CHAPTER

1 Differential color harmony: A robust approach for extracting harmonic color features and perceiving aesthetics in a large image dataset

Tousif Osman, Shahreen Shahjahan Psyche, Tonmoay Deb, Adnan Firoze, Rashedur M. Rahman

Department of Electrical and Computer Engineering, North South University, Dhaka, Bangladesh

1.1 Introduction

Understanding beauty and aesthetics has always been of interest for humans, from the very beginnings of human history. Many artists through the ages have invented their own styles of capturing beauty and have produced artwork that has dazzled, and continues to dazzle, mankind to this day. One of the difficulties of working in this domain is that there is no exact measurement of beauty as humans perceive it. Many artists have formulated their own idea of aesthetics, but none of them are absolute nor do they involve an exact scale of measurement. Therefore, most regular mathematics and computational techniques fail in this domain.

The idea of color harmony and its correlation with aesthetics is not new. One of the oldest and most notable works in this area [1], which had a great impact, was by Moon and Spencer in 1944. Many ideas and theories came after that work and today this field is quite enriched. The concept of color harmony revolves around the idea that colors maintaining a certain relation with their neighboring colors are perceived to be more appealing to the human eye [2]. This concept is very similar to the concept of musical notation, and is one of the few laws of art that does not vary from person to person. Currently, one of the most commonly used basic ideas for extracting harmonic color features (HCFs) is clustering and segmenting the colors. However, this is a resource-hungry, soft-computational approach.

In our research, we have used mathematical and computational approaches to provide a solution to this problem. We have used neural networks (NNs) and regression models to verify our extracted features and have demonstrated their significance

Systems Simulation and Modeling for Cloud Computing and Big Data Applications
https://doi.org/10.1016/B978-0-12-819779-0.00001-0

in real-world situations. Even with simpler regression models, i.e., linear regression, the results are still promising. The procedures and workflow of our algorithm have been described chronologically in the following sections.

1.2 Related works

We are not the only ones who have considered color harmony in measuring the aesthetic beauty of an image. Other researchers have walked this path before us. Lu et al. [3] have designed a statistical learning structure to train a color harmony model. They used a dataset consisting of a large number of natural images. They made Dirichlet allocation training (LDA) smoother by using the content of the images along with the visual features. Then, the harmonic color level was estimated based on the supervised/unsupervised model(s) that indicate the photo aesthetic scores. In another paper, Lu et al. [4] trained an LDA-based color harmony model that considers harmonic colors as using spatial distances.

Datta et al. [5] have developed a system that automatically infers the aesthetic beauty of an image using the visual content. They crowd sourced the ratings of images. They employed classifiers using machine-learning techniques, such as support vector machine, classification trees, and linear regression, to predict a numerical aesthetic score.

Phan et al. [6] have built statistical models that are even used in building some practical applications. These statistical models have been trained based on the coloring style of a fine arts collection. The authors have also used density estimation to determine the features of palette data, since artists usually have their own personal coloring patterns in their creations, which result in the frequent appearance of certain color palettes in multiple paintings of the same artist.

Amati et al. [7] have shown that there is no explicit linear dependence between colorfulness and aesthetics; rather, correlations arise categorically for different images: for example, “landscape,” “abstract.” As a dataset, they have compiled perceptual data from a large-scale user study.

Lu et al. [8] have proposed a color harmony model based on a Bayesian framework. Classical artists had already paved way using this technique unconsciously, without adhering to any established framework. On the other hand, learning-based models that detect the underlying patterns of a work are shown to be a possibility. In the LDA learning process, other two-color harmony models have been used during the training of the model.

Nishiyama et al. [9] claim that the existing harmonic color models consider simple coloring patterns that fail to assess photographs with complicated color patterns. To resolve this challenging problem, they have built a method in which they consider a photograph consisting of a cluster of confined regions with variation in colors that are simpler, which has ultimately led them towards developing a method to assess the aesthetics of a photo based on its harmonious colors. They have also improved the classification performance by integrating blur, edges, and saliency features.

