Zhang & Wandell, 1996; Fairchild & Johnson, 2004). The extent to which relatively simple formulae may predict perceptibility and acceptability of images is important in many areas, such as the assessment of the quality of color reproduction (Berns, 2001; Imai et al., 2003) and the quantification of spectral errors in recovering reflectance data of natural objects (Cheung et al., 2005).

Some psychophysical estimates of the visual significance of color errors in complex images have been obtained with pictorial images displayed on color monitors (Stokes et al., 1992; Song & Luo, 2000). These experiments provided useful data because they simulated realistic color distortions and quantified both notions of

perceptibility and acceptability. The test images were, however, obtained with RGB cameras, which have a constrained gamut and limited chromatic fidelity (Wu et al., 2000; Morovič & Morovič, 2003). Natural scenes have color gamuts that may extend beyond those of RGB cameras (Webster & Mollon, 1997; Párraga et al., 1998; Nascimento et al., 2002), possibly influencing the way color errors are quantified in some applications. In addition, only color-difference formulae strictly applicable to uniform stimuli have been used to quantify color errors.

The purpose of this work was to assess visual sensitivity to color errors introduced deliberately into pictures synthesized from hyperspectral images of natural scenes. The advantage of using hyperspectral images in this context is that the chromaticities of the pictures derived from them and the color errors can be rendered faithfully without being constrained by the color gamut available to an RGB image-acquisition device; in particular, there is no limit on luminance, hue or chroma. Fidelity is then limited solely by the gamut of the display device, which in the present work affected only a very small proportion of the images. Visual sensitivity to these errors was expressed in terms of distances in CIELAB color space and in its extension to spatially complex images, S-CIELAB space (Zhang & Wandell, 1996). In S-CIELAB space, the image is first transformed into an opponent-colors form, and each color dimension is convolved with a kernel the shape of which is determined by the spatial sensitivity to that dimension; the filtered representation is then transformed to CIELAB (Zhang & Wandell, 1996). Although more complex color spaces may better represent

Address correspondence and reprint requests to: Sérgio M.C. Nascimento, Department of Physics, Minho University, Campus de Gualtar, 4710-057 Braga, Portugal, E-mail: smcn@fisica.uminho.pt

color differences, these two spaces are still used in many practical applications, and their use here allows comparison with previous studies.

Psychophysical estimates of the threshold for perceptible errors were obtained from paired comparisons of the original and chromatically manipulated images displayed on a calibrated color monitor. It was found that observers' ability to discriminate between images needed on average a CIELAB color difference ΔE_{ab}^* of about 2.2. Although there was a large variation about this mean across the nine images tested in CIELAB space, this variation was much reduced when expressed as an S-CIELAB color difference ΔE_s^* .

Materials and methods

Hyperspectral images

Images of rural and urban environments were obtained by a hyperspectral imaging system (Foster et al., 2004) with a low-noise Peltier-cooled camera with a spatial resolution of 1344×1024 pixels and 12-bit intensity resolution (Hamamatsu, model C4742-95-12ER, Hamamatsu Photonics K.K., Japan). The focal length of the camera was typically 75 mm, producing an angular resolution of about 1 arc min per pixel. A fast tunable liquid-crystal filter (VariSpec, model VS-VIS2-10HC-35-SQ, Cambridge Research & Instrumentation, Inc., MA) was mounted in front of the lens, with a wavelength transmission range between 400 and 720 nm, a resolution of 1 nm and a FWHM transmission of 10 nm at 500 nm.

For each scene, thirty-three monochromatic images were acquired in 10-nm steps over 400–720 nm. A gray reference surface was introduced into each scene and the spectral-power distribution of the diffusely reflected light was measured with a telespectroradiometer (Spectra Colorimeter, PR-650, Photo Research Inc, Chatsworth, CA) just after the spectral scan. The spectral radiance data obtained from the gray reference were used to calibrate the hyperspectral images and derive the spectral radiance at each pixel. Images were corrected for dark noise, spatial nonuniformities and stray light. For further details, see Foster et al. (2004).

Nine scenes were selected from the database of images acquired in the Minho region of Portugal. The scenes represented rural and urban environments imaged at several distances, from near to far. Fig. 1 shows color pictures of the scenes tested, with those in the middle row classified as urban and the others as rural.

