Studying Empirical Color Harmony in Design

Sam Goree David Crandall
Indiana University

Abstract

Color is a major element of design, and especially in fashion. Previous work on data mining of colors has focused on individual colors or color histogram representations, but not models of the relationships between colors. We investigate hue-relative unsupervised generative modeling for color schemes and apply this model as a measurement tool to better understand how the relationship between online color scheme generators and the color schemes of real website and fashion images has changed over time.

1. Introduction

Color is an important element of most forms of visual design, from magazine covers, corporate logos and websites, to home interiors, products, and fashion. As a result, color features are often used to look for style trends in images [8, 23]. Analysis of color data in design has either focused on single colors [15] or model-free histogram representations [23], but a long tradition of work in color theory has shown that the relationships *between* colors have a significant effect on our perception [2, 19]. In this paper, inspired by these ideas from color harmony, we investigate probabilistic models for studying color schemes, separate from their component hues.

Color harmony, or the study of why certain combinations of colors look more pleasing together than others, has been a topic of active research for almost four hundred years [5]. Inspired by the success of formal models of musical harmony [2], color harmony researchers have sought to discover models which predict pleasing color schemes. These models range from older highly qualitative theories such as Itten's seven contrasts [9] to newer quantitative models like Matsuda's hue templates [14]. Template-based models are often used by artists and designers to help choose color schemes for their work, and have been applied in machine learning to classify images based on aesthetic quality [18] and to edit photos to improve their color [4]. Despite their widespread use, research shows that template-based color harmony is actually a poor model of human aesthetic preferences [19, 22], and artists and designers tend to use color templates for inspiration, then adjust them intuitively. Given their poor quality, we would like to empirically monitor the influence of these models.

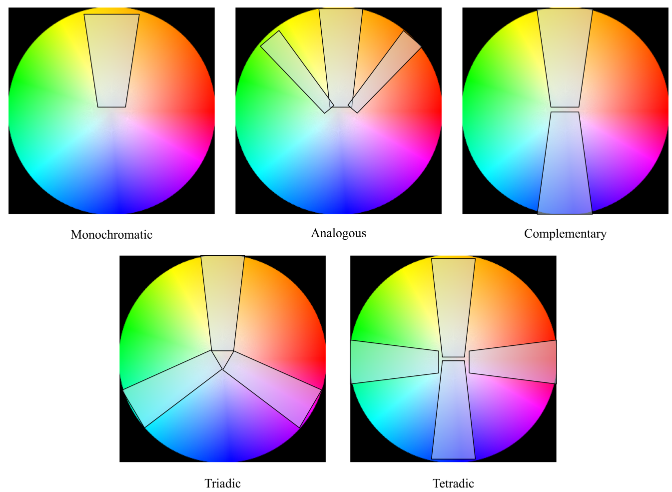

Figure 1. Five common color templates used by online color scheme generators. A color scheme from this template chooses colors within the shaded area, rotated to any center hue.

In this paper, we report on preliminary results of our ongoing work towards unsupervised generative modeling for color schemes. Note that by generative model, we do not mean a color scheme generation algorithm, which can produce new schemes, or a color harmony model, which distinguishes between good and bad schemes, but instead a probabilistic model that estimates the distribution of a collection of schemes, and measures the likelihood that a new scheme came from that distribution. This model produces a quantitative "measuring instrument" to answer questions such as:

- How likely is it that an online color scheme generator with an unknown underlying model would generate the color scheme of an image?
- To what extent does one set of images exhibit similar color-theoretic principles to another set of images?
- How has the popularity of template-based color schemes changed over time?
- Do different design media follow similar trends to-

wards or away from certain color principles?

Our approach involves fitting Gaussian Mixture Models (GMMs) to color schemes, and then using the resulting model to compute the likelihood that images were generated by this model. While conceptually simple, we show that this approach is nevertheless powerful enough to uncover trends in real-world forms of visual design. In particular, we apply our model on large-scale, time-referenced datasets of two very different types of visual artifacts — websites and fashion — and show how it uncovers clear temporal trends in both domains. To our knowledge, we are the first to present a practical technique for quantifying color combinations independent of hue.

2. Related Work

In the 20th century, two quantitative models of color harmony have been most influential: Matsuda's eight hue templates and ten lightness-chroma templates [14], and Moon and Spencer's geometric formulation of Itten's contrast principles [9, 17]. This work has inspired probabilistic models such as that of Lu *et al*. [13].

Since color is an essential element of fashion, many researchers have specifically addressed color. Matzen *et al*. investigate trends in street fashion photos and find patterns like a decrease in white after Labor Day in New York or an increase in red around Chinese New Year in China [15]. Liu *et al*. use color labels to inform their clothes parsing [11]. Vittayakorn *et al*. use color histogram features to learn a model of visual outfit similarity [23].

