https://doi.org/10.2352/CIC.2024.32.1.8

CNP2D: The Chromatic Noise Pattern Discrimination Dataset

***Trevor D. Canham,*[1] *Peter Morovič,*[2] *Richard F. Murray,*[1] *Jan Morovič,*[2] *and Michael S. Brown*[1]**
[1] ***York University, Canada;*** [2] ***HP Inc.;***

Abstract

Pixel-wise color difference metrics like ΔE_{00} have long been used in image analysis, but it remains unclear how scores should be integrated over space. To highlight this, a psychophysical experiment was conducted to characterize visual sensitivity to differences in chromatic noise patterns in different color and pattern contexts. The results demonstrated that observers were more sensitive to chromatic noise pattern (CNP) differences when similar colors were spatially dispersed over the pattern as opposed to clustered. Further analysis with common image color and texture difference metrics showed that none were sensitive to this effect. This finding highlights the need for metrics which capture the perceptual interaction between color and texture.

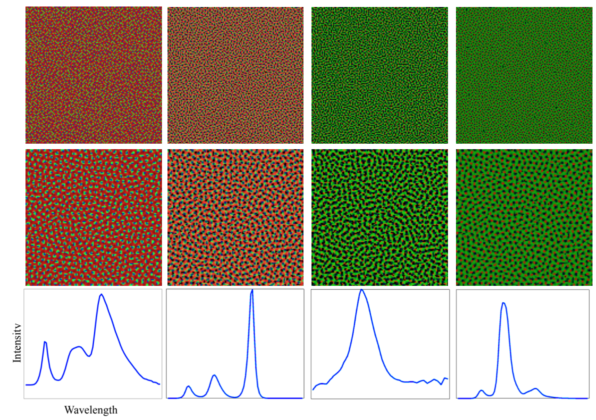

Figure 1. Noise prism [2] visualizations of metamer pairs (columns 1&2 and 3&4), rendered with two different color clustering levels using the PARAWACS technique. Global color distribution characteristics are maintained over pattern types, reflecting the distribution shown in the bottom row.

Introduction

Spatial integration of pixel-wise visual difference measurements over complex stimuli (e.g., images, textures) is an open problem. Several classes of quantitative metrics exist which focus on particular axes of complex stimuli like color [19], texture [3] and structure [17], but the proper way to sum their contributions to visual sensitivity over space and time is unknown. While a large number of experiments have been conducted with the goal of identifying decorrelated perceptual axes in human observers [1, 7, 8, 13, 15, 14], it is difficult to scale these experiments to address the full complexity of the problem.

Texture and structure (i.e., shape), for example, are not precisely distinguished by any image processing operators and the perceptual and ontological thresholds between these features is unknown. This is demonstrated in the well-cited work of Geirhos et al. [5] which aimed to analyze the classification performance of neural networks in the context of "texture" and "shape" images, but several prominent examples of stimili in the "texture" category are actually composed of the shapes of objects.

Meanwhile, many engineering applications involving the display of complex stimuli (e.g., compression, image enhancement, data visualization, etc.) require effective visual difference measures to function. In this paper, we focus on recent visualization tools that use chromatic noise patterns (CNPs) to elucidate multivariate phenomena (e.g., multispectral data [2], demographic choropleth maps [6]). This visualization technique, known as color weaving, involves assigning primary colors to multivariate data and encoding the variables' magnitude in the number of pixels of its associated primary color present in a display. An example is shown in Fig. 1, where the noise prism technique [2] is used to differentiate between illumination metamers (colors which are illuminated by spectrally different sources but look the same to observers). Despite their utility, optimizing CNPs for visualization is an open problem due to a lack of effective visual difference measures.

A possible avenue for modulating the legibility and esthetic appearance of CNP visualizations is the manipulation of their spatial distribution pattern (as between the rows of Fig. 1). However, in this application its essential to preserve the color distribution of the visualizations since they have been analytically rendered to represent physical samples. Towards this end, Parallel Random Weighted Area Coverage Selection (PARAWACS) is employed to manipulate the spatial distributions of noise prism patterns. Demonstrated in the rows of Fig. 1, PARAWACS is a pattern manipulation technique designed for printer halftoning which modulates the degree to which similar colors cluster together. The patterns shown in different rows are noticeably distinct with respect to their spatial distribution but their color histograms are maintained - a unique feature among CNP datasets.