Image manipulation

The purpose of the image manipulation was to generate images spatially similar to the originals but with variable color errors. The errors were constrained to preserve the average color and to avoid generating pixelation artifacts (as would happen if individual pixels were changed randomly). The procedure for generating color errors is illustrated in Fig. 2. First, CIELAB space was segmented into cubes of side 4 units, starting from L^* equal to zero and from the minimum of a^* and of b^* . Next, the CIELAB coordinates for each image pixel were calculated and the cube containing those coordinates identified. To each set of pixel coordinates inside each cube, a vector was added with constant CIELAB magnitude but direction that varied randomly from cube to cube, thereby ensuring that groups of similar colors were changed in the same way. The specific size of the segmenting cube defined the coarseness of the image approximation and was chosen empiri-



Fig. 1. Color pictures of the 9 scenes tested.

cally as a compromise between not introducing spatial artifacts and providing a reasonable dynamic range. For each original image, ten approximate images were thus generated with color errors varying randomly in direction within each image and of magnitude ranging from 0.5 to 5.0 in 0.5 steps across images.

Image display

A 17-inch RGB color monitor (model GDM-F400T9; Sony Corp., Tokyo, Japan) controlled by a computer with raster-graphics card

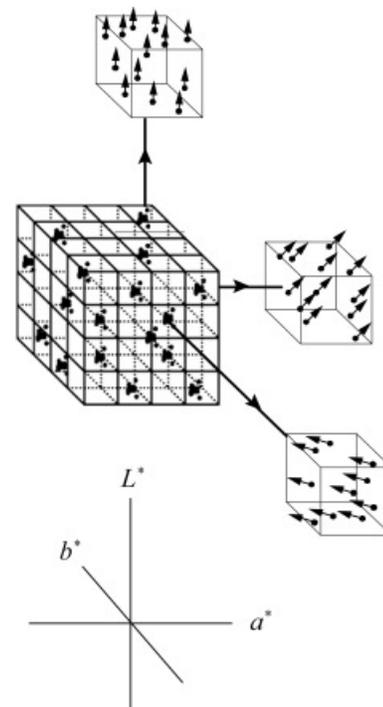


Fig. 2. Diagram of the procedure used to generate color errors. CIELAB color space was segmented into cubes of side 4 units and each image pixel was located within one of the cubes. Groups of similar colors defined by each cube were changed by adding a vector with constant CIELAB ΔE_{ab}^* magnitude but with random direction from cube to cube.

providing 24 bits per pixel in true-color mode (VSG 2/5; Cambridge Research Systems, Rochester, UK) was used to display the images. Screen resolution was 750×600 pixels and refresh rate was 80 Hz. The display system was regularly calibrated in luminance and chromaticity with a telespectoradiometer (Spectra Colorimeter, PR-650, Photo Research Inc, Chatsworth, CA). The images were all displayed with an average luminance of 15 cd m^{-2} . The percentage of pixels out of gamut in the displayed images was, on average, less than 5%. These out-of-gamut pixels were each displayed by clipping to the closest displayable RGB value; that is, they were each assigned realizable coordinates in the monitor RGB space that were as close as possible to the original coordinates. This procedure affected the color error by on average 0.03 CIELAB units.

To assess the accuracy of the CRT monitor in reproducing the small color error differences required in the experiment, a set of 40 pixels were selected at random from three original images. For each, colored patches with the CIELAB coordinates of the original pixel and the corresponding patches from the images with nominal 0.5 and 1.0 CIELAB color errors were displayed in the CRT screen and measured with the telespectoradiometer. The average color error over the 40 pixels in CIELAB units for the nominal 0.5 error was 0.53 with an SD of 0.27 and for a nominal 1.0 error was 1.03 with an SD of 0.29. Thus, the precision of the monitor in reproducing small color errors was adequate for the present experiments.

For display purposes all images were used with a spatial resolution reduced by a nearest-neighbor interpolation routine to one quarter of the original size, that is, 336×256 pixels. Each test image in the pair subtended $9^\circ \times 7^\circ$ visual angle and the pair was separated by a black gap of 0.6° . Viewing distance was 1 m.

Procedure

In each trial of the experiment a pair of images was presented to the observer. One image was always the original and the other was an approximation with a specific magnitude of color error. The pair was presented for 3 s on a black background followed by a 3-s interval before the next trial, during which the screen remained black. The location of the original image on the right or left of the screen and the error in the approximate image were randomized from trial to trial. The task of the observer was to decide whether the images were identical. Responses were made after the presentation interval with a switch box (CB6 Response Box, Cambridge

Research Systems, Rochester, UK) connected to the computer. The experiment was performed in a darkened room. Nine images with 10 color error levels were tested and each possible pair was presented to the observer in random order. Different images were tested in different sessions of about 30 min each in a different order for each observer. Each observer performed 20 trials at each error level and therefore a total of 1800 trials for the complete experiment. The experiment took 2–4 days depending on the observer.