1.3 Methodology

The basic principle of our research stands on the following idea: colors seem to be harmonized to the human eye if the color shades (Fig. 1.1A) or the base color changes (Fig. 1.1B) gradually maintain a pattern [2]. Fig. 1.1 shows three combinations of colors: (A) gradual change of color shade, (B) gradual change of color base and shade, and (C) no pattern in the change of colors. Looking at these three images, even a nonartist will find Fig. 1.1A and B more appealing than Fig. 1.1C. By nature, the human eye tends to find harmonic color "melodies" similar to music.

Commonly, artists and designers shift colors from RGB (red, green, blue) color scale to HSV/HSL (hue, saturation, value/hue, saturation, lightness). For our research, we selected the HSV scale. The RGB color scale states that every color can be represented with the three base color components of red, green and blue. In the case of HSV, in simple terms, the hue is the key component that defines the base of a color, saturation defines the whiteness of a color, and value specifies the blackness of a color. This color scale represents all possible colors in terms of these three components.

According to designer and artist guidelines, a small change in the HSV/HSL color component produces a color that maintains harmony with the original color. We have used these facts as the base hypothesis of our research. Fig. 1.2 shows a three-dimensional representation of RGB (Fig. 1.2A) and HSV (Fig. 1.2B) color space, respectively.

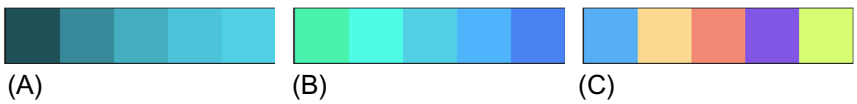


FIG. 1.1

Scale of differential colors. (A) Harmonious shade, (B) harmonious hue, (C) random color combination.

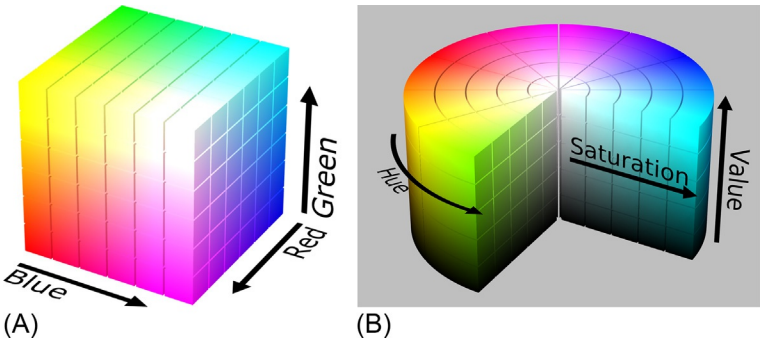


FIG. 1.2

Three-dimensional representation of color spaces. (A) RGB [10], (B) HSV [10].

After shifting space, we then calculated the gradient of the shifted color space. This gradient tells us about the rate of change in color components. Next, we applied differential operators to produce a smaller feature set that tells about the overall color correlation in an image. Finally, we applied a machine-learning model. Details of the feature extraction process and model design are explained in the following subsections.

1.3.1 Extracting differential harmonic color features

This is the initial and most crucial stage of our workflow and each step has been ordered based on performance and results (of the previous step). Shifting color spaces from RGB to HSV is the very first step of our process. We are shifting the color space so that a linear change in color component represents a change in color shade or base. This is also a common process in much color analysis research [3, 5, 8], to make the colors more aligned with how humans perceive colors. Fig. 1.3A is a sample image and Fig. 1.3B–D are its three RGB color components, respectively. Fig. 1.3E–G are the components of shifted HSV components.

Next, to produce a fixed-length feature vector, we have transformed the variable size images to a 128-pixel by 128-pixel square image using interpolation. Our dataset (described in more detail in Section 1.3.3) keeps the images in the range of 128×128, but they are not square images: an image can be 128×100 or of some other dimensions. For regular images, the size can vary in a very large domain. Therefore, if we do not change the size to a fixed-size image, the final features will not have a fixed length. We have used bicubic interpolation [11]. Fig. 1.4 shows the transformation of the sample image.

