

An Investigation into Three Visual Characteristics of Complex Scenes that Evoke Human Emotion

Xin Lu, Reginald B. Adams, Jr., Jia Li, Michelle G. Newman, and James. Z. Wang
The Pennsylvania State University, University Park, Pennsylvania, USA

Abstract—Prior computational studies have examined hundreds of visual characteristics related to color, texture, and composition in an attempt to predict human emotional responses. Beyond those myriad features examined in computer science, roundness, angularity, and visual complexity have also been found to evoke emotions in human perceivers, as demonstrated in psychological studies of facial expressions, dance poses, and even simple synthetic visual patterns. Capturing these characteristics algorithmically to incorporate in computational studies, however, has proven difficult. Here we expand the scope of previous computer vision work by examining these three visual characteristics in computer analysis of complex scenes, and compare the results to the hundreds of visual qualities previously examined. A large collection of ecologically valid stimuli (*i.e.*, photos that humans regularly encounter on the web), named the EmoSet and containing more than 40,000 images crawled from web albums, was generated using crowd-sourcing and subjected to human subject emotion ratings. We developed computational methods to the separate indices of roundness, angularity, and complexity, thereby establishing three new computational constructs. Critically, these three new physically interpretable visual constructs achieve comparable classification accuracy to the hundreds of shape, texture, composition, and facial feature characteristics previously examined. In addition, our experimental results show that color features related most strongly with the positivity of perceived emotions, the texture features related more to calmness or excitement, and roundness, angularity, and simplicity related similarly with both of these emotions dimensions.

1. Introduction

Everyday pictorial scenes are known to evoke emotions [1]. The ability of a computer program to predict evoked emotions from visual content will have a high impact on social computing. It remains unclear, however, what specific visual characteristics of scenes are associated with specific emotions, such as calmness, dynamism, turmoil, or happiness. Finding such associations is arguably one of the most fundamental research problems in visual arts,

psychology, and computer science, which has resulted in many relevant research articles [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13] over past decades. In particular, hundreds of visual characteristics, ranging from edge distributions [10], [12] and color histograms [14], [15] to SIFT, GIST, and Fisher Vector [16], [17] have been examined by recent computational studies to predict emotional responses.

Beyond those visual characteristics, many empirical studies in psychology and visual arts have investigated roundness, angularity, and complexity (which is the opposite of simplicity), and their capacity to evoke emotion. By performing experiments using facial expressions [2], dancing poses [2], and synthetic visual patterns [4], studies indicated that more rounded properties led to positive feelings, such as warmth and affection, whereas more angular properties tended to convey threat [2], [3], [18]. Meanwhile, a study in visual arts showed that humans preferred simple visual scenes and stimulus patterns [5].

To expand the scope of previous studies, herein, we investigated the *computability* of these three visual characteristics of complex scenes that evoke human emotion. While many other factors can be associated with emotional responses for complex scenes, including the semantic content of the scene and the personality, demographics, personal experience, mental state, and mood of the perceiver, from a pure visual characteristics perspective the aforementioned studies and our previous work [12] gave us confidence to explore this area further. While angularity appears to be the opposite of roundness, they are not exactly opposite in complex scenes (*e.g.*, a scene may have neither or both).

Describing these characteristics mathematically or computationally, with the purpose of predicting emotion and understanding the capacity of these characteristics to evoke emotion, is nontrivial. In our earlier research, we developed a collection of shape features that encoded the visual characteristics of roundness, angularity, and complexity using edge, corner, and contour distributions [12]. Those features were shown to predict emotion to some extent, but they were of hundreds of dimensions making them difficult to interpret. Under classification or regression frameworks, representations in different dimensions are intertwined for emotion prediction. Therefore, articulating what specific visual characteristics might be associated with a certain emotion turns out to be extremely difficult.

This work proposes novel computational methods to map visual content to the scales of roundness, angularity,

• This material is based upon work supported by the National Science Foundation under Grant No. 1110970. The authors would like to thank Michael J. Costa for assistance in data collection. Correspondence should be addressed to X. Lu and J. Z. Wang.

and complexity as three new constructs. By classifying emotional responses to natural image stimuli using the three constructs, we examined their capacity to evoke certain emotions. By computing the correlation between an individual construct and a dimension of emotional response, we examined their interrelationship statistically.