Observers

Six observers (J, L, M, P, T, and E), aged 22 to 26 years, one female and five male, performed the experiment. Three were unaware of the purpose of the experiment and three were co-authors of the paper. All had normal or corrected visual acuity and normal color vision assessed with Rayleigh anomaloscopy and the Farnsworth-Munsell 100-Hue test. Informed consent was obtained from all participants and the research was conducted in accordance with principles embodied in the Declaration of Helsinki.

Results

Fig. 3 shows, as an example, performance by two observers with two scenes. Symbols represent percentage of “same” responses as a function of the CIELAB color error ΔE_{ab}^* . Different symbols correspond to different scenes. The smooth curves were obtained by fitting a cumulative Gaussian psychometric function to the data. The 50% threshold value was estimated for each image and observer from the fitted psychometric function. Vertical bars represent ± 1 standard error (SE) of the threshold estimated by a bootstrap procedure based on 1000 replications (Foster & Bischof, 1991). These observers could distinguish the original from the approximate image with a ΔE_{ab}^* error of about 1.6 and 2.0 (Fig. 3, left panel) and 2.5 and 2.8 in the other (Fig. 3, right panel).

The open circles in Fig. 4 show 50% threshold values expressed in CIELAB ΔE_{ab}^* units for all observers and scenes tested. Horizontal dotted lines represent averages across observers. For scene 1 it was not possible to fit a psychometric function for each observer because individual data were too noisy and the data represented are from a fit across observers. Over all scenes, thresholds ranged from about 1.2 to about 4.0, with an average of 2.2 and SE of 0.28. For urban scenes, the average was 1.5 and for rural scenes 2.5. The thresholds for observer T in scenes 8 and 9

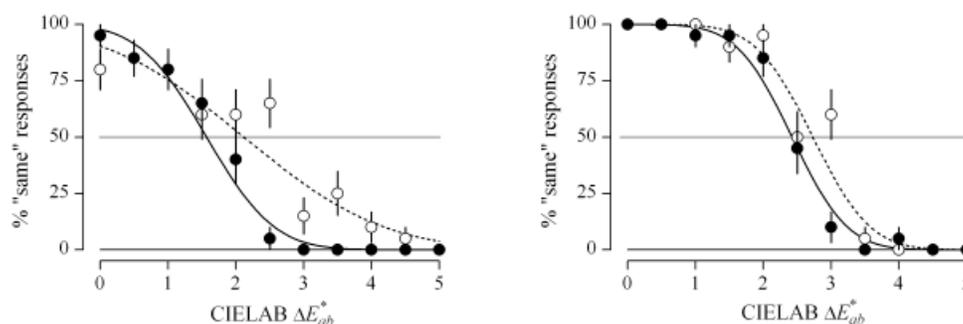


Fig. 3. Percent “same” responses as a function of color error introduced into images. Each panel contains data for one observer and two scenes (open and filled symbols). Scores were based on 20 trials at each error level, and the vertical bars show ± 1 binomial SE. The smooth curves were obtained by fitting a cumulative Gaussian psychometric function to the data.