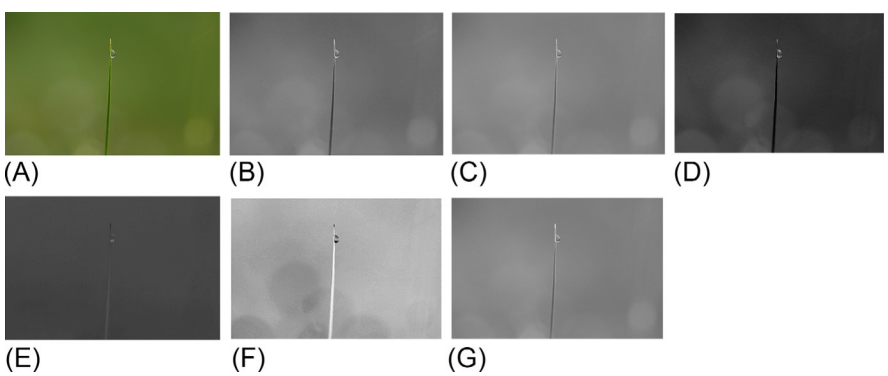


FIG. 1.3

One input image and components. (A) Input image, (B) red channel, (C) green channel, (D) blue channel, (E) hue channel in HSV, (F) saturation channel in HSV, and (G) value channel of HSV.

FIG. 1.4

Resized/resampled image version of 128 × 128 pixels from Fig. 1.3A.

The change of color component rate is formulated using a gradient image by computing the average distance of each pixel with its surrounding eight pixels. Eq. (1.1) has been applied to each pixel of the components to produce the gradient images:

$$g(x,y) = \sum_{i,j=-1,-1}^{2,2\, i,j\neq 0,0} \frac{cmp(x,y) - cmp(x+i, y+j)}{8} \tag{1.1}$$

Here x and y are the iterators of each component and *cmp* is the variable representing the component matrix. Fig. 1.5 shows the gradients of each component. Fig. 1.5A illustrates the gradient image and Fig. 1.5B–D illustrates the HSV gradient components, respectively.

Following is the pseudocode for producing gradient images:

```
For p_ij in Image Component:
  For i from −1 to 2
    For j from −1 to 2
      If i, j is not (0,0)
        avg_dist := avg_dist + (comp[p_ij] − comp[p_i + i, p_j + j])/8
    End if
   End for
  End for
   gra_comp[p_ij] := avg_dist
End for
```

Here, p_{ij} is an iterator on the color component and the *ij* suffix represents the column and row iterator variable. The *gra_comp* is the final gradient of the component for the given component. Resulting gradients hold the information about color change, but

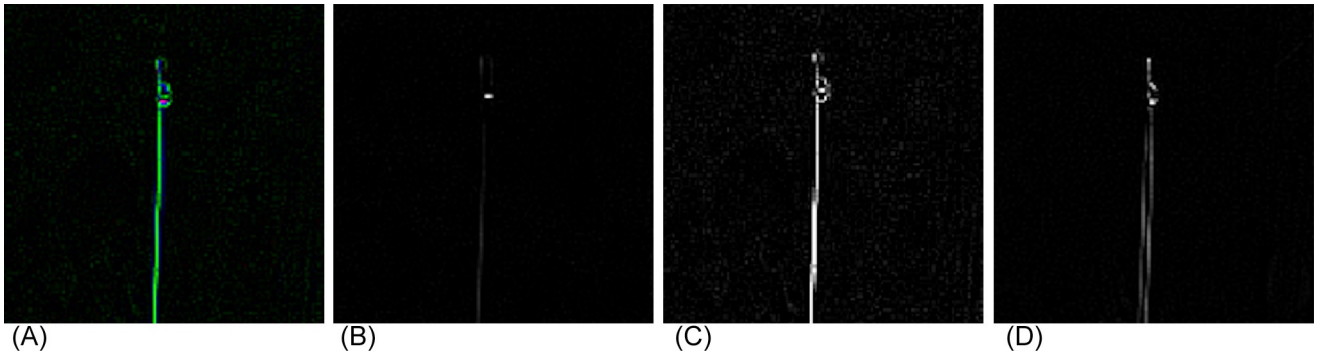


FIG. 1.5

Gradients of each color component. (A) Combined gradient, (B) hue gradient, (C) saturation gradient, (D) value gradient.

due to their higher dimensions they are problematic as features. The feature vector has a length of $3 \times 16{,}384$ and moreover we have 5000 sample images. Regular learning algorithms will fail to map the features and will require a long time to process without dimensionality reduction. We could apply principal component analysis (PCA) directly on the entire dataset, but by applying derivative and min-max normalization, we are extracting a robust feature set that has the information about the change of color in an image without much loss and can be used in any algorithm directly. However, by doing differentiation, we are losing the detail information, but we are only interested in the rate of change and continued differentiation is preserving it without much loss.