Using these stimuli, a psychophysical experiment was conducted to characterize observer sensitivity to CNP pairs with asymmetric distribution properties in the context of different colors and patterns. The experimental stimuli were also processed through color, texture and structure metrics and their relative average scores over the independent variables of illumination metamer pair, color, and pattern type are compared to those of observers.

The psychophysical experiment showed all independent variables had a significant effect on the discriminability of stimuli. In particular, there was a significant interaction between color and pattern type. The results of the metrics experiment showed that none of the those tested were able to consistently predict the relative sensitivity to different colors or predict the effect of spatial pattern type in the correct direction despite demonstrating high overall correlations to observer data.

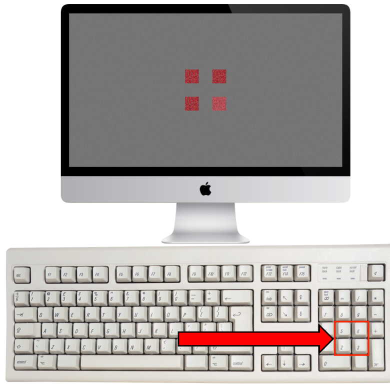

Figure 2. *Laboratory setup and stimuli for 4AFC experiment. Observers are asked to choose which patch is different from the other three.*

Background

In his work which served as the foundation for the field of texture analysis, Julesz [7] showed that textures can be well approximated by matching statistical, topological, or heuristic properties and identified limits for the discriminability of random Markov patterns along these axes. In particular, his findings showed that changes in higher-order marginal moments of the luminance distribution like skewness and kurtosis are not discriminable so long as the mean and variance of the distribution are held constant. However, Pratt et al. [13] showed in later experiments that with alternative statistical techniques noise patterns could be created which are discriminable but differ in skewness alone. This finding demonstrated the dependence of experiments of the type of Julesz [7] on their stimulus generation techniques.

In a similar vein to the proposed experiments, te Pas et al. [15] compared discriminability thresholds between solid colors and color texture patches. Stimuli separated by a hard edge resulted in color texture discrimination thresholds that were 15 times those of the solid color stimuli, but the threshold disparity was reduced to 1.5 times when the hard edge was smoothed over. Seybold [14] & Meininger [8] both explored chromatic noise pattern visibility for video applications and observed low pass characteristics with respect to spatial and temporal frequency. However, these experiments focus on noise stimuli with the distribution characteristics of image sensors.

Inspired by the prism, Canham et al. [2] use chromatic noise patterns (CNPs) to visualize multispectral samples, where each spectral band is represented by a unique in-gamut color and its relative intensity is encoded in pattern area coverage. The method's novelty in the context of existing multispectral visualization techniques is that primaries representing the spectral bands are distributed spatially instead of additively blended [12]. Hagh-Shenas et al. [6] showed that color blending has an inferior information-carrying capacity for general multivariate data compared to spatially distributed primaries (color weaving).

The noise prism generation process was outlined in the following way. The spectral power distribution of a given color sample is area normalized and re-scaled according to the desired total number of pixels in the visualization, resulting in a pixel count for each band. A pattern is then populated with the associated tristimulus values of the CIE 1931 2° standard observer at each band. These tristimulus values are then gamut mapped to the destination display space via saturation scaling in IPT space (where hue and saturation are well decorrelated [4]). Finally, a channel-wise exponent is applied to compensate for color balance shifts from the original sample.

This prior work features a four alternative forced choice (4AFC) psychophysical experiment to demonstrate that by visualizing image regions encoded as CNPs, color sample pairs that would otherwise be perceived by a standard observer as equivalent due to trichromatic integration (metamers) could be discriminated (as in Fig. 1). The stimuli were generated from a set of illumination metamer pairs. For each pair, a four-square arrangement of patches subtending two degrees was shown, where three patches were generated from one spectral power distribution and the fourth from a metameric spectral power distribution. Observers were asked to report the position of the unique patch and their accuracy was recorded. The experiment showed that illumination metamer pairs in different color contexts ranged from obviously discriminable (100% accuracy across all observers and trial repeats) to completely indiscriminable.

