IMAGE HARMONY FOR CONSUMER IMAGES

Elena Fedorovskaya, Carman Neustaedter and Wei Hao

Eastman Kodak Company, Research Laboratories, Rochester NY 14650, USA

ABSTRACT

Images elicit a variety of emotional responses related to image content, overall aesthetic appeal, or a combination of both. One aspect of aesthetic appeal is harmony: the pleasing or congruent arrangement of parts producing internal calm or tranquility. We conducted a series of experiments to identify what factors, if any, could predict harmony in an image. Subjective judgments of image harmony were collected for images representative of typical consumer photography. Our results show that for simplified images (pixelated to control for emotional responses) reasons for image harmony are fairly dependent on the viewer, but typically involve edge contrast, average lightness, range of lightness, or the inclusion of Gestalt principles. Extraction of global image features may help to explain results with black and white and color images.

Index Terms— image harmony, image quality, computational assessment, consumer photography

1. INTRODUCTION

Research has shown that images elicit different emotional responses in individuals, often linked to the subject matter, depicted objects or scenes contained within the image [l, 2]. It is also known that aesthetically pleasing images must follow certain design principles, which many artists consider universal [3]. The growing field of computational aesthetics is focusing on revealing or applying these principles toward the analysis of visual and photographic art, music and other media [4]. At the same time, new multidisciplinary scientific fields such as neuroaesthetics are studying brain mechanisms involved in appreciation and emotional reaction to art [5].

Our focus is on understanding how aesthetic responses to images can be modeled for use within the field of computational aesthetics. There exist several open research problems of interest to us in this area. First, it is not clear if aesthetic response to an image can be experimentally separated from emotional reaction to the object or a situation depicted in an image. This may largely depend on the emotional life experiences of the observer, yet we do not know for certain. Second, we do not know if it is possible to understand and model individual aesthetic and emotional preferences to media. If available, such models could be applied toward consumer and amateur photography and visual art.

To investigate these problems, we designed a series of experiments aimed at understanding the aesthetic appeal of images as perceived by an individual observer. More specifically, they attempt to uncover what features of an image, if any, can be used to predict image harmony. Webster's New Collegiate Dictionary defines harmony as "a) pleasing or congruent arrangement of parts; b) correspondence, accord; c) internal calm, tranquility."

First, we describe findings from our first experiment which aims to uncover what factors can be used to predict image harmony in *simplified images*. Next, we compare these results to experiments involving *black and white images*, followed by experiments utilizing a series of *colored images*. Our findings show that even though image harmony is largely based on individual preferences, these preferences are often dictated by common image properties.

2. IMAGE HARMONY FOR SIMPLIFIED IMAGES

We first conducted an experiment on a simplified set of images as a baseline for predicting image harmony.

2.1. Data Sets and Pre-Processing

We selected thirty-four digital images (768 x 512 pixels in size) as a representative sample of common consumer photography. The content of the images varied between portraits and landscapes, indoor and outdoor scenes, different seasons, people, animals, and images taken under different lighting conditions (e.g., flash, bright sun, shadows, etc.). Images were then pre-processed to create Mondrian-like stimuli in order to control for the effects of emotional reactions to images and their semantic content, as opposed to aesthetic appeal. This process (sometimes used in psychophysical research) consists of transforming each image into a series of unidentifiable rectangular patches. In our experiment, we created Mondrians by first pixelating each image where code values in 20 x 20-pixel regions were averaged. The resulting images were then randomly flipped or rotated to make it difficult to recognize the original pictures. Figure 1 shows an example of an existing image (left) and the resulting Mondrian-like stimulus (right).

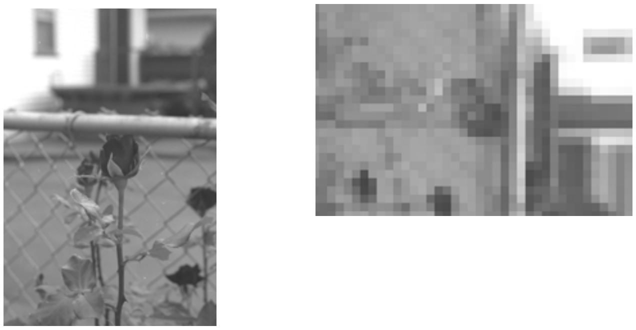

Figure 1.