Meanwhile, we have been working on developing a large collection of ecologically valid stimuli for the research community. We named the dataset the **EmoSet**. Unlike the widely used The International Affective Picture System (IAPS), for which 1,082 images were rated with emotional responses, the EmoSet is much larger and all images are complex scenes that humans regularly encounter in life. Specifically, we collected 43,837 images and manually-supplied emotion labels (both dimensional and categorical) following strict psychological subject study procedures and validation approaches. These images were crawled from more than 1,000 users' web albums using 558 emotional words as search terms. The 558 emotional words were summarized by Averill [19]. We used the 558 words to search Flickr. We introduce the data collection in Section 2.1.

The EmoSet marks the first time a human subject study, examining *perceived emotion* triggered by visual stimuli, was performed in an uncontrolled environment using common photographs of complex scenes. The EmoSet is also different from large-scale affective datasets introduced in [17] and [20], where researchers crawled user-generated content, including pictorial ratings or associated affective tags, to indicate the affective intention of images. Whereas those datasets were of large scale, the emotional labels were not generated by human subjects under strict psychological procedures. Psychological conventions were applied when recording human perceived emotions, *e.g.*, aroused emotions are best recorded within six seconds after subjects view each visual stimulus [1]. More details will be provided later.

We quantized and investigated roundness, angularity, and complexity by the three constructs using the EmoSet. Our **main findings** were as follows.

- There were statistically significant correlations between the three constructs and emotional responses, *e.g.*, complexity and valence, and angularity and valence.
- The capacity of the three constructs to classify the positivity of emotional responses was established. When combined with color features, the constructs achieve comparable classification accuracy on the positivity of emotions as a set of over 200 shape, texture, composition, and facial features. This reduces the number of features required for classification by about two orders of magnitude.
- The three constructs were *completely interpretable* and could be used in other applications involving roundness, angularity, and simplicity/complexity of visual scenes.
- Our experimental results indicated that among the color, texture, shape, facial, and composition features, color features showed higher capacity in classifying the positivity of emotional responses, whereas texture features showed higher capacity in distinguishing calmness from excitement. The three constructs showed consistent capacity in classifying both dimensions of emotions.

2. The Approach

2.1. Creating the EmoSet Dataset

To have a large collection of photographs with complex scenes, we crawled more than 50,000 images from Flickr.com, one of the most popular web albums (Fig. 1). We chose this site because of its large and highly diverse user base, and its focus on managing rather than sharing personal photos. Images of certain categories were removed (explained below). We performed a human subject study on those photographs and developed a large-scale ecologically valid image stimuli, *i.e.*, the EmoSet. The human subject study was empowered by crowdsourcing and computational tools in order to recruit a diverse population of human subjects. We incorporated strict psychological procedures into the User Interface (UI) design.



Figure 1. Example images crawled. Images with faces were removed.

As a result, the EmoSet contains 43,837 color images associated with emotional labels, including dimensional labels, categorical labels, and likeability ratings. Subjects' demographics were also collected such as age, gender, ethnic groups, nationality, educational background, and income level. In addition, we collected semantic tags and other metadata associated with the images in the EmoSet.

2.1.1. Crawling Ecologically Valid Image Stimuli. To collect a large quantity of image stimuli, we took 558 emotional words summarized by Averill [19] and used those words to retrieve images by triggering the Flickr image search engine. For each emotional word, we took the top 100 returned images to ensure a high correlation between images and the query. The crawled images were generated by ordinary web users and contained complex scenes that humans may encounter in daily life. We removed duplicates, images of bad taste (*e.g.*, highly offensive, sexually explicit), images with computationally detectable human faces, and the ones primarily occupied by text.

2.1.2. The Human Subject Study. In our efforts to establish a large labeled set of image stimuli, we leveraged crowdsourcing and computational tools, and incorporated Lang and Bradley’s methods in creating and validating the IAPS [1] into our study design. We detail the design rationale below.