The present work presents a similar experiment with the additional independent variable of pattern type. Parallel Random Weighted Area Coverage Selection (PARAWACS) is employed to generate experimental stimuli which differ in pattern (spatial distribution) but maintain roughly the same color distribution. Originally developed for the purpose of halftoning in the context of the HANS printing pipeline [9], PARAWACS uses a single pattern matrix (or selector matrix) to make choices from a probability distribution of possible states at a given pixel. The pattern matrix encodes a continuous-tone range of uniform-likelihood values, distributed in a desired spatial arrangement pattern, such as blue-noise (dispersed) and green-noise (clustered), with others also possible. Given a pattern matrix with a tone value at a given pixel, the choice of state (e.g., spectral primary) is then made by comparing the pixel value against the cumulative probability distribution at the pixel (e.g., the spectral power distribution). This, coupled with the uniform distribution of the pattern matrix values, ensures that the encoded phenomenon (e.g., the spectral power distribution) is represented correctly over some patch area and that the pattern of the source matrix is present in the noise prism.

Methodology

A four-alternative-forced-choice (4AFC) experiment was conducted to assess the degree of visual difference between chromatic noise pattern visualizations of metameric spectral power distribution pairs generated with the noise prism and PARAWACS techniques over different colors and pattern types. The same stimuli were passed to image color, texture and structure difference metrics to compare their predictions with observer judgements.

Participants

21 students and staff between the ages of 21 and 34 from York University participated in the experiment. Each observer was verified to have normal color vision with an Ishihara color deficiency test. They received a small gift as incentive for their participation.

Apparatus

The experiment employed a 2014 iMac as a computing platform and display, running software coded in MATLAB using the Psychtoolbox [11]. The software managed the cadence of the experiment, displaying stimuli, querying the input device, and saving participant performance data. The display output was measured for primary colors and white at 16 drive values with a PhotoResearch PR-655 spectrophotometer to verify additivity and consistency across color regions. The display's gamma function and primary matrix were derived from the measurements to accurately represent noise pattern colorimetry. The experiment was conducted in a dark environment for ease of repeatability. Observer responses were submitted on a standard QWERTY keyboard.

Stimuli

Observers were presented with four noise prism patterns corresponding to the same patch reflectance. Each pattern subtended two degrees of visual angle at the display's native resolution and were arranged as in Fig. 2. Three out of the four patterns were generated with the same illumination source, while the fourth was generated with a metameric alternative. The patch reflectances were a subset of X-Rite Macbeth color checker patches representing a range of important memory colors (e.g., skin tones, foliage, etc.). The light sources included the CIE D65 "daylight" standard and measured SPDs from CRT, OLED, and two LED devices which are shown in Figure 3. Display spectra were chosen since their primaries could be conveniently re-scaled on a patch-by-patch basis to produce the same CIEXYZ values as the D65 illumination case, rendering all stimuli for the same patch reflectance as metameric matches. Given these SPDs, the stimuli are computed using the noise prism methodology described in the previous section. The full CIE D65-illuminated macbeth chart passed through the stimuli generation processes is shown in Figure 4.

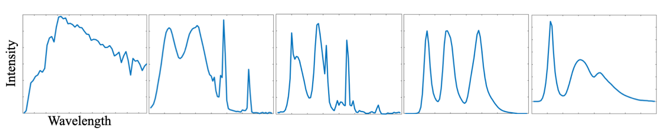

Figure 3. *Illuminant spectral power distributions measured from 380-780 nanometers. From left to right, the illuminants are D65, CRT, LED, OLED, and an alternative LED (LED2).*

Procedure

At the start of each session the experimental instructions were read aloud, detailing the cadence of the experiment and task (i.e., to select the patch that is different from the other three). Before starting the body of the experiment, the observers practiced the experiment with 5 random trials. For the main experiment, the conditions were presented to observers in a randomized order. Patches were flashed for 1 s, then disappeared. The next trial began immediately after participants submitted their responses. Observers completed the experiment in 15 - 30 minutes.