From our research results, we have concluded that applying discrete differentiation on the gradients holds the color correlation for the entire image from a lower dimension. Eq. (1.2) is the discrete differentiation operator we have used to reduce the dimensions. X_1, X_2, X_3, X_{m-1}, X_m are the elements of the feature vector, where X_1 is the first element and X_m is the last element.

$$\Delta y = [X_2 - X_1, X_3 - X_2, \ldots X_m - X_{m-1}] \tag{1.2}$$

This operator reduces one dimension of a series. We have calculated lower dimensional 5×5 gradient matrices by calculating 123rd derivatives and subsequently calculating 123rd derivatives on the transpose of the previously calculated matrix.

$$z_i = \frac{x_{i-\min(x)}}{\max(x) - \min(x)} \tag{1.3}$$

We have applied min-max normalization to normalize the results in Eq. (1.2) (using Eq. 1.3). Fig. 1.6A–C represent differential matrices derived from each component.

Finally, we have changed the shape 5×5 2D matrix to a 1×25 1D matrix. These final matrices are our calculated HCF vectors, which we used to train models in the following sections. Fig. 1.7A–C represents the nonnormalized feature scores. If we observe the original sample image (Fig. 1.4), we can see the image has a base color of green, where the greatest part of the image is composed of shades of green (light gray in print version), which implies the rate of change of hue is less, whereas the saturation and value have a higher rate of change. Our calculated HCF also complies with the expected result: hue has a more stable rate of change than the other two. Fig. 1.8A

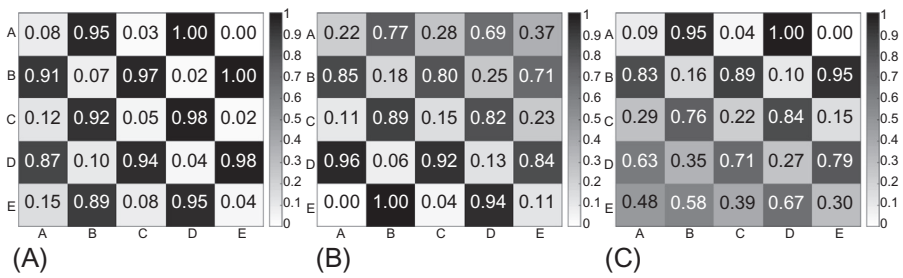


FIG. 1.6

Differential features. (A) Hue, (B) saturation, and (C) value component.

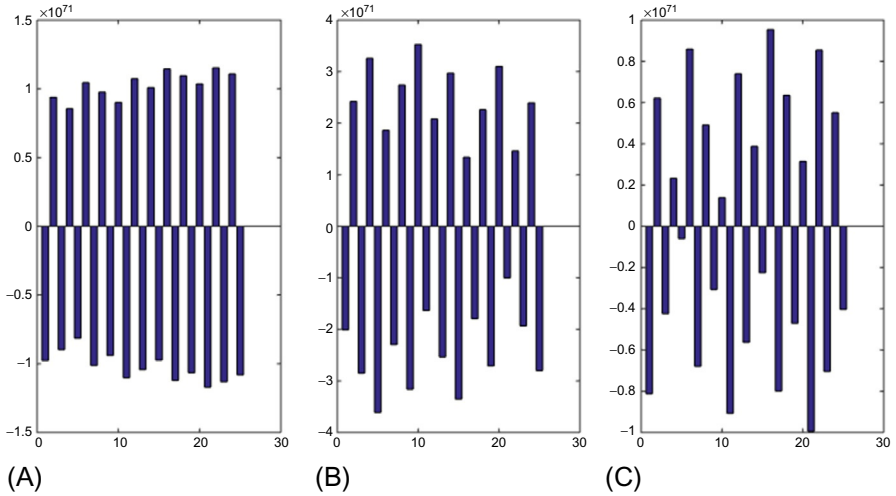


FIG. 1.7

Harmonic color features (HCF) plots: (A) hue, (B) saturation, (C) value.

