Pacific Graphics 2019
C. Theobalt, J. Lee, and G. Wetzstein
(Guest Editors)

Volume 38 (2019), Number 7

A Color-Pair Based Approach for Accurate Color Harmony Estimation

B. Yang[1], T. Wei[1], X. Fang[2], Z. Deng[3], F. W. B. Li[4], Y. Ling[1], and X. Wang[1]

[1]Zhejiang Gongshang University, China
[2]Anhui University, China
[3]University of Houston, USA
[4]University of Durham, United Kingdom

Abstract

Harmonious color combinations can stimulate positive user emotional responses. However, a widely open research question is: how can we establish a robust and accurate color harmony measure for the public and professional designers to identify the harmony level of a color theme or color set. Building upon the key discovery that color pairs play an important role in harmony estimation, in this paper we present a novel color-pair based estimation model to accurately measure the color harmony. It first takes a two-layer maximum likelihood estimation (MLE) based method to compute an initial prediction of color harmony by statistically modeling the pair-wise color preferences from existing datasets. Then, the initial scores are refined through a back-propagation neural network (BPNN) with a variety of color features extracted in different color spaces, so that an accurate harmony estimation can be obtained at the end. Our extensive experiments, including performance comparisons of harmony estimation applications, show the advantages of our method in comparison with the state of the art methods.

CCS Concepts

•***Computing methodologies*** → ***Perception;***

1. Introduction

Color harmony studies the visually harmonious color combination for arising a harmoniously emotional and environmental perception on a scene or an object. Obtaining an expected harmony of a color combination is critical for design; however, this task is challenging even for professionals. Therefore, designers often prefer seeking inspirations from different sources, such as artistic works, photographies, or color palette guides. There are some generally accepted aesthetic rules, e.g., complementary coloring and warm-versus-cool coloring. But these rules or guidelines need non-trivial manual involvements or judgments.

Previous harmony estimation studies [OL06, OCLM11, SBS10] use psychological analyses from a limited number of volunteers with limited color ranges and combinations, which can easily lead to an unstable model. Large database based learning works [LKPL14, OAH11, OAH14, GLS17] fall short of obtaining accurate results for real-world datasets [OAH11, COL18a, COL18b, KM16] in comparison with ground-truth user evaluations. Indeed, despite the above progresses have been made on quantifying the harmony level of a color set, this problem is far away from being resolved.

We performed statistical analysis on some publicly available, color harmony datasets, and found single colors based methods would be difficult to reliably estimate the color harmony of a color theme. On the other hand, ideally, we should collect large-scale user evaluations on all possible or at least a large number of N-color themes (N$\geq$3, where N is typically set to 5, since existing publicly available color harmony datasets are often based on 5-colors themes). Due to the combination issue, it is impractical to conduct and collect such large-scale user evaluations.

As a trade-off, in this work we utilize the concept of *color pairs* as the basis to effectively measure the color harmony of a color theme. Specifically, we propose a more accurate color harmony prediction model based on statistical inferences and learning from the color pairs. In our approach, the distribution of each color pair is utilized so that a two-layer maximum likelihood estimation (MLE) based method can be used to compute an initial harmony estimation. The initial estimation is then further refined with a set of selected color features in different color spaces through a back-propagation neural network (BPNN). To this end, an accurate color harmony prediction can be obtained. We performed comprehensive experiments as well as comparisons with existing methods to demonstrate the accuracy and robustness of our method.

The main contributions of this work include:

- To the best of our knowledge, our work is the first to report the usefulness of color pairs for harmony estimation.

- We introduce a novel color-pair based approach for an accurate color harmony estimation, where a two-layer MLE and BPNN are utilized to obtain a coarse-to-fine harmony estimation.

The remainder of this paper is organized as follows. The related work and the overview of our method are first introduced in Section 2 and Section 3, respectively. Then, the key observation on the color pair basis is discussed in Section 4. After that, the technical details of our method including the two-layer MLE based initial estimation and the BPNN based on refinement are presented in Section 5 and Section 6, respectively. Experimental results and applications are introduced in Section 7 and Section 8 separately, with the whole paper concluded in Section 9.

2. Related Work

In this section, we first review the development of color harmony theories and estimation, and then briefly review its recent applications.

2.1. Color Harmony

Color harmony theories can date back the early 20th century along with the development of the modern color theory [Ost69, Mun69]. Perhaps the most popular method is the color wheel which links color harmony to the relative color positions around a circle. Itten [Itt61] designed a 12-hue color wheel where two complementary colors are separated apart by 180°. Matsuda [Mat95] introduced an 8-hue color wheel consisting of different hue templates. These wheels and various extensions, such as N-color harmony, have been widely adopted for some fields, including color design [COSG*06, TMWW11, LZNH15] and color harmony theory [OAH11]. However, color wheels only consider color harmony without any quantification for harmony estimation. They may suggest ambiguous color pairs and make the combination of colors a tedious and non-trivial task.

Some studies investigated color ordering or relationships from various properties, such as hue, lightness, and chromaticity [Nem93, vG70, Che81, MS44b, MS44c, MS44a, Alb13]. Palmer [PS10] statistically analyzed the color preferences of volunteers and presented an ecological valence theory. However, the above studies are still subjective by nature and do not provide any quantitative ways to quantify color harmony, which is similar to the above color wheels based methods.

Numerical representations of color harmony have been studied physiologically in some controlled situations where color preferences of volunteers are collected to build harmony prediction models. For example, researchers invited volunteers to observe and score color pairs through physical experiments for possible factors affecting color harmony [OL06, OCLM11, SBS10]. However, their models are obtained with limited numbers of participants, color ranges and combinations. Consequently, their results can hardly be applicable in practice.