Design

The experiment employed an $8 \times 10 \times 2$ between-subjects design with the following independent variables and levels:

- 8 patch reflectances (dark skin, light skin, blue sky, foliage, purplish blue, moderate red, green, and blue)
- 10 illumination metamer pairs (all unique combinations of Figure 3)
- 2 spatial pattern arrangements (PARAWACS [10] blue noise and green noise)

The dependent variable was discrimination accuracy. Higher accuracy scores indicate larger perceptual differences between stimulus pairs. The experiment was conducted over two separate sessions, where half of the color levels were tested on each day. Seven observers participated in both sessions, while the remaining 14 participated in one of the two sessions. The total number of trials was 16,960 (= 21 observers $\times$ 8 $\times$ 10 $\times$ 2 $\times$ 4 to 16 repeats)

Metrics

A set of commonly used image color, texture, and structure metrics were tested for their ability to quantify relative perceptual differences between CNPs. We used three different metrics representing each of the above categories to evaluate stimuli pairs and measured the correlation of the resulting scores to the observer data in terms of Pearson Linear Correlation Coefficient (PLCC), Spearman Rank Order Correlation Coefficient (SROCC) and coefficient of determination (R^2).

Our chosen metrics were as follows. First, we applied a spatial extension of CIELAB (ΔE_S) proposed for quantifying halftone visibility [18]. The measured white point of the monitor from the experimental setup was used to properly compute CIELAB values. Next we applied Deep Image Structure and Texture Similarity (DISTS), a state of the art texture metric from Ding et al. [3] trained on color images for resistance to pixel-wise registration errors. Finally, we computed the structural similarity index metric (SSIM [17]) to individual CIELAB channels and compute their average.

Results

The overall mean accuracy score was 75%. Observers were able to discriminate pairs of stimuli with blue noise spatial patterns more accurately (77%) than pairs with green noise spatial patterns (72%). The effect of independent variables pattern type ($F_{1,160} = 7.57$, $p < 0.005$) shown in Fig. 5, color ($F_{7,160} = 92.28$, $p < 0.001$) shown in Fig. 6 and illumination metamer pair ($F_{9,160} = 488$, $p < 0.001$) shown in Fig. 7 were identified as significant according to an N-way ANOVA test. The test also revealed a significant interaction between color and pattern type ($F_{7,160} = 2.45$, $p < 0.05$) shown in Fig. 8. The interaction between illumination metamer pair and pattern type was not found to be significant ($F_{9,160} = 0.94$, $p > 0.05$). The results of the chosen metrics ΔE_S [18], DISTS [3] and SSIM [17] were averaged over independent variable levels (Fig. 9). The correlation scores for each metric over the full stimuli set are presented in Table 1.

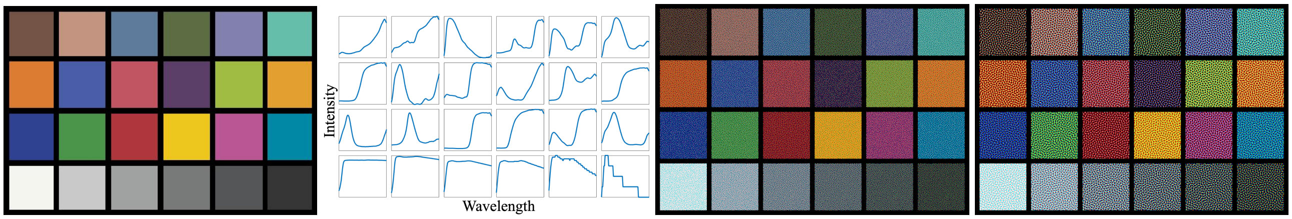

Figure 4. *From left to right, CIE D65 illuminated X-Rite Macbeth color checker chart, reflectance as a function of wavelength, noise prism rendering with PARAWACS blue noise spatial pattern, and noise prism rendering with PARAWACS green noise spatial pattern.*