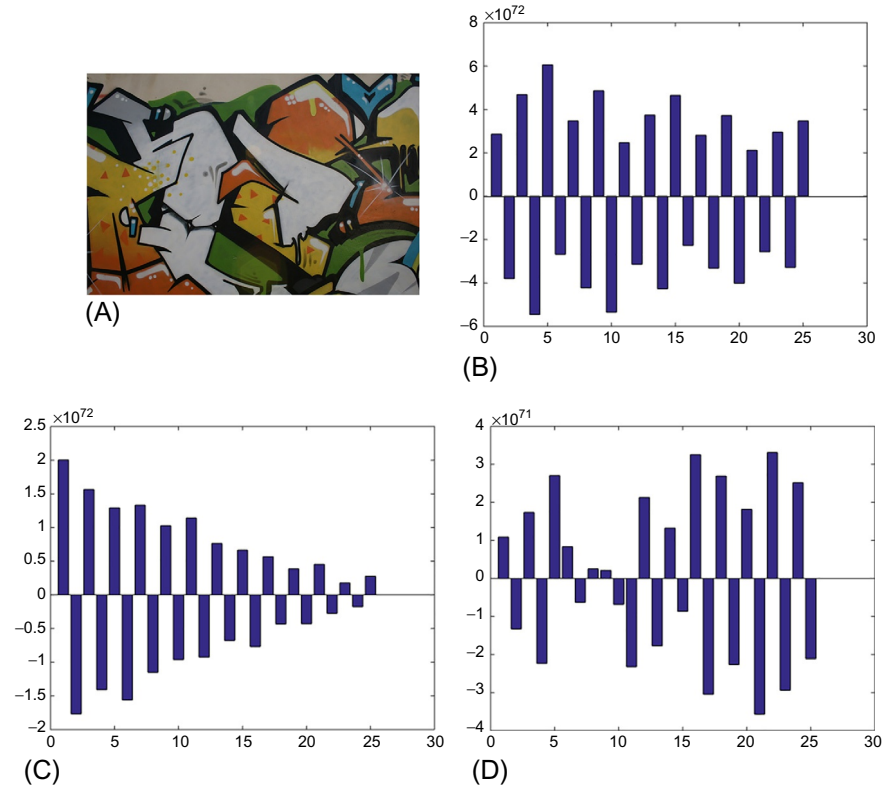


FIG. 1.8

Input image and harmonic color features (HCF) plots. (A) Input image, (B) hue, (C) saturation, (D) value.

represents another sample image where color in the image changes more rapidly. Fig. 1.8B–D represent their differential feature components.

1.3.2 Regression model design

We have applied our feature vectors on a real-world dataset and predicted an aesthetics score using these features. A multilayer NN model was used to train our system. We constructed one hidden layer with a hidden layer size of 5 in our NN. We trained the model with.45 learning rate and 500 training cycles. Combining three components, we used 75 features as our color features. We have preprocessed the feature set using PCA and validated our results using bootstrapping. We have taken 7 PCAs that capture 95% of variance. Fig. 1.9A shows a sample of our NN. Seven components are the input of the neural net and the predicted score is the output of the neuron. We used this network to produce regression values. We made a 70%–30% split (i.e., 3500 images against 1500 images) of the dataset for training and testing, respectively. We trained our model with validation, first training the model with 3150 images and validating with 350 images. We used the trained model to predict the score of the 1500 test images and calculated our prediction score.