Inspired by the concept of semantic meaning, defined in Charles E. Osgood’s Measurement of Meaning as “the relation of signs to their significants” [21], we asked human subjects to evaluate a series of color images from three perspectives: (I) by rating them along the three dimensional scales — valence, arousal, and dominance, (II) by selecting one or more categorical emotions if relevant, and (III) by selecting their level of like/dislike toward every presented image (*i.e.*, likeability).

In part I, we adopted a dimensional approach in an attempt to understand the emotional characteristics that people associate with a vast array of images. The dimensional approach was also used in the creation of the IAPS [1], whose strengths have been indicated by recent studies in psychology [22], [23], [24]. In line with the IAPS study, we utilized the Self-Assessment Manikin (SAM) instrument, recording a rating for the three dimensions: valence, arousal, and dominance. A 9-point rating scale was employed to quantify the emotional ratings on the three dimensions. Instead of the static SAM instrument used in the IAPS, we implemented a dynamic SAM instrument, which could easily be manipulated by sliding a solid bubble along a bar. This was motivated by Lang’s claim that “SAM was presented to subjects as an oscilloscope display under the control of a joy-stick on his chair arm” and “An initial instruction program associates the visual anchors of each display with the polar adjectives used by Mehrabian to define the semantic differential scales for the three affective factors” [25]. A gradually changed expression on the dynamic SAM allowed for a more “natural” rating experience for human subjects. As a result of making SAM dynamic, it was necessary to display a single SAM figure for each dimension, minimizing the clutter that would otherwise exist with three rows of static SAM figures, varying slightly in expression.

We collected categorical emotion labels in part II, where eight basic emotions discussed in [26] were included. We displayed the emotions with a checkbox next to the sentence “Click here if you felt any emotions from the displayed image”. Participants were allowed to enter one or more emotions that were not included in the list provided by selecting the checkbox next to the word “Other”, whereby a blank text box would appear.

We collected likeability ratings in part III. We included likeability as an affective measure of images to indicate the extent to which subject liked images. To quantify the likeability, we included a scale for participants to select: like extremely, like very much, like slightly, neither like nor dislike, dislike slightly, dislike very much, and dislike extremely.

Motivated by the subjective nature of analyzing image affect, we also collected demographics of the human sub-

jects, including their age, gender, ethnic groups, nationality, educational background, and income level. Such information helps understand the generality of our findings regarding the population.

2.1.3. Human Subject Study Procedures. Detailed procedures of the human subject study are introduced here. Once a participant clicked the “agree” button on the consent form, we presented him/her with the instructions for participating in the study. We allowed five to ten minutes for participants to review the instruction, and each subject was asked to evaluate 200 images in total. We briefly summarize the human subject study procedures below.

- *Step 1* The subject clicks the “Start” button. After five seconds the subject will be presented with an image.
- *Step 2* The subject views the image that displays for six seconds.
- *Step 3* A page with three parts will display, and the subject is allotted about 13–15 seconds to fill out these parts. For part I, the subject was asked to make a rating on each scale (valence, arousal, and dominance), based on how they *actually felt* while they observed the image. The subject was asked to complete part II only if they felt emotions by selecting one or more of the emotions they felt and/or by entering the emotion(s) they felt into “Other.” For part III, the subject was asked to rate how much they liked or disliked the image. The subject then clicked “Next” in the lower right hand corner when finished with all three parts.
- *Step 4* The subject repeated “Step 2” and “Step 3” until a button with the word “Finish” was displayed.
- *Step 5* The subject clicked the “Finish” button.

In *Step 1*, we followed psychological convention in [1], and set six seconds as the default value in *Step 2* for image viewing. This was because our intention was to collect *immediate* affective responses from participants given the visual stimuli. If subjects needed to refer back to the image, they were allowed to click “Reshow Image” in the upper left part of the screen, and click “Hide” to return to the three-part questionnaire. We expect this function to be used only occasionally.

2.1.4. Dataset Statistics. We statistically analyzed the EmoSet, including the collected emotional labels and subjects’ demographics. Each image in the EmoSet was evaluated by at least three subjects. To reduce low-quality ratings, we removed ratings with a total viewing duration shorter than 2.5 seconds. Borrowing from the procedures used in the IAPS, valence, arousal, and dominance were rated on a scale from 1 to 9; likeability was rated on a scale from 1 to 7, borrowing from the photo.net website. We showed the distributions of mean values in valence, arousal, dominance, and likeability in Fig. 2. Statistics of the IAPS are in Fig. 3.