Statistical learning based approaches have also been proposed, including the hierarchical unsupervised learning model [LKPL14], regression [OAH11, LH13], collaborative filtering [OAH14], etc. Gramazio et al. [GLS17] proposed a five-color palette generation system, Colorgorical. It takes iteratively semi-random sampling to pick colors from the CIELAB space and customizes a valid palette based on the color preferences of different users. However, the preferred color palette generated with limited users may not be effective as harmonious colors. Kita and Miyata [KM16] evaluated any color combinations based on LASSO regression [OAH11]. However, this method can produce less harmonious colors than the five-color based method [OAH11]. Lu et al. [LPZL16] proposed a statistical learning framework for image harmony from a large number of natural images.

2.2. Applications of Color Harmony

Color harmony has been intensively applied for image recoloring and harmonization [LZNH15, COSG*06, TMWW11, LZNH15]. Recently, Tan et al. [TEG18] decomposed an image into a set of additive mixing layers, each of which corresponds to a color palette with varying weights. Through simultaneously recoloring the five-dimensional space RGBXY, it can be several orders higher of magnitude in efficiency than the work by Cohen et al. [COSG*06].

Chang et al. [CFL*15] allowed users to recolor photos by changing some palette colors interactively. Junho et al. [CYLC17] applied a deep learning based recoloring algorithm based on a target palette by optimizing the Euclidean and adversarial losses. Zhang et al. [ZXST17] and Mairéad [GDS17] took color decomposition and L_2 based cost function [CFL*15] for recoloring.

Color harmony has also been applied for dress-up. Yu et al. [YYTC12] developed a probabilistic framework based on the model of O'Donovan et al. [OAH11], which exploits a Bayesian network trained on example images of real-world outfits. In contrast, our method can suggest the appropriate colors for clothing by maximizing the color harmony degree.

3. Overview

In this section, we briefly describe the pipeline of our method and the datasets used for our analysis and learning.

Pipeline: Our color harmony estimation method for an input color theme consists of two main steps (Figure 1). The first step is a coarse estimation, which consists of two sub-steps: First substep is to estimate the harmony scores of all the color pairs that are enclosed in the input color theme while the subsequently second sub-step is to estimate the harmony score of the whole color theme, through a novel two-layers MLE scheme. For the second step, the outcomes from the first step, as well as supplemental features from different color spaces, are further fed into a novel BPNN model to obtain the final harmony score of the whole input color theme.

Datasets: Two publicly available datasets are used for our work: Kuler [COL18a] and MTurk [OAH11]. The Kuler dataset consists of 46,137 themes created by the users of Adobe Color CC, where each color theme is rated between 1 and 5 by at least two users. The MTurk dataset is based on Amazon Mechanical Turk, which includes finely selected 10,743 Kuler topics covering a variety of highly and poorly scored themes, with each graded by at least three random users.

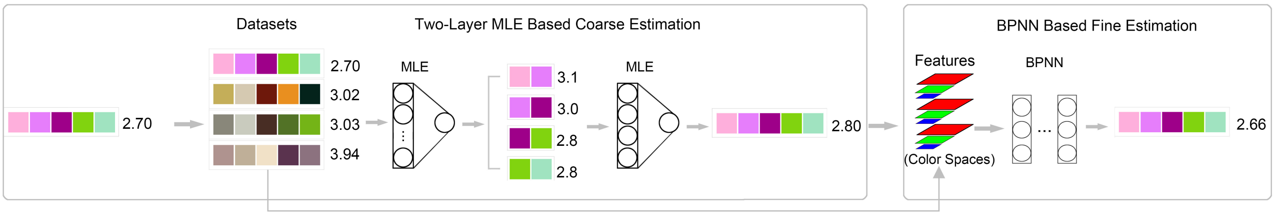


Figure 1: *The pipeline of our method. The number on the right side of each color set, including the five-color palettes and color pairs, represents the harmony score of the corresponding set.*

4. Color Pairs: a Novel Basis for Harmony Estimation

This section introduces the key finding of this work: color pairs can be an effective basis for color harmony estimation. Specifically, we first analyze the limitation of using single colors as the evaluation basis, and then discuss the new finding on the use of color pairs for harmony estimation.

4.1. Analysis of Single Colors Basis

In some scenarios we may prefer one color over others. However, it is unclear whether such a single colors based preference can be reliably used for estimating the color harmony of a color theme. To answer this question, we statistically look into the relationship between individual colors and the harmony scores.

In our work, assuming $S_i = (C_{i1}, C_{i2}, C_{i3}, C_{i4}, C_{i5})$ is one of the thousands of different color schemes and its obtained harmony value is h_i. Then, we further assume the harmony values of all the individual colors enclosed in S_i (that is, C_{i1}, C_{i2},..., C_{i5}) are h_i. We take the probability and average of all colors with the same hue value of H_c as the probability and average harmony value of color c (refer to Figure 2).

As shown in Figure 2a, the distributions of both Kuler and MTurk are very similar, peaking at warm colors (e.g., red, orange, green, and cyan). However, their harmony scores of the same colors are substantially different, as shown in Figure 2b. In this figure, the average scores of all the colors in the two datasets are at two different levels, despite they distribute almost uniformly. This experiment shows that the estimated harmony scores for the same color can be different from different data sources and systematically biased towards different ranges. Therefore, user preferences on individual colors cannot form a sufficient and robust basis to quantify the harmony of a color theme.

4.2. Analysis of Color Pairs Basis

We perform statistical analysis on color pairs, which are obtained by dividing each five-color theme into four neighboring matches (i.e., four color pairs). Figure 3 shows the percentages of the re-use times of color pairs in MTurk and Kuler. In other words, some color pairs are heavily re-used (that is, occurrences in different color schemes) while others may be seldom re-used. About 7% and 20.5% color pairs are used more than ten times in the MTurk and Kuler datasets, respectively. However, their color themes are not evenly distributed. Table 1 shows the distributions of color themes based on their harmony score intervals. Most of the color themes are scored in the middle (2-4), with even 99.46% in the 2-4 interval for the MTurk dataset.