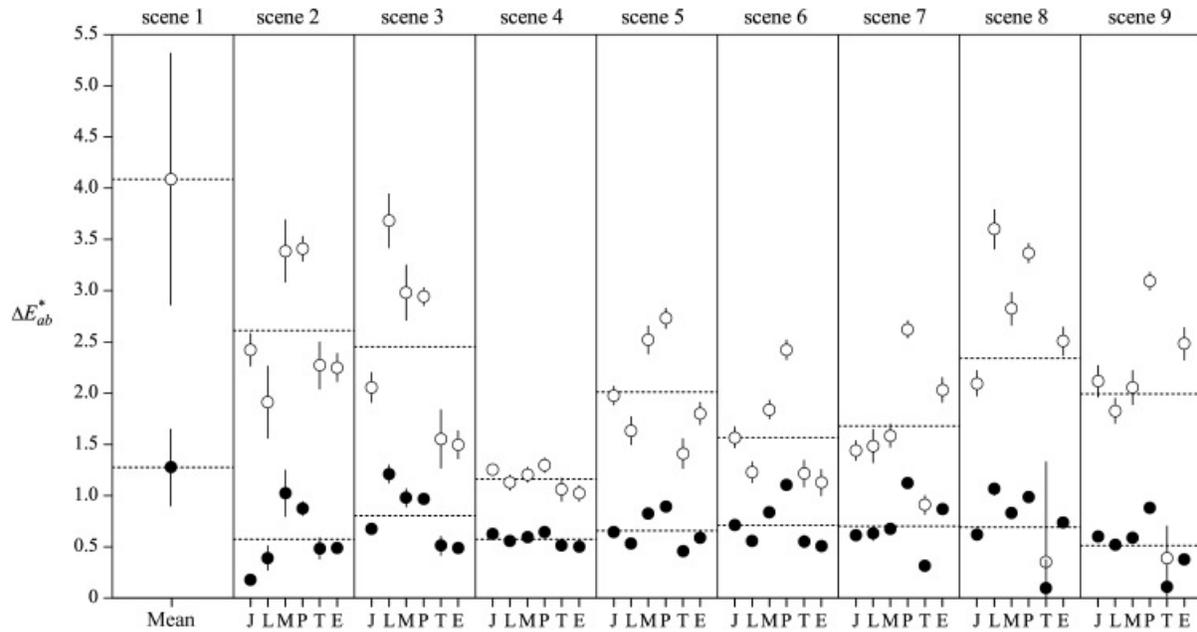


Fig. 4. Threshold color errors corresponding to 50% discrimination for all observers and scenes expressed in CIELAB units (open circles) and S-CIELAB units (filled circles). Error bars represent ± 1 SE estimated from a bootstrap based on 1000 replications. Horizontal dotted lines represent averages across observers. (For scene 1 it was not possible to fit a psychometric function for each observer because individual data were too noisy and the data represented are from a fit across observers.)

were low and may have been caused by some localized feature identified by the observer; they are, however, clearly different from the thresholds of the rest of the group. The variation in performance across scenes was statistically significant (repeated measures ANOVA excluding scene 1, $F(7,35) = 5.4$ $P < 0.01$). The average slope of the psychometric function across images was -1.4 with a standard deviation of 1.2 ; no significant correlation was found between slope and thresholds ($P > 0.2$). Threshold values obtained from the first half of the trials were similar to those obtained from the second half, suggesting that learning effects were not important.

The filled circles in Fig. 4 show 50% thresholds expressed in S-CIELAB ΔE_s units. To compute these threshold values, the average error in S-CIELAB units was computed for each of the 10 images of each scene and threshold was then computed in the same way as for CIELAB. The parameters used for S-CIELAB computations were those of Zhang & Wandell (1996). Over all scenes, thresholds ranged from about 0.5 to about 1.2, with an average of 0.7 and SE of 0.07. For urban scenes, the average was 0.6 and for rural scenes 0.8. The variation in performance across scenes was not statistically significant (repeated measures ANOVA excluding scene 1, $F(7,35) = 1.7$ $P > 0.1$). The average slope of the psychometric function across images was -3.5 with a standard deviation of 1.9 ; no significant correlation was found between slope and thresholds ($P > 0.5$).

Discussion

The psychophysical data presented here show that an average CIELAB color difference ΔE_{ab}^* of at least 2 is necessary for a color error to be detected in complex images of natural scenes derived from hyperspectral data. This result is consistent with data from previous studies (Stokes et al., 1992; Song & Luo, 2000) using

images from RGB cameras. In a study of the number of spectral-reflectance basis functions needed to reproduce natural scenes (Nascimento et al., 2005) observers were able to distinguish pairs of images with average CIELAB color differences $\Delta E_{ab}^* < 1$ but with local differences > 3 .

As noted earlier, because the images were here obtained from hyperspectral data, their color gamut was unconstrained by the physical limits that would normally be set by an RGB image-acquisition device, and were determined solely by the display device, affecting about 5% of pixels. The size of the average threshold for color errors suggests that the accurate reproduction of the original scene may not be a fundamental issue.

There was, however, a significant variation in performance across scenes when color errors were expressed in CIELAB space, with smaller threshold values for urban scenes, probably because of the presence of large uniform surfaces. This result is consistent with the fact that urban scenes need slightly smaller differences to become chromatically indistinguishable (Nascimento et al., 2005). No dependence on scene content was found in another previous study by Stokes et al. (1992) but their test image set contained only one image of a rural environment.