The human subject study involved both psychology students within Penn State University and users of the Amazon Mechanical Turk, which ensured a diverse population of emotional ratings. Among the 4,148 human subjects we recruited, there were 2,236 females and 1,912 males, with

ages ranging from 18 to 72, various ethnic groups including American Indian or Alaska Native, Asian, African American, native Hawaiian or Other Pacific Islander, Hispanic or Latino, and Not Hispanic or Latino. Participants also had diverse income and education levels. Due to space we cannot provide a dissection.

2.2. The Three New Computational Constructs

To articulate a specific relationship between roundness, angularity, and simplicity with human emotion, this paper proposes computational methods to map images to the scales of roundness, angularity, and simplicity as three new computational constructs. We detailed the three constructs in this subsection.

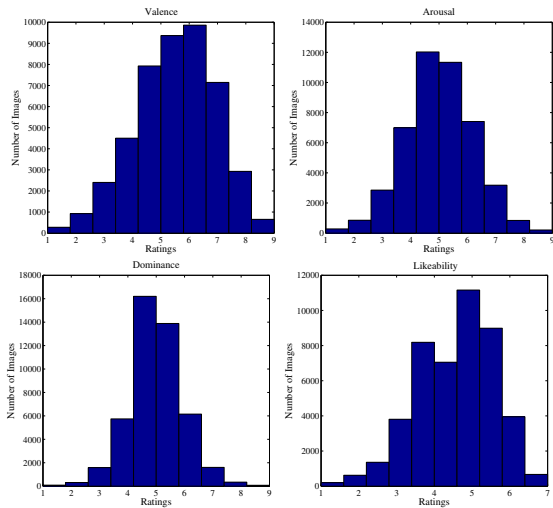


Figure 2. Mean value distributions of valence, arousal, dominance, and likeability in the EmoSet.

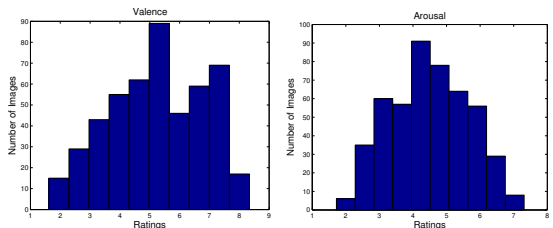


Figure 3. Mean value distributions of valence and arousal in the IAPS.

2.2.1. Roundness. Roundness was defined as “a measure of how closely the shape of an object approached that of a circle.” [27]. To compute the roundness score of an image, we first segmented the image into regions, then traced their boundaries, and finally computed the goodness of fit to a circle for each region. The step-by-step procedure was:

- 1) The segmentation approach in [28] was adopted. Suppose the segments are $\mathcal{S} = \{S_1, S_2, \dots, S_N\}$, where the number of segments N was automatically determined by the algorithm. Let the set of boundary points of segment S_i be $B_i = \{(x_j, y_j)\}$.

- 2) The Pratt Algorithm [29] was applied to find the circle C_i best fitted to B_i . Denote the center of the circle by (c_i, d_i) and radius by u_i . The Pratt Algorithm was applied because of its capacity to fit incomplete circles, *i.e.*, arcs of any degree.
- 3) For each segment, we defined the roundness disparity of S_i by $r_i = \sigma(d(B_i, C_i))$. Denoted by $d(B_i, C_i)$ a set of distance between each point in B_i to C_i , and denoted by σ the standard deviation of that set. The distance between a point (x_i, y_i) and a circle C_i was computed by the absolute difference between the radius u_i and the Euclidean distance from the point to the center of the circle.
- 4) The roundness disparity of an image I was denoted by $r_I = \min_{i=1}^N r_i e^{-\lambda r_i / \max(v, h)}$. Denoted by v the number of rows and h the number of columns of the image I .

In the experiments, we set $\lambda = 0.5$ and normalized the roundness disparity values to the interval $[0, 1]$. We quantified the roundness score as $1 - r_I$, so the closer r_I was to 1 meant that the image was associated with an obvious round property, and to 0 the opposite. We present examples of images and their roundness scores in Fig. 4. The images with highest roundness scores are shown in the first row; images with medium ranges of roundness scores in the second row; and images with lowest roundness scores in the third row.