If produced, for example, by averaging pixel values within rectangular areas of a certain size, Mondrians' statistics preserve, to a certain degree, the global statistics of the original images. Thus, rather than seeing a realistic consumer image that would likely create emotional reactions, subjects instead saw an image that contained similar visual features as the original image yet did not have any preconceived emotional bias. The use of Mondrians also aids analysis because the simplified nature of the images (e.g., well-defined edges, and absence of objects) can help identify image-based features that may be useful for predicting image harmony. This could then be applied to more natural scenes.

2.2. Experimental Setup

The experiment was performed in a lab setting in order to control the viewing distance, ambient illumination, display characteristics and image rendering properties. Stimuli were viewed by subjects on a 20" CRT monitor with a white point set to D50. Screen maximum luminance was 69.7 cd/m^2 and the gamma of the monitor was 2.14. Screen resolution was 1152 x 870 pixels. Subjects sat at a distance of 75 cm from the screen in a dimly illuminated dark room. Ambient indirect lighting of 5000 K from a 60-watt bulb produced luminance of 0.0398 cd/m^2 at the screen. An adapting neutral field with a luminance of 15 cd/m^2 served as a background for image presentation and was displayed between trials (this level corresponded to the L* value of 50). The visual angle of the stimuli was 17.7 x 11.57 degrees for a landscape-oriented image and 11.8 x 17.3 degrees for a portrait-oriented image. The screen size, and therefore the size of the adapting field, was 26.27 x 19.67 degrees of visual angle.

In comparison to web-based experiments, our stimuli and viewing conditions are well characterized, thereby reducing experimental noise. The drawback to this approach, however, is that a relatively small number of images and participants can be studied. Yet a smaller sample also provides an opportunity to examine the data for each individual participant more thoroughly. This becomes especially important in cases where intra- and inter-individual variability exist.

2.3. Experimental Procedure

Seventeen subjects (eleven male and six female) participated in the experiment. All had normal or corrected-to-normal visual acuity and their ages ranged from 20 to 55 years. Subjects were first given instructions that included a definition of image harmony as: "the congruency and agreement of various parts and attributes when viewing the image that produces a pleasing sensation of the adequacy and comfort to the viewer's eye."

Next, subjects completed a trial session that was used to assess the subject's degree of confidence and response consistency. Here each subject viewed 10 to 20 stimulus images from a test set, one at a time. For each, they were asked to rate the level of visual harmony contained within the image. Perceived image harmony was evaluated using a free modulus magnitude estimation technique, where observers are not given a reference image for the attribute; instead, they use their own internal reference. Subjects were told that the maximum visual harmony they could imagine would be equal to 100, and the lowest would be 0.

Finally, subjects were randomly presented each of the 35 stimulus images, one at a time, where each image was repeated four times per subject (interspersed amongst the other images). Thus, each subject saw a series of 140 images (including repetitions) and assessed the harmony found in each using the same metric as the trial sessions. Time presentation for every image was not limited, however, subjects were asked to reply as quickly as they possibly could. The average time for an experiment was 45-60 minutes, which included a break to avoid fatigue.

2.2. Analysis: Feature Extraction

We averaged the four harmony values that subjects gave for each stimulus image and then analyzed our data for key features which could potentially be used to predict image harmony. First, we extracted three edge and region related features [6] after converting an image to the CIELAB space.. We also tested other features, such as 1) standard deviation of L*, as a variant of edge contrast measure, 2) mean L*, 3) difference between mean lightness of an image and a background lightness of the monitor screen of 50L*, and 4) a number of segmented regions as the measure of complexity. Further, several symmetry measures were calculated. We used the SIFT[7] features as the local descriptors extracted from 8-by-8 pixels patches. The K-mean clustering was used to form the visual vocabulary. The number of clusters (vocabulary size) was set to be 200.

Each image was then split into 4 equal regions . The visual word counting histograms were created for each region. The similarity measures between each pair of regions were defined as an intersection of their histograms [8]. Those similarity measures were used as representing visual symmetry between corresponding regions.

In total, 9 symmetry related features (for horizontal, vertical and diagonal correlation between top, bottom, right and left part of the image), and 7 edge and region related features were computed.

In addition to this, we also evaluated the potential effect of spatial frequency information by computing amplitude and power spectra in different spatial visual frequency bands using 2D FFT analysis. However, none of these features significantly contributed to visual harmony prediction, so this data will not be discussed in this paper. Due to the small stimulus set and modified nature of our images we are considering our results as preliminary and exploratory, and find that the linear regression analysis is the most appropriate at this stage.