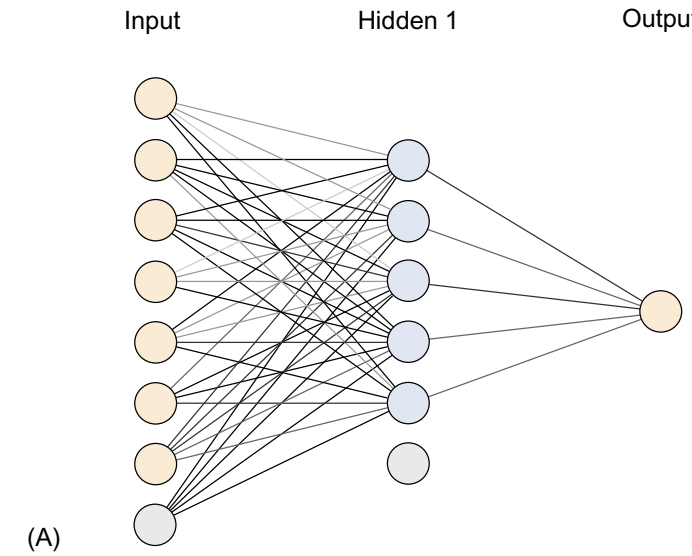
1.3.3 Dataset and user study

We used MIR Flicker [12] as our data source, and selected 5000 random images from all the available categories. Next, we conducted a user study to score selected images, with 374 participants in the survey, from whom we collected 12,748 scores on the 5000 images. The participants were given the following instruction: Rate the picture you are seeing, where 5 is the best score and 1 is the worst score.

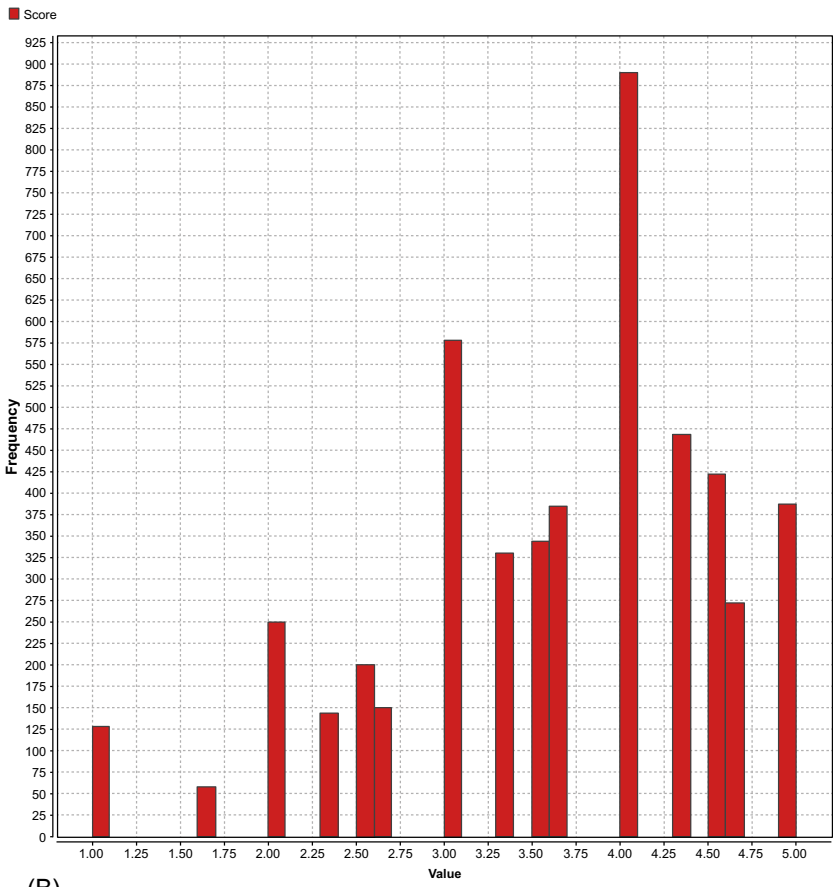
Fig. 1.9B shows the histogram of user score against image count. At least 2 and a maximum of 3 users scored all the images to eliminate bias. We took the average of the user score as our target value.

1.4 Results and analysis

After applying our model to the training data, we have calculated prediction performance, which is listed in Table 1.1. A scatter plot of our predicted score against the user given score has been presented in Fig. 1.10A and its density has been plotted in Fig. 1.10B. The X axis of the scatter plot shows the user given score, the Y axis shows the predicted scores, and the points are the 1500 predicted scores. The color scheme of Fig. 1.10B represents the predicted score, where red is the high score and blue is the higher score. If we observe the plot, we can notice that most of our prediction lies in the range of the score 3 to 4.5. This is because the majority of user given scores are within this range. We can verify this by observing Fig. 1.10B. Also, we can observe that the lower score has a higher density from 3 on the plot, where higher score densities reach towards a score of 5. From this, we can conclude that images not having a



(A)



(B)

FIG. 1.9

Diagram of NN and dataset distribution. (A) Diagram of NN; (B) histogram of dataset.