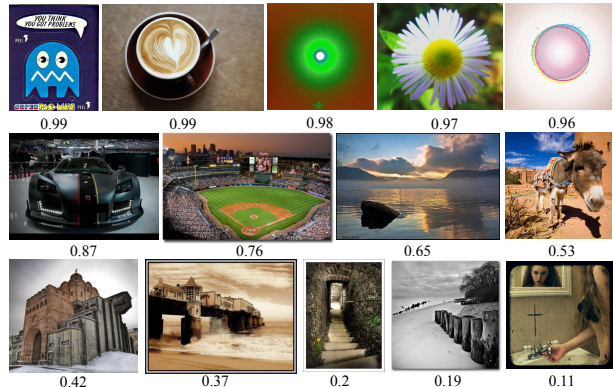


Figure 4. Example images and their roundness scores.

2.2.2. Angularity. In the Merriam-Webster Dictionary, angularity is defined as “the quality of being angular,” and angular is explained as being lean and having prominent bone structure. We also interviewed five individuals, including one undergraduate student, three graduate students, and one faculty member. They considered angularity to be related to “sword-like”, tall, thin, or narrow objects. These clues motivated us to examine how similar object boundaries were to long ellipses. Similar to roundness, an image was segmented into regions, for each of which an angularity measure was computed. We approximated the quality of being lean and having prominent bone structure by the elongatedness of fitted ellipses. Specifically, the angularity score of an image was computed as following:

- 1) For each set B_i , least-squares criterion was used to

estimate the best fit to an ellipse E_i^1 . We denoted the center of the ellipse by (c_i, d_i) , semimajor axis by m_i , semiminor axis by n_i , and angle of the ellipse by e_i .

- 2) For each image segment S_i , we denoted the angularity of a region i by $a_i = m_i/n_i$. As our goal was to find lean ellipses, we omitted horizontal and vertical ellipses according to e_i and ellipses that were too small in area.
- 3) We computed angularity of the image I , denoted by $a_I = \max_i^N a_i$.

Angularity scores for images in the EmoSet were computed and normalized to $[0, 1]$. The closer a_I was to 1 meant that the image showed an obvious angular property. Examples of images and their angularity scores are presented in Fig. 5.

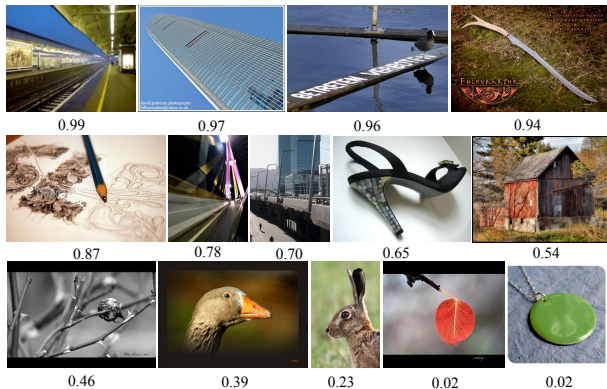


Figure 5. Example images and their **angularity** scores.

2.2.3. Simplicity (Complexity). According to [5], simplicity (complexity) of an image is primarily depending on two objective factors: minimalistic structures that are used in a given representation and the simplest way of organizing these structures. Motivated by such concept, we used the number of segments in an image as an indication for its simplicity. We defined the complexity score by $s_i = |S|$ and normalized the scores to $[0, 1]$ for images in the EmoSet. Simplicity and complexity, being opposite, were essentially represented by the same construct. We thus omitted complexity in the later presentations. We present examples of images and their scores of simplicity in Fig. 6.

3. The Primary Findings

In this section, we present the three major findings of this study, *i.e.*, statistical correlations between roundness, angularity, and simplicity and human emotion (Section 3.1), the capacity of the three constructs to classify the positivity of perceived emotion (Section 3.2), and the power of various visual features to classify the positivity and calmness of perceived emotion (Section 3.3).