3. RESULTS

As one might expect, harmony scores varied substantially across the subjects, which resulted in low average pairwise correlation between individual assessments (0.14). Figure 2 presents the similarity-dissimilarity relationship between individual subjects' data revealed using Ward's clustering method. The graph beneath the dendrogram (Figure 2, bottom) shows the percentage of the total variance accounted by a chosen number of clusters. Four clusters marked with different colors present a possible solution. These four clusters (red, pink, blue, and green) account for a substantial portion of the variance, and further increases in the number of the clusters produces a slow improvement for the percentage of the explained total variance.

To understand the factors underlying our experimental data and their possible connection with the image features, we performed principal component analysis with subsequent varimax rotation. Three principal components together account for about 65% of the total variance, four principal components represent about 75% of the total variance, and five components represent about 80% of the variance. We performed varimax rotation for each of those solutions separately and inspected the obtained configurations and their correlations with the computed contrast- and lightness related features to understand the possible perceptual nature of the factors. Rotations were used to better align the directions of the factors with the original variables so that the factors may be more interpretable.

The four factors were interpreted as correlating with the following features: edge contrast (Factor#1, R^2=0.72), average lightness (Factor#2, R^2=0.39), range of lightness (Factor#3, R^2=0.65), and a "good figure" factor or "*prägnanz.*" The "good figure" factor represents the principle of perceptual organization discovered by Gestalt psychologists and reflects the tendency toward preference for "good figure", where "good" means "regular", symmetrical," "stable", or in short "simple". This factor appears to be relevant to the picture composition and correlates with the diagonal symmetry measure derived from visual vocabulary description (R^2=0.52). Figure 3a shows three images with the highest values for this factor, and Figure 3b shows two images with the lowest values.

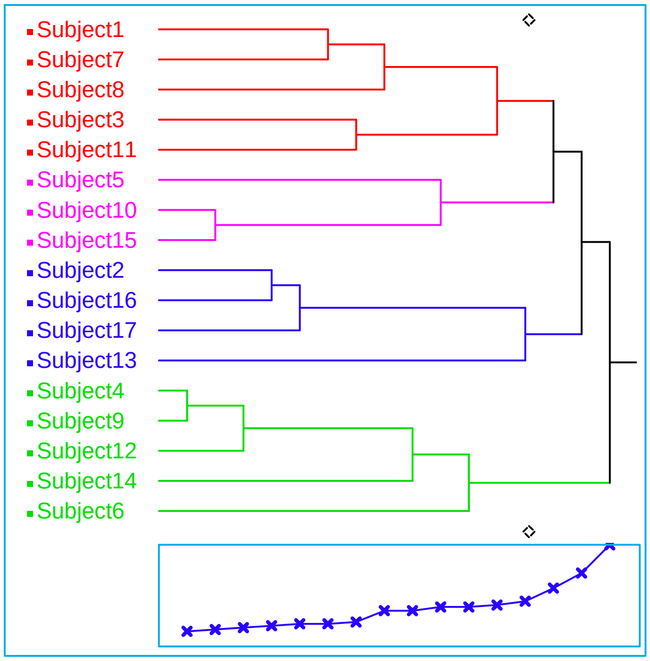

Figure 2.

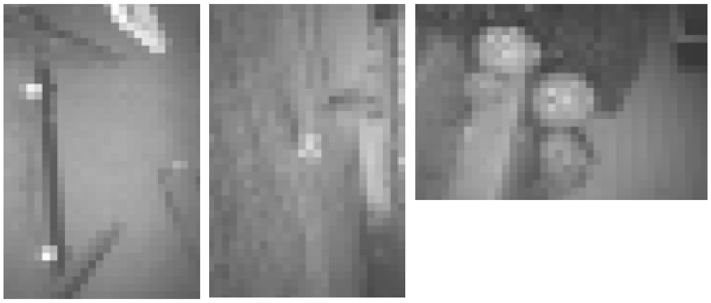

Figure 3a.

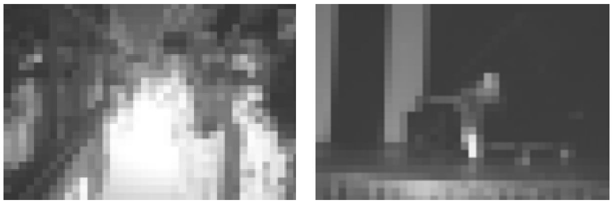

Figure 3b.

Comparison of mean harmony scores for each subject groups in Figure 2 with the four factors described above shows that the third group of subjects based their judgments of harmony almost exclusively on the edge contrast (Factor #1), while the second group tends to negatively respond to the same factor and positively to other factors, with the most emphasis on the "good figure" factor.