Table 1.1 Performance score of NN and LR.

Attributes	Neural network	Linear regression
Root mean squared error	1.115 ± 0.196	0.940 ± 0.013
Absolute error	0.868 ± 0.170	0.761 ± 0.007
Relative error	33.87% ± 6.84%	28.37% ± 0.88%
Prediction average	3.615 ± 0.017	3.609 ± 0.013

higher score does not imply that those images do not have color harmony, but images with the higher score have a certain level of color harmony. This insight is similar to our result from previous work [13], where we showed that images with higher score maintain a certain level of rule of thirds, but the opposite is not true. This finding revalidates the findings of Mai et al. [14] in a comprehensive manner for the first time in computer vision literature, to the best of the authors' knowledge. Along with our NN results in Table 1.1, we have also listed results using a linear regression (LR) prediction model. Although LR seems to have a better performance, its result is more biased. Its predictions lie in the range of 3.50 to 3.74, as the majority of the dataset lies in this region, while root mean squared error (RMSE) is lower. Eq. (1.4) is the equation for LR.

$$x = -0.002 * \text{pca}_1 - 0.009 * \text{pca}_2 - 0.006 * \text{pca}_3 + 0.053 * \text{pca}_5 + 0.024 * \text{pca}_6 + 0.027 * \text{pca}_7 + 3.615 \tag{1.4}$$

1.5 Discussion and future work

In this research, we have successfully developed a faster calculative method for extracting harmonic color features. We also have constructed a simplified method for perceiving beauty in images. The outcome of this research can be used not only in the domain of predicting aesthetics but also in other domains, i.e., cognitive model development, analysis of human perception of colors, color compositions, etc. We plan to create a model based on the human context of perceiving beauty as our future work. Along with this, we plan to combine features of this and our previous work [13] and create a robust model to perceive aesthetics with greater accuracy.

Acknowledgments

This research work is funded by North South University and we would like to thank all the participants from North South University who have participated in the user study of scoring the aesthetics of several images.