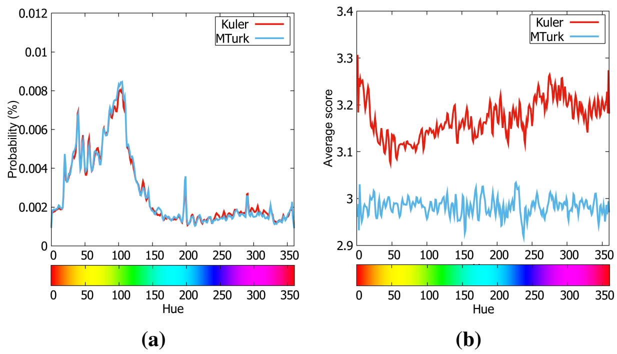


Figure 2: *Statistical analysis of the relationship between single color preferences and the color harmony estimation for different datasets. **(a)**: The color probability, and **(b)** the average harmony scores of all the colors. Note that the color is represented by hue.*

Table 1: *Ratios of themes in different score intervals for MTurk and Kuler*

Score Interval	1-2	2-3	3-4	4-5
MTurk	0.49%	49.39%	50.07%	0.05%
Kuler	6.28%	28.26%	46.03%	19.43%

We also validated the effectiveness of color pairs for the evaluation of theme harmony. Here, we visualize the number of the occurrences of color pairs in the Kuler dataset in Figure 4a. In this figure, each point represents a color pair, and a darker color denotes a larger number of its occurrences. Note that some areas are apparently denser than others, representing more occurrences of the corresponding pairs than others. From this figure, we can see users generally prefer pairs of similar types of colors (e.g., a pair of two warm colors, or a pair of two cold colors) than pairs of mixed colors (e.g., a pair composed of a warm color and a cold color). Such a distribution is consistent with previous studies [OL06, OCLM11] and conforms to the classical psychology based color harmony theory.

We also took a close look at two randomly selected example

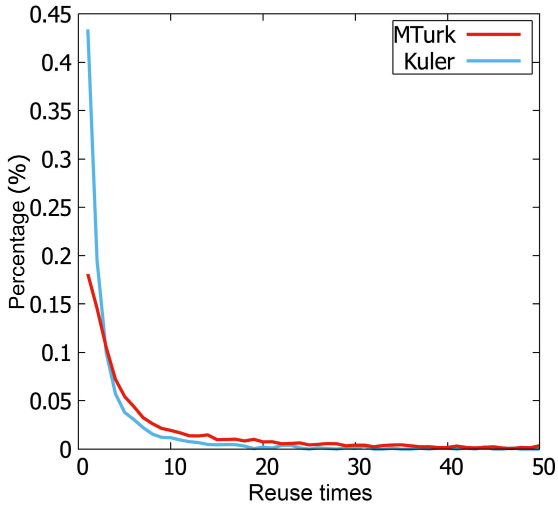
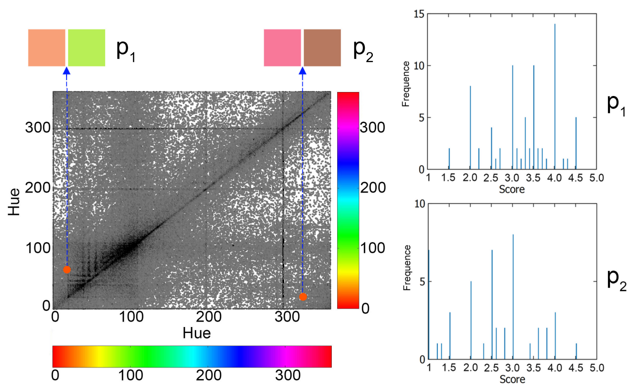


Figure 3: *The percentages of color-pairs under different reuse times*

.



(a) **(b)**

Figure 4: *Visualization of the reuses of color pairs in Kuler. (a): the visualization; and (b): the usage frequencies for the two example color pairs shown in (a).*

pairs, p_1 and p_2 (see Figure 4b). They are respectively selected from highly dense and relatively sparse areas in Figure 4a, e.g., p_1 has more occurrences than p_2. We can see that most of the scores of p_1 are bigger than 3 while most of the scores of p_2 largely is smaller than 3. This observation could lead to an interesting hypothesis: color pairs with higher occurrences are generally included into higher harmonious themes; and vice versa. Therefore, we hypothesize that color pair could be an effective basis to estimate the harmony of a color theme.

5. Two-Layer MLE based Coarse Harmony Estimation

In our two-layers MLE based coarse harmony estimation scheme, the first layer estimates the harmony score of each color pair according to its distribution, and the second layer further computes the harmony score of the whole color theme based on the obtained estimations of the color pairs. Our MLE-based method is based on the pair-wise probability distributions of the large-scale dataset, making the predicted harmony scores more accurate than traditional statistical methods.

The order of each color appearing in a color theme is important for its overall harmony, i. e., two themes consisting of the same set of five colors but different color orders could have possibly different harmony scores. Consequently, the order of the two colors in a color pair also matters when considering the color pair as a building block of a color theme: two colors with different orders represent different color pairs. Therefore, in our work, four consecutively neighboring pairs of each five-color theme in the dataset are taken for our experiments, where each pair keeps the same color order as in its original theme.

5.1. Two-layer MLE Estimation

In the first MLE layer, the harmony estimation of each color pair enclosed in a color theme is computed based on the statistics of all the pairs in all the color themes in the dataset.