In recent years, color harmony models have inspired a number of online color scheme generators. Some generate harmonious color schemes stochastically [3, 20, 21] while others ask the user to choose a single color, then derive other harmonious colors (e.g., Adobe Color [1]) or allow users to create, share, and search for color schemes [16]. Of these examples, paletton [21] and Adobe Color both explicitly use template-based color harmony models. Five common templates are visualized in Figure 1.

Despite the elegance of these models and the convenience of online generators, studies of human color harmony preferences [19, 22] have shown that template-based models are not very accurate descriptions of human aesthetic preferences for color combinations: preferences vary between individuals and depend on the hue values (not just their relative positions in color-space) [19].

3. Methodology

To learn a probabilistic approximation of a template-based color generator from images, we need to extract the principal colors from the image, order all the colors, convert them to a representation invariant to lightness, chroma, and hue translations, as well as lightness and chroma dilations,

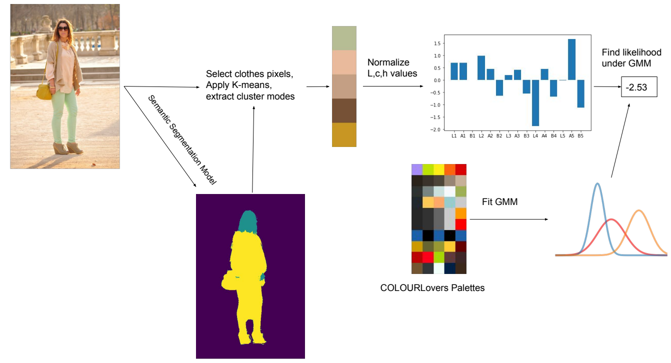

Figure 2. Our approach to compute the likelihood that a fashion color scheme came from a given color scheme dataset. We use the same process, without semantic segmentation, for website images.

and apply a probabilistic model to those features via maximum likelihood estimation. If only a subset of the original image is of interest, as with fashion images, we apply semantic segmentation first to isolate the pixels which correspond to clothing. We now describe each of these steps.

Extracting Color Schemes: To extract color schemes from images, we use weighted K-means clustering on the images pixels, where the weight for each pixel is equal to its chroma value and $k = 5$ for consistency with O'Donovan *et al*. [19] and the Paletton generator [21]. Despite the existence of better methods [6, 10], we found that the (chroma-weighted) mode of each cluster served as an effective representation of the color scheme of an image.

Normalization: In order to cluster together color schemes which were generated by the same color templates, we need a stable representation which captures the relative positions of each color, while being invariant to translation or dilation of lightness, chroma, or hue. To this end, we apply the following normalization procedure. Consider a color scheme $X = x_1, ..., x_k$ where each $x_i = l_i, a_i, b_i$ is a point in CIEL*a*b* colorspace, and colors are ordered from most to least frequent. We express the chroma $c_i = \sqrt{a_i^2 + b_i^2}$ and hue $h_i = \text{atan2}(a_i, b_i)$ (two argument inverse tangent). We compute the stable lightness, hue, and chroma,

$$l'_i = \frac{l_i - \bar{l}}{\sqrt{\frac{1}{k}\Sigma_j^k (l_j - \bar{l})^2}}, \quad c'_i = \frac{c_i}{\bar{c}}, \quad h'_i = (h_i - h_1) \mod 2\pi,$$

where $\bar{l}$ and $\bar{c}$ indicate the mean lightness and chroma, respectively, guaranteeing a chroma mean of 1, principal hue of 0, lightness mean of 0, and standard deviation of 1. Finally, we convert back to rectangular coordinates, to avoid the hue discontinuity around 0,

$$a'_i = c'_i \cos(h'_i), \quad b'_i = c'_i \sin(h'_i).$$

This approach has two important consequences. First, it ensures that the patterns we observe are not caused by general trends in hue, chroma, or lightness, such as a trend towards lighter colors or more reds. Second, it makes color templates (e.g., Figure 1) linearly separable. We find that when tested on synthetic data generated from five templates, a softmax regression model trained to distinguish between color templates improves from 39% to 69% accuracy.

Modeling: To capture the different kinds of color templates used in a dataset, we use a Gaussian Mixture Model (GMM) over the $3k$ dimensional normalized features. Note that the modal regions of the distribution correspond to common color patterns on the color wheel, not colors themselves, and thus high likelihood examples exhibit common color patterns and low likelihood examples do not. We use Expectation Maximization to fit GMMs with 10 components.