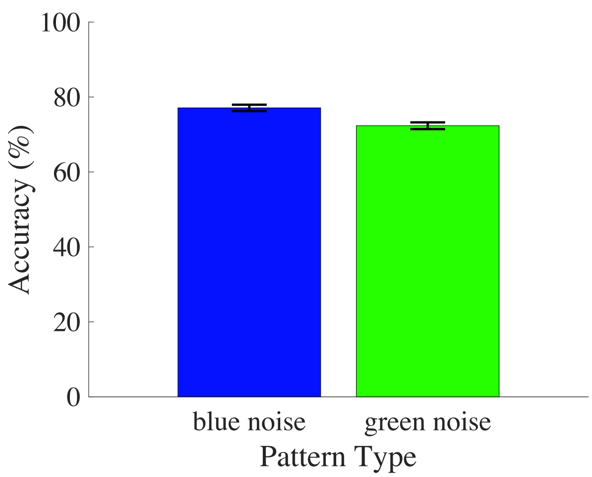

Figure 5. *Mean observer accuracy as a function of pattern type. Pairs of blue noise patterns were easier to discriminate on average than pairs of green noise patterns.*

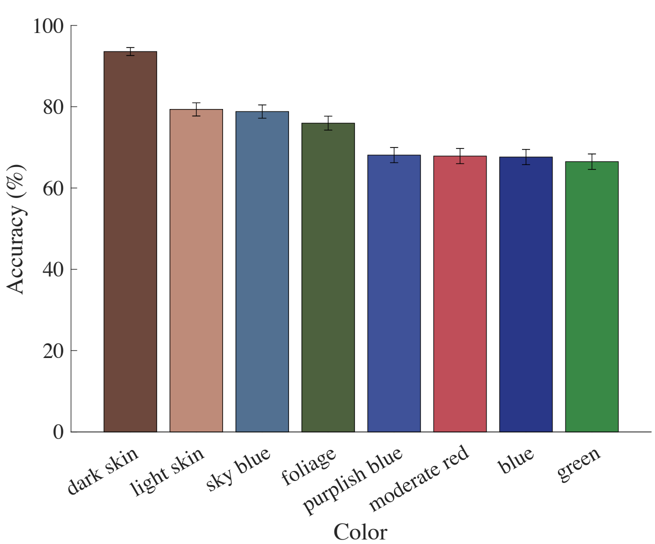

Figure 6. *Mean observer accuracy as a function of color.*

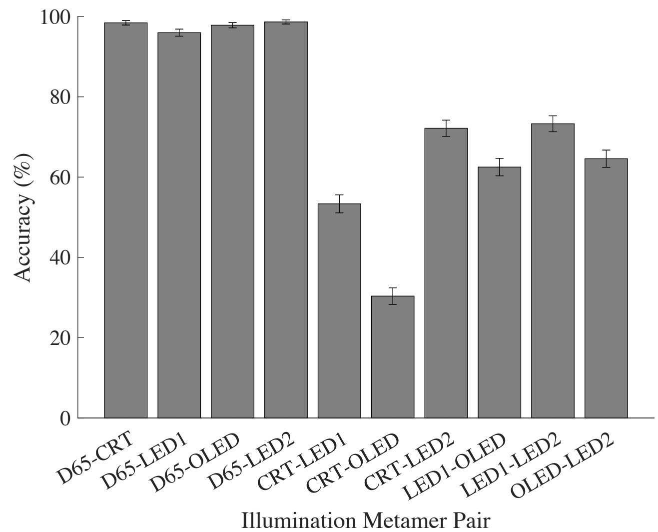

Figure 7. *Mean observer accuracy as a function of Illumination Metamer Pair. Broad band vs. narrow band illumination metamer pairs (e.g. D65-CRT) are easy for observers to discriminate, while pairs of narrow band spectral power distributions are harder.*