Assume a color theme $C = (c_1, c_2, \ldots, c_N)$ consists of N different colors. Each of its M $(= N-1)$ color pairs can be denoted as $d_m | m \in \{1, 2, \ldots, M\}$ with its corresponding harmony value, $h_m | m \in \{1, 2, \ldots, M\}$. As the underlying assumption of this step, we take the known harmony score of the color scheme C as the sample value of each of its enclosed color pairs. Then, the MLE estimator of h_m, $\hat{h}_m$, can be formulated as:

$$\hat{h}_m = \arg\max_{h_m} \ln \mathcal{L}(h_m; S) \propto \arg\max_{h_m} \ln \sum_{i=1}^{L} p(s_i | h_m), \qquad (1)$$

where $S = \{s_1, s_2, \ldots, s_L\}$ denotes the combined set of the total L scores of d_m, and $\mathcal{L}()$ and p are the likelihood and probability functions, respectively.

h_m in Equation (1) is the unknown variable determined by the distribution of d_m, which can be estimated by the well-known Kolmogorov-Smirnov (K-S) test. As a non-parametric test, K-S compares the i.i.d. ordered observations of c_i with a reference probability, and consequently the K-S statistic for a given cumulative distribution function $P(c_i)$ can be modeled as:

$$\lim_{c_i \to \infty} D_e = \lim_{c_i \to \infty} \sup_{c_i} |P'(c_i) - P(c_i)|, \qquad (2)$$

where $P'(c_i)$ is the empirical distribution function of c_i. This statistic suggests that $P'(c_i)$ can be represented by $P(c_i)$ when D_e is very close to zero.

To find the distributions of each color pair, all possible distributions of $P'(c_i)$ are fitted with the data one by one according to Equation (2) under the confidence level $\alpha = 95\%$. The results show that about 80% data can be approximated by normal distributions, 15% data conform to the extreme value distributions, and the remaining 5% data can be fitted with other distributions, such as the gamma distribution and the Weibull distribution (Figure 5). Consequently, Equation (1) can be reformulated to compute the pair-wise harmony score once the distribution is confirmed.

The coarse harmony score of each color theme C in the second

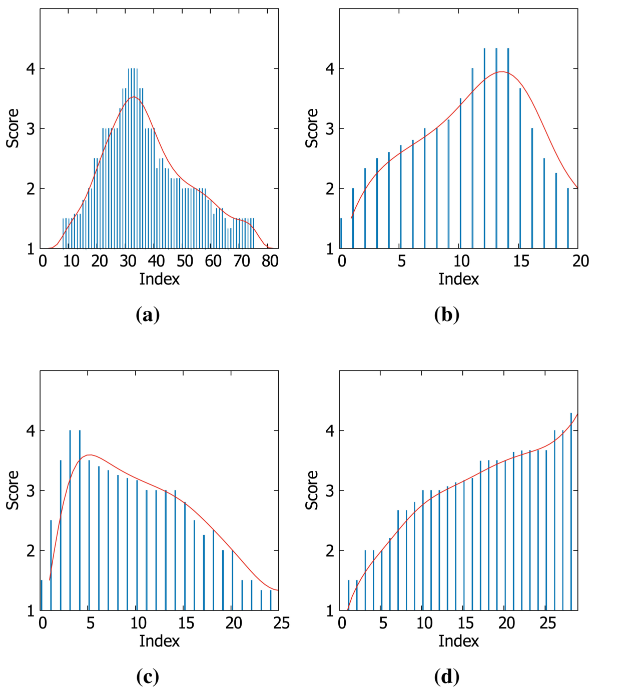


Figure 5: *The fitted distributions in K-S test according to the scores (S) of four example color pairs, including Gaussian (a), Extreme (b), Gamma (c) and Weibull (d) distributions. A number on the X-axis denotes the index of a color pair in the subset.*

layer of the MLE estimator, h_C, is estimated based on the pairwise harmony estimations from the first layer. This time we assume each color pair obeys the i.i.d. Gaussian distribution $G(\mu,\sigma)$. Therefore, h_C, can be estimated with MLE according to the following equation:

$$h_C = \arg\max_{\mu} \ln \prod_{i=1}^{M} G(h_i|\mu,\sigma) \propto \arg\min_{\mu} \ln \sum_{i=1}^{M} (h_i-\mu)^2. \quad (3)$$

5.2. Denoising Using Multivariate Linear Regression

To further improve the estimation accuracy, a multivariate linear regression is applied to eliminate noise, where the standard deviation of the hue channel is found to be effective for the weighting purpose. Assume $\sigma(C)$ is the standard color deviation of C. Accordingly, the denoised harmony $\hat{h}_C$ can be formulated as:

$$\hat{h}_C = \boldsymbol{w}^T \boldsymbol{u}_C + b, \quad (4)$$

where $\boldsymbol{u}_C = (\sigma(C), h_C)^T$ with $\boldsymbol{w}$ and b being the weights and the error.

Let's assume y_i is the ground truth score of one of the N unique themes in the dataset. Equation (4) can be estimated with the least squares method as:

$$E = \frac{1}{2}\sum_{i=1}^{N}(y_i - \boldsymbol{w}^T \boldsymbol{u}_C - b)^2, \quad (5)$$

which can be rewritten in the following form for a fast solution by taking the partial derivative of E w.r.t. $\boldsymbol{w'}$:

$$E \propto tr[(\boldsymbol{y} - \boldsymbol{x}\boldsymbol{w'})^T(\boldsymbol{y} - \boldsymbol{x}\boldsymbol{w'})]. \quad (6)$$

Then, $\hat{h}_C$ can be updated with Equation (4). Consequently, the denoised harmony estimation is obtained.

Table 2 shows the the prediction errors are noticeably reduced for both datasets after applying our multivariate regression. The above two-layer MLE based harmony estimation focuses only on the color (hue) part of the theme and thus may not be robust. We also need take more comprehensive features to obtain more accurate predictions, which is described below.

Table 2: *The changes of mean squared error (MSE) after the multivariate linear regression based denoising is applied.*

Dataset	Mturk	Kuler
MSE before regression	0.0635	0.4363
MSE after regression	0.0227	0.3410

6. BPNN based Harmony Refinement

Figure 6 shows the pipeline of the BPNN based harmony refinment, where additional color features from the color theme are extracted to augment the two-layer MLE based initial harmony estimation.