Whereas psychological conventions treated roundness and angularity as opposite properties, many natural photographs showed neither of the properties. As the goal of the study was to examine the capacity of roundness,

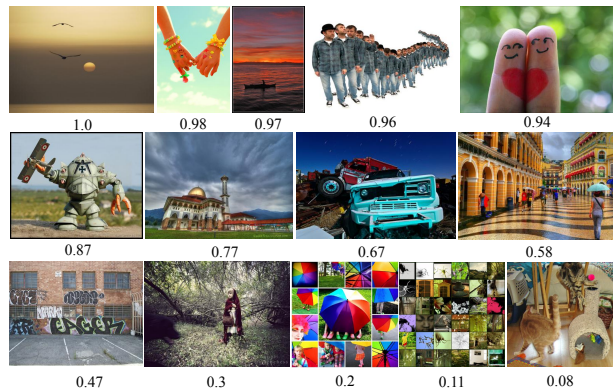


Figure 6. Example images and their **simplicity** scores.

and simplicity in evoking human emotion, we targeted visual stimuli with at least a non-zero construct of roundness or angularity. We thus removed 12,158 images from the EmoSet because they were associated with zero constructs of both roundness and angularity. The process resulted in 31,679 retained images. It is possible to define the constructs differently or to include additional ones so that more images can be incorporated.

3.1. Statistical Correlation Analysis

To examine the intrinsic relationship between the three constructs and evoked emotion, we computed correlations between one construct, such as simplicity, roundness, and angularity, and a dimension of the emotional response, such as valence, arousal, and dominance. All the correlations² were considered statistically significant, except for the correlation between roundness and likeability. Results are shown in Figs. 7, 8, and 9 (also in Table 1). The red number at the top left corner indicates the statistically significant correlations in terms of p-value.

TABLE 1. CORRELATION COEFFICIENTS

	Simplicity	Roundness	Angularity
Valence	0.11	0.01	0.05
Arousal	0.09	0.02	0.03
Dominance	0.04	0.02	0.02
Likeability	0.07	-0.01	0.05

In particular, the strongest correlation coefficient was between simplicity and valence, *i.e.*, 0.11. Whereas the correlation coefficients are small numerically from a psychological perspective, they were computed on 31,679 images containing complex backgrounds and evaluated in uncontrolled user subject study settings. These facts make the correlation analysis statistically meaningful. As the p-value was much smaller than 0.0001, the intrinsic relationships between simplicity and four dimensions of perceived emotions were regarded as present. The correlation coefficients between angularity or roundness and perceived emotion were smaller than those between simplicity and perceived

1. <http://www.mathworks.com/matlabcentral/fileexchange/3215-fit-ellipse>

2. $\alpha = 0.05$ for 95% confidence intervals.

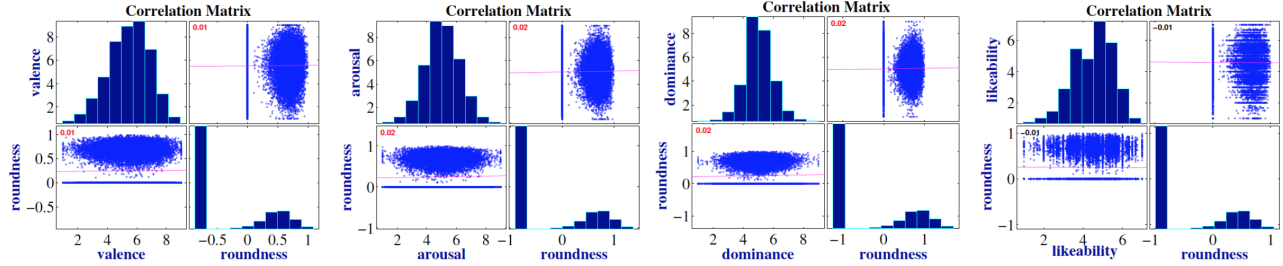


Figure 7. Correlation between **roundness** and valence, arousal, dominance, and likeability in natural photographs.