Semantic Segmentation: For fashion images, in order to reason about the colors of clothes pixels, rather than background or skin pixels, we employed a semantic segmentation model [12] trained on the CFPD dataset [11].[1] We reduced the labels in the dataset from 22 to 3 classes: (1) skin, face, hair, and sunglasses, (2) background, and (3) all remaining labels. Our analysis is only conducted on pixels from the third class. Our model achieved 93% accuracy on the CFPD test set for this greatly simplified clothes parsing problem.

Datasets: We fit our model to three color schemes datasets, and then observe color scheme trends in two diverse image datasets: website images and fashion images. Our color scheme datasets include our own synthetic template-based color schemes which use the five templates in 1 ($n = 10000$), the ColourLovers dataset ($n = 383938$) of human-uploaded color schemes from [19], and schemes we scraped from colormind.io ($n = 5000$), which generates schemes using a neural network [20]. Our dataset of website images ($n = 50232$) consists of screenshots from the Alexa top 500 US websites which we collected from 2004 to 2016 using the Internet Archive.[2] Our dataset of fashion images is the Street Fashion Styles dataset [7].

4. Results

After fitting our model to the three color scheme datasets, we can visualize the kinds of schemes each model prefers by examining the most likely and unlikely training examples. The likelihood scores for twenty four examples are visualized in Figure 3. We observe (subjectively) that all three models assign higher likelihood to template-like color schemes and lower likelihood to more scattered color schemes. The template model prefers highly saturated schemes, the ColorMind model prefers triadic color schemes, while the ColourLovers model prefers color gradients. All three assign very low likelihood to the single-color orange scheme at the bottom.

[1] We adapted code from https://github.com/minar09/Fashion-Clothing-Parsing

[2] https://archive.org

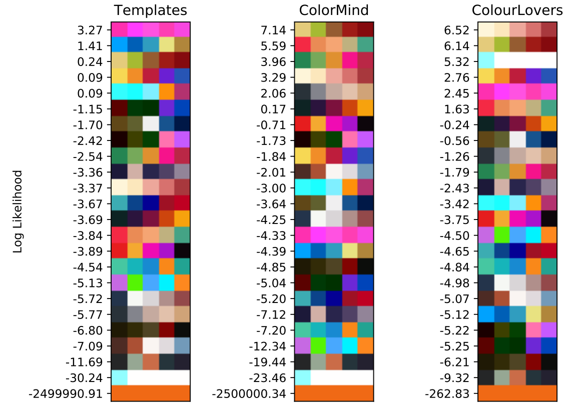

Figure 3. Twenty-four 5-color schemes, eight from each of the three color scheme datasets, ranked three ways according to their likelihood under a model trained on the entirety of one dataset.

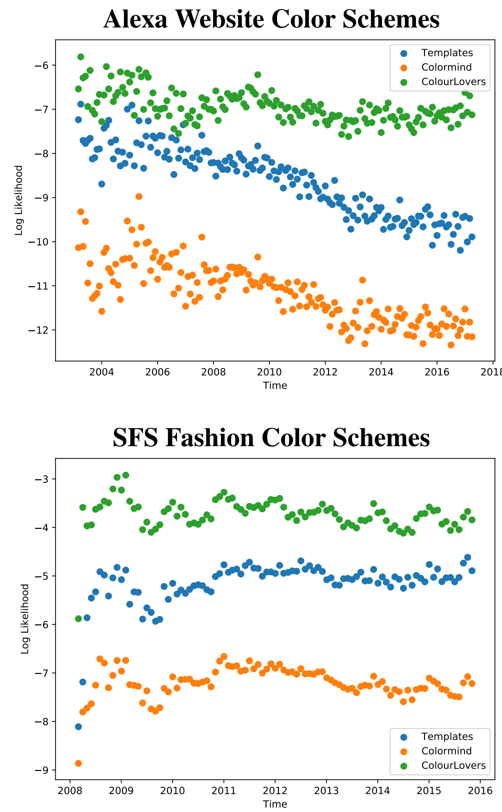

Figure 4. Average log likelihood for each month of images in each of our datasets, according to each of our GMMs. The SFS dataset has significantly lower sample size in 2008, causing outliers.

Our analysis of website and fashion images is shown in Figure 4. The website images have relatively high likelihood under all three models, but exhibit downward trends, meaning that the kinds of color schemes websites use are becoming less likely under color templates. The fashion images have higher likelihood across the board, do not exhibit broad trends, but do exhibit seasonality.

5. Discussion & Conclusion

We applied GMMs to learn a hue-relative generative model of color schemes from online color scheme websites and used them to measure the likelihood of real color schemes over time. We found four patterns:

1. Websites are shifting away from the template-based color schemes found from online generators.
2. Fashion color schemes exhibit little change over time but show a high degree of seasonality, where color schemes for winter fashion are more likely than for summer fashion.
3. The model based on human-created color schemes assigns higher likelihood to real color schemes from our datasets overall.
4. Both models assign higher likelihood to fashion color schemes than website color schemes.