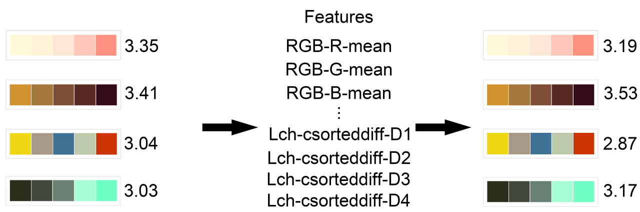
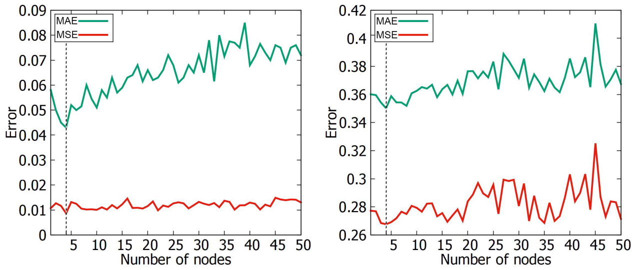


Figure 6: *Pipeline of the BPNN based harmony refinement, including feature extraction and BPNN prediction.*

6.1. Complementary Color Features

Here, some basic features [OL06, OAH11], such as lightness and chrominance, are included because of their relationship to color harmony, while insignificant features identified by previous studies are ignored to ensure the efficiency of the refinement.

Eight types of features are mainly computed as input to BPNN, including mean, standard deviation, maximum, minimum, range, median, mode, and color moment. They are computed from ten domains, including each channel of the three popular color spaces (RGB, Lab and HSV) and the chromaticity in the LCH color space, and consequently, there are 80 features. Additional two types of features are proposed to counter the illumination affection. One type of features are those computed with the Euclidean differences of adjacent colors in the theme for three popular color spaces (RGB, Lab and HSV). Then, there are additionally 24 features. The other type are the differences between adjacent colors for the ten domains, which adds 40 features to the feature list. These 40 features are also ordered decreasingly to improve the discrimination



(a) *MTurk* **(b)** *Kuler*

Figure 7: *Performances of different numbers of nodes in the hidden layer for MTurk and Kuler.*

of the feature vector, which adds another 40 features to the feature list. Finally, the normalized harmony estimation from the two-layer MLE is included as a feature. Therefore, totally, there are 185 features extracted for each theme, which form as a 185-element feature vector and input to BPNN for further harmony refinement.

6.2. Estimation with Back Propagation Neural Networks

BPNN is a multi-layer feedforward network trained with backward error propagation, which is capable of realizing any complex nonlinear mapping. It adopts gradient descent algorithm to gradually adjust weights and thresholds of the network by the backward error propagation for reducing errors. Predicted results can well approach the targets with minimum square errors if there exists proper design of network model structure and training samples.

In our work, the input layer of BPNN composes of the 185 neurons for the feature vector while the output layer is only one neuron for the prediction. We use one hidden layer to minimize the training time without scarifying accuracy. Deciding number of hidden layer nodes n_{id} is challenging because too few neurons cannot ensure the convergence while too many neurons may lead to excessive training time and even overfiting [Gur14, JMM96, HN92].

The optimal number of hidden layer nodes is empirically obtained through experiments. Figure 7 shows the effects of applying different number of hidden layer nodes on two datasets, measured by MSE and mean absolute error (MAE). MSE for MTurk (Figure 7a) are always at around 0.01. But its MAE gradually increases when increasing the hidden layer nodes and eventually leads to the overfitting and increase of the training time. This observation can also be got for Kuler (Figure 7b): Both MSE and MAE gradually increases when node number is larger than six with the minimal error appearing for four hidden layer nodes. Therefore, we take four as the optimal number of hidden layer nodes.

We demonstrate the benefits of using one hidden layer through experimental comparisons. Table 3 shows the experimental comparison of one hidden layer and three hidden layers. We select four hidden nodes in the one hidden layer, and the numbers of the hidden nodes in the three hidden layers are 4, 2 and 4, respectively. The comparison results show that the results are similar for the MTurk dataset, and the result by the three hidden layers in the Kuler dataset is worse than that by the one hidden layer. Therefore, we argue that adding more hidden layers may not substantially help the prediction accuracy of our model, besides inducing significantly more computational load.

Learning rate is also important, which determines the weight change in each training cycle. Bigger rates will worsen the prediction accuracy while smaller rates will cost longer running time. Experimentally, we set the learning rate $\eta = 0.1$ to balance the prediction accuracy and time consumption.

The optimal number of iterations is critical to obtain the best prediction results. Experimentally, we set it to 15000 and adopt the built-in validation checking function of MATLAB with the default parameters. This configuration makes the training process stop if the error curve does not decrease for six consecutive iterations. Additional training may not improve the prediction accuracy further and may even lead to over-fitting.

7. Experiment Results

In this section, we describe the experimental results by our approach. To train BPNN, we split the datasets into a training set and a test set by the ratio 4:1. We compared our prediction results with the LASSO regression based method by O'Donovan et al. [OAH11]. Figure 8 shows the comparison of the predicted harmony scores of some random color themes selected from the MTurk dataset by our method and [OAH11]. Obviously, the prediction results by our approach are consistently more closer to the ground-truth scores and thus more accurate than those by O'Donovan et al. [OAH11].

Note that in the work of [OAH11], O'Donovan et al. have already compared their method with some previous methods including those based on SVM, LS and KNN, and they demonstrated that the method in [OAH11] can only obtain slight improvements over the above mentioned methods. As described above, our experiment results in Figure 8 show that our method can substantially outperform the method in [OAH11]. Therefore, comparisons separately for each of those previous methods except [OAH11] were not taken.