A linear regression analysis between harmony ratings and feature values was then performed for every individual subject. We have observed that in most cases we were able to successfully explain individual harmony ratings on the basis of just one, two or three features, though the feature sets utilized in the linear regression, their sign and relative magnitude were different for different subjects. Subject T's harmony ratings, for example, were negatively correlating with the edge contrast measure (R^2=0.86), while subject S's data, on contrary, positively correlated with the ratings ($R^{2=}$0.73). The final feature subset that contributed to analysis for all subjects contained 8 features from 20 computed. The R^2 for 9 subjects exceeded 0.56 and averaged 0.66. For the remaining 8 subjects prediction was less successful ($R^2 < 0.46$). Although we attempted to use symmetry features and region-based features as predictors, they do not sufficiently capture global structural properties of the images, which may influence visual harmony [2].

To obtain insights regarding additional image characteristics that affect harmony judgments, and also to learn how the simplified images compare with natural scenes, we conducted two additional experiments with black-and-white and color images, where we asked observers to view and rate overall perceived attributes, including quality for the black-and-white images, and harmony and quality for the color images. The experimental procedure was similar to the one described for mondrian experiments. Five subjects who viewed mondrians also evaluated black-and-white images, from which the mondrians were created. Forty six images were the same scenes in the black-and-white and color image sets. We supposed that this overlap would allow us to evaluate how increasing complexity of images and acquisition of properties of natural scenes affect judgments of harmony. We also hoped that it would help us ascertain other computationally extractable image characteristics related to visual harmony and aesthetics. In the present paper we plan to only provide initial qualitative analysis of the data and suggest hypotheses for further testing.

The experimental results showed that there was little correlation between quality judgments for black and white images and harmony judgments for mondrian images (r=0.2). At the same time there was a strong correlation between those attributes for color images (r=0.92). This discrepancy was particularly evident for two image groups. The images in the first group had significantly higher harmony scores for mondrians than the quality scores for the corresponding black and white images. The images in this group were blurred and had low contrast. The second group, in reverse, contained images with considerably higher quality ratings compared to the harmony ratings for the corresponding mondrians. Those images were high contrast outdoor or indoor scenes, all of them mirrored or rotated to create mondrians. For the outdoor scenes of this group the resulting mondrians often had a darker upper section and a lighter lower section due to change in the sky orientation. For the indoor close-up portraits, mondrian rotation created strong vertical fragmentation of the image. To explain such results global image features need to be considered [9]. Unusual orientation of an image may delay recognition which is based on extraction of high probability global features. This will influence perception of an image to appear less natural and less harmonious. We plan in depth investigation of the influence of global image features on aesthetic quality and visual harmony in future experiments.

4. REFERENCES

[1] P. J. Lang, M.K. Greenwald, M.M. Bradley, & A.O.Hamm. (1993). "Looking at pictures: Affective, facial, visceral, and behavioral reactions", *Psychophysiology*, **30**, 261–273.

[2] E.A.Fedorovskaya, P. Miller, G. Prabhu, C. Horwitz, T. Matraszek, P. Parks, R. Blazey, and Endrikhovski, S. "Affective imaging: psychological and physiological reactions to individually chosen image," *Proceedings of the SPIE Conference*, 4299, pp.524-532, 2001.

[3] M.Freeman. *The Image*. Collins, 1989, p.192.

[4] R. Datta, D. Joshi, J. Li and J. Wang, "Studying Aesthetics in Photographic Images Using a Computational Approach" in "*Lecture Notes In Computer Science*," **3953**, pp 288-301 Springer, 2006.

[5] S. Zeki, *Inner Vision; An Exploration of Art and the Brain*, Oxford University Press, 1999.

[6] E. Fedorovskaya, "Perceived Overall Contrast and Quality of the Tone Scale Rendering for Natural Images," Human Vision and Electronic Imaging VII, *Proc. SPIE Conference*, **4662**, pp. 119-128, 2002.

[7] D.Lowe. Towards a computational model for object recognition in IT cortex. In *Biologically Motivated Computer Vision*, pages 20–31, 2000.

[8] M.Swain and D.Ballard. Color indexing. *IJCV*, 7(1):11–32, 1991.

[9] A.Oliva, A.Torralba, "Modeling the Shape of the Scene: Holistic Representation of the Spatial Envelope" in *International Journal of Computer Vision* 42(3), pp. 145-175, 2001