Our analysis is preliminary and has some limitations. For example, there is a potential racial bias arising from skin/clothing parsing algorithms like the one we use, especially when color is the subject of research. While we did not notice significantly different accuracy based on race, we note that the CFPD dataset appears to over-represent American, European, and Asian women, and does not contain demographic labels to measure performance bias precisely.

When modeling images as mixtures of colors, Latent Dirichlet Allocation seems like a natural model, but initial experiments found that the color topics this method finds do not capture meaningful combinations of colors. There may be another mixed-membership model which can better capture the relationship between pixel values, colors, color schemes, and images.

There are numerous opportunities for future work with this approach to color. We were unsuccessful at finding satisfactory explanations for the trends we observed, future work is required to find more suitable examples and visualizations. We also do not address other design media, like interior or graphic design, which may also exhibit interesting color scheme trends as well. While we do not use our model to assess color harmony or propose new color schemes, such an application is possible and may be interesting to study.

Despite the limitations of modeling subjective aesthetic preferences, color harmony has a long and rich history of creative thought and influence on design. We provide a first look at probabilistic models for measuring that influence and apply those models to investigate how the likelihood of website and fashion images under online color scheme generators has changed over time. We hope our work inspires more nuanced future quantitative analysis of color combinations in design.

References

[1] Adobe color, 2019. https://color.adobe.com/create. 2

[2] J. Albers. *Interaction of color*. Yale Press, 1975. 1

[3] F. Bianchi. Coolors, 2014. https://coolors.co/. 2

[4] D. Cohen-Or, O. Sorkine, R. Gal, T. Leyvand, and Y.-Q. Xu. Color harmonization. In *ACM Transactions on Graphics (TOG)*, volume 25, pages 624–630. ACM, 2006. 1

[5] L. da Vinci. *A treatise of painting*. J. Taylor, 1721. 1

[6] J. Delon, A. Desolneux, J. L. Lisani, and A. B. Petro. Automatic color palette. In *ICIP*, 2005. 2

[7] X. Gu, Y. Wong, P. Peng, L. Shou, G. Chen, and M. S. Kankanhalli. Understanding fashion trends from street photos via neighbor-constrained embedding learning. In *ACM Multimedia*, pages 190–198, 2017. 3

[8] S. C. Hidayati, K.-L. Hua, W.-H. Cheng, and S.-W. Sun. What are the fashion trends in new york? In *ACM Multimedia*, pages 197–200, 2014. 1

[9] J. Itten. *The elements of color*. John Wiley & Sons, 1970. 1, 2

[10] S. Lin and P. Hanrahan. Modeling how people extract color themes from images. In *CHI*, pages 3101–3110, 2013. 2

[11] S. Liu, J. Feng, C. Domokos, H. Xu, J. Huang, Z. Hu, and S. Yan. Fashion parsing with weak color-category labels. *IEEE Trans. Multimedia*, 16(1):253–265, 2013. 2, 3

[12] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, pages 3431–3440, 2015. 3

[13] P. Lu, X. Peng, R. Li, and X. Wang. Towards aesthetics of image: A bayesian framework for color harmony modeling. *Signal Processing: Image Communication*, 39:487–498, 2015. 2

[14] Y. Matsuda. Color design. *Asakura Shoten*, 2(4):10, 1995. 1, 2

[15] K. Matzen, K. Bala, and N. Snavely. Streetstyle: Exploring world-wide clothing styles from millions of photos. *arXiv preprint arXiv:1706.01869*, 2017. 1, 2

[16] D. Monsef. Colourlovers, 2005. https://www.colourlovers.com/. 2

[17] P. Moon and D. Spencer. Geometric formulation of classical color harmony. *JOSA*, 34(1):46–59, 1944. 2

[18] M. Nishiyama, T. Okabe, I. Sato, and Y. Sato. Aesthetic quality classification of photographs based on color harmony. In *CVPR*, 2011. 1

[19] P. O'Donovan, A. Agarwala, and A. Hertzmann. Color compatibility from large datasets. In *ACM SIGGRAPH*, pages 1–12. 2011. 1, 2, 3

[20] J. Qiao. Colormind, 2017. http://colormind.io/. 2, 3

[21] P. Stanicek. Paletton, 2002. https://paletton.com/. 2

[22] F. Szabo, P. Bodrogi, and J. Schanda. Experimental modeling of colour harmony. *Color Research & Application*, 35(1):34–49, 2010. 1, 2

[23] S. Vittayakorn, K. Yamaguchi, A. C. Berg, and T. L. Berg. Runway to realway: Visual analysis of fashion. In *IEEE WCACV*, 2015. 1, 2