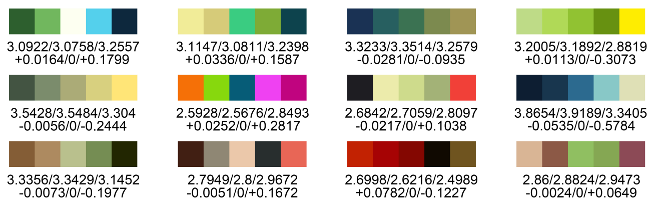


Figure 8: *Comparison of the color theme harmony prediction between O'Donovan et al. [OAH11] and our method. The three numbers in the first row under each color theme are the harmony scores of our method, of the ground truth, and of the results by O'Donovan et al. [OAH11], respectively, with the numbers in the second row being the corresponding difference between each method and the ground truth (i.e. the score of each method minus the score of the ground truth).*

In addition, two additional versions of our method were considered to show the advantages of our two-layer MLE and BPNN

Table 3: *Estimation accuracy comparison among one hidden layer and three hidden layers in two datasets.*

Method \ Dataset	One hidden layer (MTurk)	Three hidden layers (MTurk)	One hidden layer (Kuler)	Three hidden layers (Kuler)
MSE	0.0086	0.0084	0.2676	0.2683
MAE	0.0438	0.0462	0.3502	0.3541

based methodology: Our method without BPNN (MLE-only) , our method without MLE (BPNN-only), and only use one layer of MLE (One MLE layer). BPNN-only takes only the supplementary features to evaluate the harmony. Table 4 and Table 5 show the performance comparisons among different methods using MSE (Mean Squared Error) and MAE (Mean Absolute Error). Our prediction results with only BPNN or MLE can lead to large errors, while combing the two (i. e., the our method in Table 4 and Table 5) can perform significantly better, with the errors decreased even more than 60% for MTurk. This comparison shows that our method is more robust than the other four methods (O'Donovan et al., Kita and Miyata, SVM, and Auto-encoder) for MTurk and Kuler. Different from user-defined features, we use auto-encoder automatically extract some most representative features from color schemes and use the features for BPNN training. This conclusion can also be observed in Table 6, where Pearson correlation coefficient [OL06] (denoted as R) is used to evaluate the harmony prediction accuracy.

Validation User Study. We have conducted a user study to evaluate our work according to users' intuitive visual perception. We invited 30 participants (15 men and 15 women; ages from 20 to 25) to evaluate 20 sets of color themes. In our study, we obtained two different ratings for each set of the color themes: the first one was predicted by our computational model (denoted as OURS in Figure 9), while the second one was obtained by using a five-point (1-5) rating system to describe the participants' perceptions on the color theme (denoted as USERS in Figure 9). We depict the results of the user study by averaging the obtained ratings. As shown in Figure 9, the results predicted by our framework are reasonably comparable with those obtained by the participants.

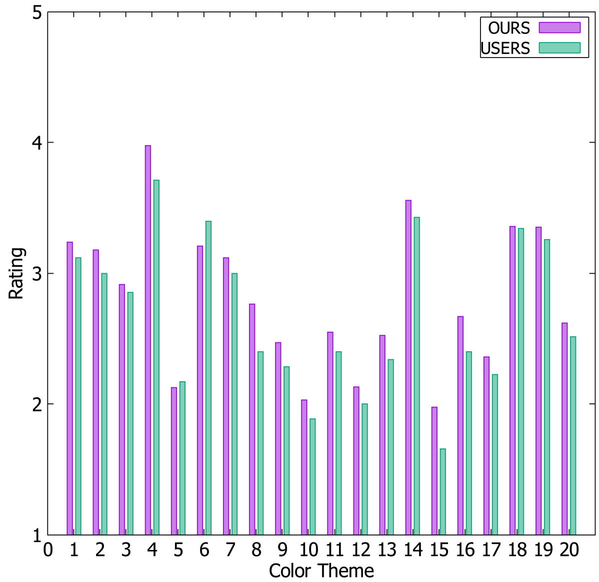


Figure 9: *Comparison of Outputs from Our model and Users.*

Figure 10: *Comparison of the harmony estimation with extracted color themes between O'Donovan et al. [OAH11] and our method. The numbers under each theme are listed in the same ways as Figure 8.*

8. Selected Applications

We also demonstrate how to apply our method to a few selected applications. Specifically, the state of the art methods and ours are trained separately with the MTurk dataset for comparison. We also employed the work of [LH13] to extract the color theme of an image for the applications described below.

Harmony Estimation. First, we applied harmony estimation to evaluate the extracted color themes. Here, each image is segmented to compute the color theme by extracting six different types of 79 features using LASSO regression. Figure 10 shows example results of color theme extraction from images of different styles. The estimations by Our method are consistently more accurate than the work of [OAH11].

Color Suggestion. Harmony estimation can also be applied for color suggestion so that suitable colors can be recommended to replace existing colors for a harmonious effect. Figure 11 shows the experimental results with the source palettes created by the Colorgorical system [GLS17]. The source palettes have different styles. For example, the third one is warm while the fifth is cold. For simplicity, we replace only one color among the five colors for suggesting a better harmonious theme, i. e., the new theme with the new color will have the max harmony value among all permutations of its five colors. 16 equal-distanced colors from RGB color space are included as the possible colors for suggestion. It can be seen that the new palettes with our suggested colors are more harmonious with higher scores than those by Kita et al. [KM16] .

Recoloring Images. We also experimented with recoloring images with the suggested colors. In this experiment, two colors among the five theme colors of each image are suggested for a better harmony effect. The best and worst harmonious themes among

Table 4: *Color harmony estimation performances of different methods in MSE.*

Method / Dataset	MLE-only	BPNN-only	One MLE layer	Our method	O'Donovan et al. [OAH11]	Kita and Miyata [KM16]	SVM	Auto-encoder
MTurk	0.0227	0.0222	0.0124	0.0086	0.0518	0.0568	0.0304	0.0542
Kuler	0.3410	0.4034	0.3078	0.2676	0.4456	0.4884	0.4525	0.4372

Table 5: *Color harmony estimation performances of different methods in MAE.*

Method / Dataset	MLE-only	BPNN-only	One MLE layer	Our method	O'Donovan et al. [OAH11]	Kita and Miyata [KM16]	SVM	Auto-encoder
MTurk	0.2435	0.1144	0.0775	0.0438	0.4456	0.4639	0.0974	0.1003
Kuler	0.5656	0.4882	0.3925	0.3502	0.5191	0.5393	0.4707	0.5039