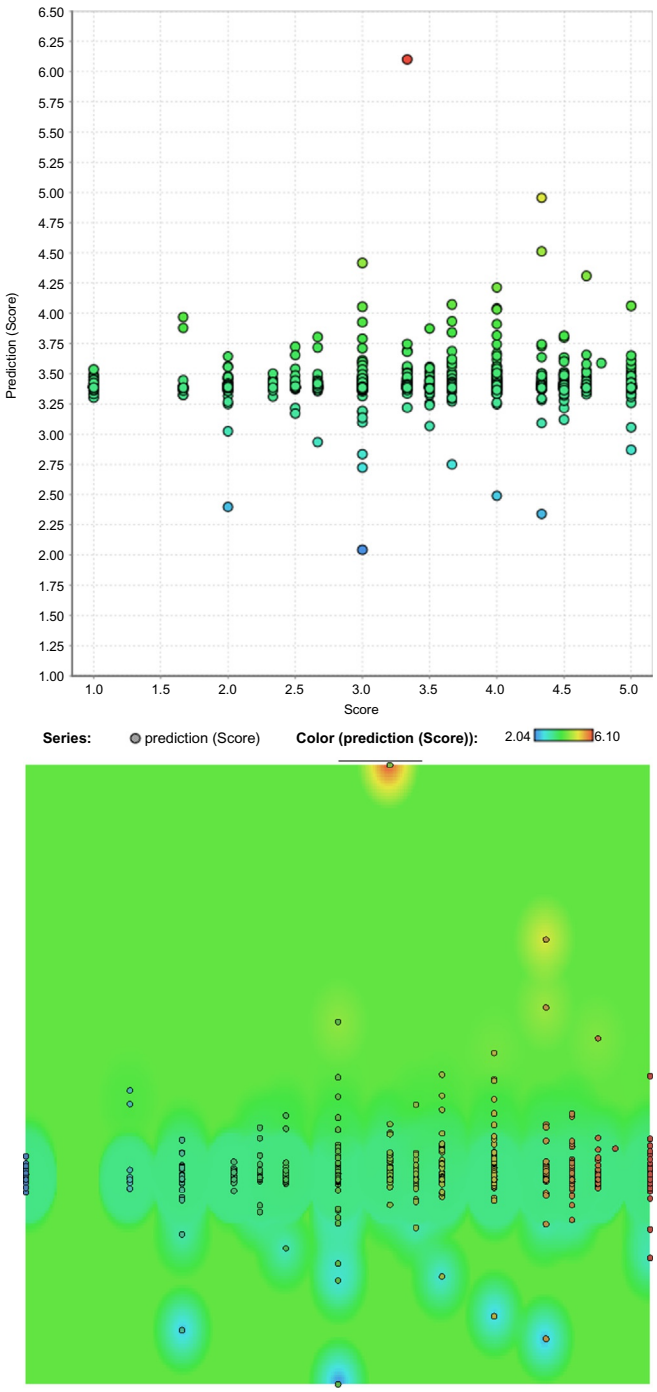


FIG. 1.10

Scatter and density plot of predicted scores. (A) Scatter plot; (B) density plot.

References

[1] P. Moon, D.E. Spencer, Geometric formulation of classical color harmony*, J. Opt. Soc. Am. 34 (1) (1944) 46.

[2] T. Stone, S. Adams, N. Morioka, Color Design Workbook: A Real World Guide to Using Color in Graphic Design, Rockport, 2017.

[3] P. Lu, X. Peng, X. Zhu, R. Li, An EL-LDA based general color harmony model for photo aesthetics assessment, Signal Process. 120 (2016) 731–745.

[4] P. Lu, X. Peng, C. Yuan, R. Li, X. Wang, Image color harmony modeling through neighbored co-occurrence colors, Neurocomputing 201 (2016) 82–91.

[5] R. Datta, D. Joshi, J. Li, J.Z. Wang, Studying aesthetics in photographic images using a computational approach, in: Computer Vision – ECCV 2006, Springer Berlin Heidelberg, 2006, , pp. 288–301.

[6] H. Phan, H. Fu, A. Chan, Color orchestra: ordering color palettes for interpolation and prediction, IEEE Trans. Vis. Comput. Graph. 24 (6) (2017) 1942–1955.

[7] C. Amati, N.J. Mitra, T. Weyrich, A study of image colourfulness, in: Proceedings of the Workshop on Computational Aesthetics—CAe'14, 2014.

[8] P. Lu, X. Peng, R. Li, X. Wang, Towards aesthetics of image: a Bayesian framework for color harmony modeling, Signal Process. Image Commun. 39 (2015) 487–498.

[9] M. Nishiyama, T. Okabe, I. Sato, Y. Sato, Aesthetic quality classification of photographs based on color harmony, in: CVPR 2011, 2011.

[10] Wikipedia. n.d. User:Datumizer—Wikimedia Commons. Retrieved from https://commons.wikimedia.org/wiki/User:Datumizer.

[11] R. Keys, Cubic convolution interpolation for digital image processing, IEEE Trans. Acoust. Speech Signal Process. 29 (6) (1981) 1153–1160.

[12] M.J. Huiskes, M.S. Lew, The MIR Flickr retrieval evaluation, in: ACM International Conference on Multimedia Information Retrieval (MIR'08), Vancouver, Canada, 2008.

[13] A. Firoze, T. Osman, S. Psyche, R.M. Rahman, Scoring photographic rule of thirds in a large MIRFLICKR dataset: a showdown between machine perception and human perception of image aesthetics, in: 10th Asian Conference on Intelligent Information and Database Systems (ACIIDS), Lecture Notes in Computer Science (LNCS), Vietnam, 2018.

[14] L. Mai, H. Le, Y. Niu, F. Liu, Rule of thirds detection from photograph, in: Proceedings of the IEEE International Symposium on Multimedia (ISM), Dana Point, CA, USA, 2011, pp. 91–96.