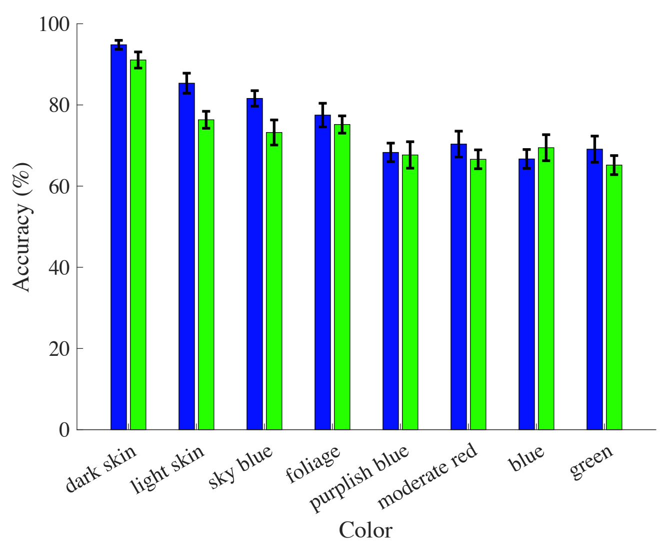

Figure 8. *Mean observer accuracy as a function of color split by pattern type. Differences between pattern types are significant with dark skin, light skin and sky blue but not with more saturated colors.*