Table 6: *Estimation accuracy comparison among different methods in Pearson correlation coefficient (%).*

Method / Dataset	MLE-only	BPNN-only	Our method	O'Donovan et al. [OAH11]	Kita and Miyata [KM16]
MTurk	78.69	80.12	96.03	73.61	68.35
Kuler	45.87	61.38	70.02	37.65	34.96

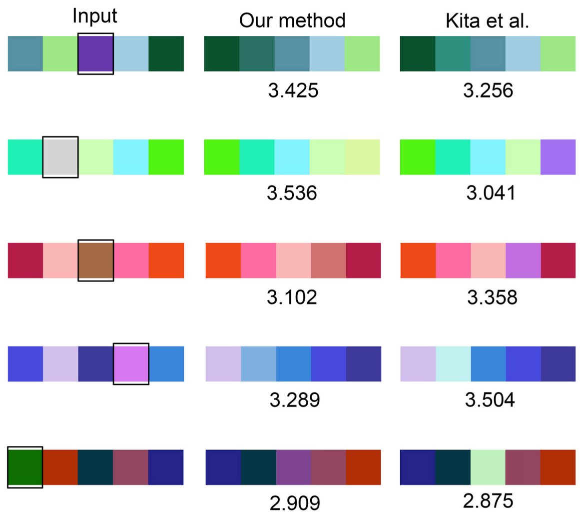


Figure 11: *Color suggestions for five different palettes. Each row shows the original palette, and harmonized palettes by our method and Kita et al. [KM16] respectively. The color in a black rectangle is the color to be replaced.*

top 100 suggested color themes obtained by both O'Donovan et al. [OAH11] and our method are selected for comparison. Figure 12 shows the results for images with different styles: simple graphics, natural scenes, buildings, and home scenes. Each image was recolored based on the colors of a given color theme using the algorithm by Lin et al. [LWL*18]. Our method generally outperformed [OAH11] with higher scores and visually more harmonious effects for both the best and worst cases. Especially, for the inharmonious natural scene shown in the second row of Figure 12, our method produced more harmonious results than O'Donovan et al. [OAH11], because ours selected colors with higher brightness and saturation than theirs.

Virtual Shopping. We can also apply harmony estimation to virtual shopping for a better color combination of a cloth set. As shown in Figure 13. Our method can produce more harmonious costume colors than O'Donovan et al. [OAH11] and thus improve the costume appearance. This could be a useful feature for virtual or online shopping applications.

9. Discussion and Conclusion

In this paper, we present a new color harmony estimation method by discovering the importance of color pairs in harmony evaluation. Specifically, we introduce a two-layers MLE based method by utilizing the statistical distribution of color pairs, so that a robust initial harmony estimation of a color theme can be obtained. Then, the initial estimation and other selected color features are further inputted into a BPNN to produce the final harmony value. Our experiment results show that better harmony estimations and application results can be obtained by our method than state of the art methods.

Our current method has some limitations. First, the BPNN in our current methodology is relatively simplified; therefore, its performance can be limited when dealing with very large training datasets. For this case, the widely studied deep neural networks can be a potential model to consider. Second, our current model is designed for the harmony evaluation of five-color themes. However, our model could be extended for themes with more than five colors. In this case, simply extracting more features or reshaping the two-layer MLE and BPNN to fit with the bigger themes may not work well. We need find a more creative way to cope with this type of themes in the future.

10. Acknowledgments

This work was partly supported by a research grant from the National key R&D Program of china (Grant No:2018YFB 1403200). Zhigang Deng is in part supported by US NSF IIS-1524782.

References

[Alb13] ALBERS J.: *Interaction of color*. Yale University Press, 2013. 2

Figure 12: *Color suggestions for image examples with four different styles (from top to bottom): simple graphics, natural scenes, buildings, and home scenes. Each row shows the original image (leftmost) and the most harmonious and the most inharmonious results from both O'Donovan et al. [OAH11] and our method. Note that (1) the original images came from COLORLovers and colorpalettes [pi18], and (2) the color to be replaced is shown in the same way as Figure 11.*

[CFL*15] Chang H., Fried O., Liu Y., DiVerdi S., Finkelstein A.: Palette-based photo recoloring. *ACM Transactions on Graphics 34*, 4 (2015), 139. 2

[Che81] Chevreul M.: The principles of harmony and contrast of colors. translation by c. martel (1854) from the french edition of 1839. reprinted, 1981. 2

[COL18a] COLOR A.:, Accessed JUN, 2018. URL: https://color.adobe.com/. 1, 2

[COL18b] COLORLOVERS:. Accessed JUN, 2018. URL: http://www.colourlovers.com/. 1

[COSG*06] Cohen-Or D., Sorkine O., Gal R., Leyvand T., Xu Y.-Q.: Color harmonization. *ACM Transactions on Graphics 25*, 3 (2006), 624–630. 2

[CYLC17] Cho J., Yun S., Lee K., Choi J. Y.: PaletteNet: Image

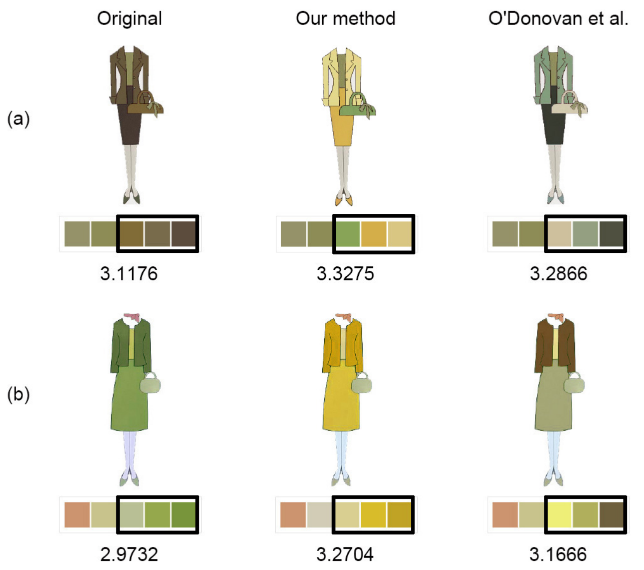


Figure 13: *Comparison of costume coloring suggestion between our method and O'Donovan et al. [OAH11] The color to be replaced is shown in the same way as Figure 11.*

recolorization with given color palette. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2017), pp. 1058–1066. 2