When thresholds were expressed as S-CIELAB color differences ΔE_s , the variation across scenes was smaller than that in CIELAB space and not statistically significant, an indication that the image manipulations used here correctly took into account the differing spatial structures of the scenes. The threshold values obtained in this representation are consistent with values obtained for halftone images (Zhang et al., 1997). For most practical applications with complex colored images a single number independent of the image content seems to be sufficient to describe the perceptibility of color differences as long as they are expressed in S-CIELAB color space.

Acknowledgments

This work was supported by the Fundação para a Ciência e Tecnologia, (grant no. POSI/SRI/40212/2001), by the Centro de Física of Minho University, Braga, Portugal, and by the Engineering and Physical Sciences Research Council (grant nos. GR/R39412/01 and EP/B000257/1). João M.M. Linhares was supported by the Fundação para a Ciência e a Tecnologia, Portugal.

References

- BERNS, R.S. (2001). The science of digitizing paintings for color-accurate image archives: A review. *Journal of Imaging Science and Technology* **45**, 305–325.
- CHEUNG, V., WESTLAND, S., LI, C.J., HARDEBERG, J. & CONNAB, D. (2005). Characterization of trichromatic color cameras by using a new multi-spectral imaging technique. *Journal of the Optical Society of America A-Optics, Image Science and Vision* **22**, 1231–1240.
- FAIRCHILD, M.D. (2005). *Color Appearance Models*. John Wiley & Sons Ltd.
- FAIRCHILD, M.D. & JOHNSON, G.M. (2004). iCAM framework for image appearance, differences, and quality. *Journal of Electronic Imaging* **13**, 126–138.
- FOSTER, D.H. & BISCHOF, W.F. (1991). Thresholds from psychometric functions: Superiority of bootstrap to incremental and probit variance estimators. *Psychological Bulletin* **109**, 152–159.
- FOSTER, D.H., NASCIMENTO, S.M.C. & AMANO, K. (2004). Information limits on neural identification of colored surfaces in natural scenes. *Visual Neuroscience* **21**, 331–336.
- IMAI, F.H., WYBLE, D.R., BERNES, R.S. & TZENG, D.Y. (2003). A feasibility study of spectral color reproduction. *Journal of Imaging Science and Technology* **47**, 543–553.
- MOROVIČ, J. & MOROVIČ, P. (2003). Determining colour gamuts of digital cameras and scanners. *Color Research and Application* **28**, 59–68.
- NASCIMENTO, S.M.C., FERREIRA, F.P. & FOSTER, D.H. (2002). Statistics of spatial cone-excitation ratios in natural scenes. *Journal of the Optical Society of America A-Optics, Image Science and Vision* **19**, 1484–1490.
- NASCIMENTO, S.M.C., FOSTER, D.H. & AMANO, K. (2005). Psychophysical estimates of the number of spectral-reflectance basis functions needed to reproduce natural scenes. *Journal of the Optical Society of America A-Optics Image Science and Vision* **22**, 1017–1022.
- PÁRRAGA, C.A., BRELSTAFF, G., TROSCIANKO, T. & MOOREHEAD, I.R. (1998). Color and luminance information in natural scenes. *Journal of the Optical Society of America A-Optics, Image Science and Vision* **15**, 563–569.
- SONG, T. & LUO, R. (2000). Testing color-difference formulae on complex images using a CRT monitor. *Proceedings Eighth IS&T/SID Color Imaging Conference, IS&T*, pp. 44–48.
- STOKES, M., FAILCHILD, M.D. & BERNES, R.S. (1992). Colorimetrically quantified tolerances for pictorial images. *TAGA part 2*, pp. 757–778.
- WEBSTER, M.A. & MOLLON, J.D. (1997). Adaptation and the color statistics of natural images. *Vision Research* **37**, 3283–3298.
- WU, W.C., ALLEBACH, J.P. & ANALOUI, M. (2000). Imaging colorimetry using a digital camera. *Journal of Imaging Science and Technology* **44**, 267–279.
- ZHANG, X., SILVERSTEIN, D.A., FARRELL, J.E. & WANDELL, B.A. (1997). Color image quality metric S-CIELAB and its application on half-tone texture visibility. In *COMPCON97 Digest of Papers*, pp. 44–48. IEEE.
- ZHANG, X. & WANDELL, B.A. (1996). A spatial extension of CIELAB for digital color image reproduction. *Proceedings of the SID Symposiums*, pp. 731–734.