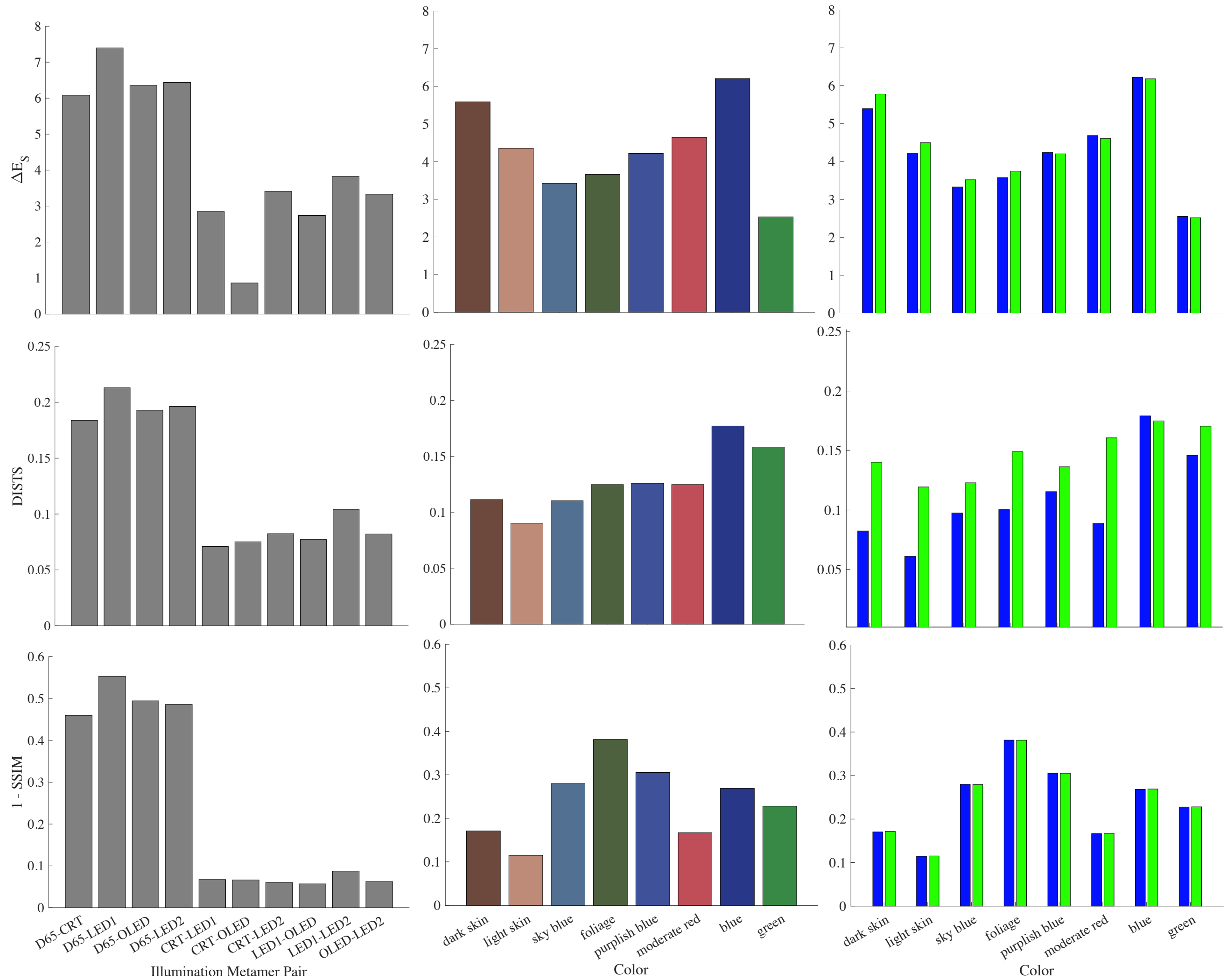

Figure 9. Metric results over experimental stimuli pairs. From top to bottom, ΔE_S, DISTS and SSIM mean results over independent variables illumination metamer pair, color, and color split by pattern type.

Table 1. Correlation between metrics sensitive to color (ΔE_S), texture (DISTS) and image (SSIM) distortions and observer accuracy scores.

	PLCC ↑	SROCC ↑	R^2 ↑
ΔE_S [18]	**0.73**	**0.87**	**0.54**
DISTS [3]	0.52	0.56	0.27
SSIM [17]	0.58	0.71	0.33

Discussion

The experiments showed that observers were more sensitive on average to chromatic noise pattern (CNP) differences with blue noise patterns (which minimize color clustering) than green noise patterns (which encourage clustering) in the context of the stimulus presentation scheme and task of these experiments (Fig. 5). However, Figs. 1 & 4 show blue noise patterns have the effect of encouraging spatial integration of pixels, while the green noise pattern makes the individual primaries more obvious.

One explanation for this result is the short stimulus presentation time of 1 s. During this time, observers have to assess four spatially displaced patterns before they disappear. Discrimination may be encouraged by spatial integration in this context because the timescale of the experiment targets pre-attentive processing which operates on summary statistics over larger spatial extents [16]. The work of Balas et al. [1] suggests that the marginal moments of a texture synthesis algorithm are the more salient when lesioned than various primaries of spatial distribution (e.g., auto correlation), but these experiments were conducted at a 250 ms timescale. Indeed, many Julesz-type studies share the original author's focus on features which constitute "immediate, effortless perception, without scrutiny" by adopting similarly short stimulus presentation timescales [7]. This said, an interesting future direction of this work would be to extend the stimulus presentation timescale with an alternative task to allow for cognitive analysis of the patterns.

Interestingly, Fig. 8 shows that the effect of spatial pattern arrangement is more pronounced in some colors over others - in particular skin tones and blue sky, whereas more saturated colors were not significantly affected by the spatial patterns tested here. Also, the illumination metamer pairs which served as asymmetric distribution basis functions around a common color center had no cross effect with pattern type.

Fig. 7 shows that the accuracy scores averaged over illumination metamer pairs had a strong bimodal distribution between comparisons against D65 (a broad band illumination source) and between the various narrow band sources shown in Fig. 3. Fig. 9 shows that all of the tested metrics were sensitive to this bimodal distribution. ΔE_S performed particularly well in this regard, identifying the correct ranking for nearly all significant differences across the variable set. This adherence to measured phenomena is reflected in the overall correlation results. However, the average accuracy scores for different color centers were not properly reflected by any metrics. Also, while DISTS and ΔE_S were sensitive to the effect of pattern type, neither metric predicted the results in the proper direction. Finally, SSIM was not sensitive to the effect of pattern type. This is expected as the metric targets distribution moments over local areas and the color distributions of different pattern types were roughly the same. These results highlight the need for a new metric to account for the interaction between color and texture in the human visual system.

Conclusion

A psychophysical experiment was conducted to characterize sensitivity to chromatic noise patterns (CNPs) with varying distribution properties in different color and pattern contexts. The stimuli were generated with a visualization technique which allows users to discriminate between illumination metamers [2], and a pattern manipulation technique which allowed for the clustering of common colors in the pattern to be modulated [10]. Observers were tasked with picking out the metamer in a four-alternative-forced-choice experiment. The results demonstrated that observers were more sensitive to CNP differences when the stimuli were generated with a pattern formation function which discourages clustering of similar colors than when the stimuli were generated with a function which encouraged clustering, particularly in the context of important memory colors like sky blue and skin tones.