[GDS17] GROGAN M., DAHYOT R., SMOLIC A.: User interaction for image recolouring using L_2. In *European Conference on Visual Media* (2017), pp. 6:1–6:10. 2

[GLS17] GRAMAZIO C. C., LAIDLAW D. H., SCHLOSS K. B.: Colorgorical: Creating discriminable and preferable color palettes for information visualization. *IEEE Transactions on Visualization and Computer Graphics 23*, 1 (2017), 521–530. 1, 2, 7

[Gur14] GURNEY K.: *An introduction to neural networks*. CRC press, 2014. 6

[HN92] HECHT-NIELSEN R.: Theory of the backpropagation neural network. In *Neural networks for perception*. Elsevier, 1992, pp. 65–93. 6

[Itt61] ITTEN J.: The art of color, 1961. 2

[JMM96] JAIN A. K., MAO J., MOHIUDDIN K.: Artificial neural networks: A tutorial. *Computer*, 3 (1996), 31–44. 6

[KM16] KITA N., MIYATA K.: Aesthetic rating and color suggestion for color palettes. In *Pacific Conference on Computer Graphics and Applications* (2016), pp. 127–136. 1, 2, 7, 8

[LH13] LIN S., HANRAHAN P.: Modeling how people extract color themes from images. In *SIGCHI Conference on Human Factors in Computing Systems* (2013), pp. 3101–3110. 2, 7

[LKPL14] LU P., KUANG Z., PENG X., LI R.: Discovering harmony: A hierarchical colour harmony model for aesthetics assessment. In *Asian Conference on Computer Vision* (2014), pp. 452–467. 1, 2

[LPZL16] LU P., PENG X., ZHU X., LI R.: An EL-LDA based general color harmony model for photo aesthetics assessment. *Signal Processing 120* (2016), 731–745. 2

[LWL*18] LIN X., WANG X., LI F. W. B., YANG B., ZHANG K., WEI T.: Image recoloring for home scene. In *Proceedings of the 16th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry* (New York, NY, USA, 2018), VRCAI '18, ACM, pp. 29:1–29:9. 8

[LZNH15] LI X., ZHAO H., NIE G., HUANG H.: Image recoloring using geodesic distance based color harmonization. *Computational Visual Media 1*, 2 (2015), 143–155. 2

[Mat95] MATSUDA Y.: Color design. *Asakura Shoten 2*, 4 (1995), 10. 2

[MS44a] MOON P., SPENCER D. E.: Aesthetic measure applied to color harmony. *Journal of the Optical Society of America 34*, 4 (1944), 234–242. 2

[MS44b] MOON P., SPENCER D. E.: Area in color harmony. *Journal of the Optical Society of America 34*, 2 (1944), 93–103. 2

[MS44c] MOON P., SPENCER D. E.: Geometric formulation of classical color harmony. *Journal of the Optical Society of America 34*, 1 (1944), 46–59. 2

[Mun69] MUNSELL A. H.: *A grammar of color*. Van Nostrand Reinhold, 1969. 2

[Nem93] NEMCSICS A.: *Colour dynamics: Environmental colour design*. Hong Kong Polytech. Univ., 1993. 2

[OAH11] O'DONOVAN P., AGARWALA A., HERTZMANN A.: Color compatibility from large datasets. *ACM Transactions on Graphics 30*, 4 (2011), 63. 1, 2, 5, 6, 7, 8, 9, 10

[OAH14] O'DONOVAN P., AGARWALA A., HERTZMANN A.: Collaborative filtering of color aesthetics. In *Workshop on Computational Aesthetics* (2014), pp. 33–40. 1, 2

[OCLM11] OU L.-C., CHONG P., LUO M. R., MINCHEW C.: Additivity of colour harmony. *Color Research & Application 36*, 5 (2011), 355–372. 1, 2, 3

[OL06] OU L.-C., LUO M. R.: A colour harmony model for two-colour combinations. *Color Research & Application 31*, 3 (2006), 191–204. 1, 2, 3, 5, 7

[Ost69] OSTWALD W.: The colour primer, 1969. 2

[pi18] PALETTE IDEAS C.:, Accessed JUN, 2018. URL: `http://colorpalettes.net/`. 9

[PS10] PALMER S. E., SCHLOSS K. B.: An ecological valence theory of human color preference. *National Academy of Sciences 107*, 19 (2010), 8877–8882. 2

[SBS10] SZABÓ F., BODROGI P., SCHANDA J.: Experimental modeling of colour harmony. *Color Research & Application 35*, 1 (2010), 34–49. 1, 2

[TEG18] TAN J., ECHEVARRIA J., GINGOLD Y.: Palette-based image decomposition, harmonization, and color transfer. *arXiv preprint arXiv:1804.01225* (2018). 2

[TMWW11] TANG Z., MIAO Z., WAN Y., WANG Z.: Color harmonization for images. *Journal of Electronic Imaging 20*, 2 (2011), 023001. 2

[vG70] VON GOETHE J. W.: Theory of colours, 1970. 2

[YYTC12] YU L.-F., YEUNG S.-K., TERZOPOULOS D., CHAN T. F.: DressUp!: Outfit synthesis through automatic optimization. *ACM Transactions on Graphics 31*, 6 (2012), 134. 2

[ZXST17] ZHANG Q., XIAO C., SUN H., TANG F.: Palette-based image recoloring using color decomposition optimization. *IEEE Transactions on Image Processing 26*, 4 (2017), 1952–1964. 2