The stimuli pairs were then measured by well known image color and texture metrics [18, 3, 17] and compared to the observer accuracy scores across independent variables. While the metrics roughly followed the effect of illumination metamer pairing, none properly captured the color and pattern effects, highlighting the need for new metrics which account for the interaction between color and texture properties in human vision. This is one promising direction for future work. Another promising direction would be a longer time-scale experiment using these stimuli.

Acknowledgments

The authors would like to thank the many observers who participated in the experiment. This work was supported by the Canada First Research Excellence Fund for the Vision: Science to Applications (VISTA) program, an NSERC Discovery Grant, and the Canada Research Chair program.

References

[1] Benjamin J Balas. Texture synthesis and perception: Using computational models to study texture representations in the human visual system. *Vision Research*, 46(3):299 – 309, 2006.

[2] Trevor D. Canham, Javier Vazquez-Corral, David L. Long, Richard F. Murray, and Michael S. Brown. Noise prism: A novel multispectral visualization technique. In *CIC*, 2023.

[3] Keyan Ding, Kede Ma, Shiqi Wang, and Eero P. Simoncelli. Image quality assessment: Unifying structure and texture similarity. *IEEE TPAMI*, 44(5):2567–2581, 2022.

[4] Fritz Ebner and Mark Fairchild. Finding constant hue surfaces in color space. In *Photonics West '98 Electronic Imaging*, pages 4207–4216, 1998.

[5] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A Wichmann, and Weiland Brendel. Visibility of spatiotemporal noise in digital video. In *CIC*, 2016.

[6] Haleh Hagh-Shenas, Sunghee Kim, Victoria Interrante, and Christopher Healey. Weaving versus blending: a quantitative assessment of the information carrying capacities of two alternative methods for conveying multivariate data with color. *IEEE TVCG*, 13(6):1270–1277, 2007.

[7] Bela Julesz. Visual pattern discrimination. *IRE transactions on Information Theory*, 8(2):84–92, 1962.

[8] Catherine Meininger. Determining visibility thresholds for spatial and spatiotemporal chromatic noise. *SMPTE Motion Imaging Journal*, 128(2):31–40, 2019.

[9] Jan Morovič, Peter Morovič, and Jordi Arnabat. Hans: Controlling ink-jet print attributes via neugebauer primary area coverages. *IEEE TIP*, 21(2), 2012.

[10] Peter Morovič, Jan Morovič, Jay Gondek, and Robert Ulichney. Direct pattern control halftoning of neugebauer primaries. *IEEE TIP*, 26(9), 2017.

[11] Dennis G. Pelli. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, 10:437–442, 1997.

[12] Gerrit Polder and Gerie W.A.M. van der Heijden. Visualization of spectral images. In *SPIE*, 2001.

[13] William K Pratt, Olivier D Faugeras, and Andre Gagalowicz. Visual discrimination of stochastic texture fields. *IEEE TSMC*, 8(11):796–804, 1978.

[14] Tamara Seybold, Betina L. Koelln, Aynur Pasha, and Harald Brendel. Visibility of spatiotemporal noise in digital video. In *CIC*, 2016.

[15] Susan F te Pas and Jan J Koenderink. Visual discrimination of spectral distributions. *Perception*, 33(12):1483–1497, 2004.

[16] Anne Treisman. Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing*, 31, 1985.

[17] Zhou Wang and A.C. Bovik. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 13(4), 2001.

[18] Xuemei Zhang, D.A. Silverstein, J.E. Farrell, and B.A. Wandell. Color image quality metric s-cielab and its application on halftone texture visibility. In *IEEE COMPCON*, 1997.

[19] Xuemei Zhang and Brian A. Wandell. A spatial extension of CIELAB for digital color-image reproduction. *Journal of the Society for Information Display*, 5(1):61–63, 1997.