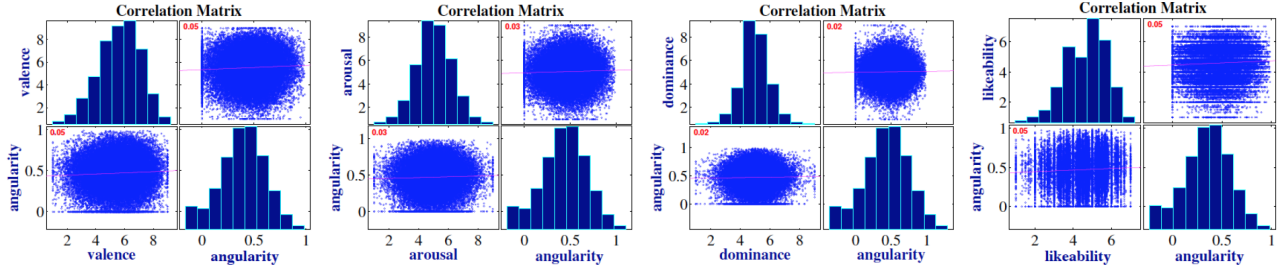


Figure 8. Correlation between **angularity** and valence, arousal, dominance, and likeability in natural photographs.

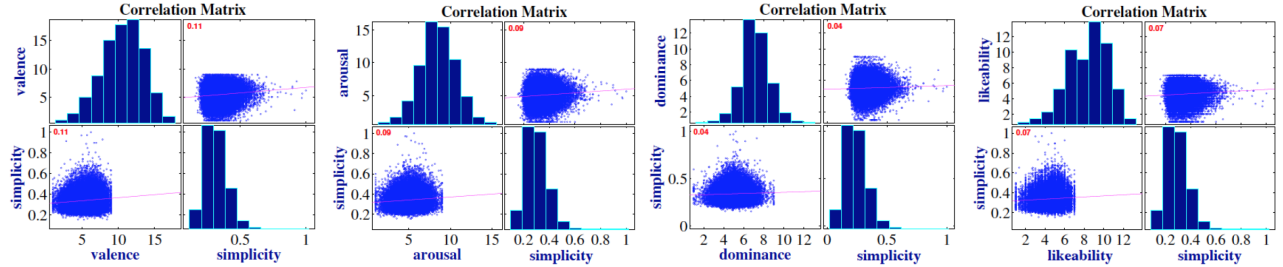


Figure 9. Correlation between **simplicity** and valence, arousal, dominance, and likeability in natural photographs.

emotion, which implied that simplicity related more strongly with perceived emotion compared to angularity and roundness on an arbitrary photograph.

3.2. The Capacity of the Three Constructs

To examine the capacity of the three constructs to classify the emotional responses to natural image stimuli, we formulated a classification task to distinguish positive emotions from negative ones, *i.e.*, high vs. low valence. This task has high application potential. We did not use a regression because even classification was proven difficult for emotions. The scale of valence ranged from 1 to 9, where 1 referred to the lowest value in valence and 9 the highest. Images with a medium-range score, such as 5, showed neither positive nor negative emotions. Following conventions in aesthetics studies that a gap may be applied to facilitate classifier training, we adopted a gap of 1.87 to divide image collections into two groups, images arousing positive emotions (valence > 6.63) and negative emotions (valence < 4.5). The gap is selected as a reasonable compromise between separation power and portion of images retained. To adjust the classifier parameters and evaluate the trained classifier, we randomly divided the data into training, validation, and testing sets, where the number of images

with positive and negative emotions were equal. Specifically, we randomly selected 70% of the data used for training, 10% for validation, and 20% for testing. This resulted in 12,600 images for training, 1,800 for validation, and 3,600 for testing. The SVM classifier with the RBM kernel, one of the popular classifier training approaches, was applied with 10-fold cross-validation. Among the 144 pairs of parameter candidates, the best performing c and g were selected given their performance on the validation dataset.

Various visual features were used in the classification. Color, texture, facial, and composition features were computed as presented in [8] and shape features as in [12]. The three constructs were computed as described in Section 2.2. “3-constructs” refers to the concatenation of the three constructs, and “color+3-con” denotes the concatenation of color features and the three constructs. “C.T.C.F.+3-con” refers to a concatenation of color, texture, composition, facial, and the three constructs. “C.T.C.F.+shape” refers to the concatenation of color, texture, composition, facial, and shape feature. While there are many other ways to combine, space limitation allow us to show only example results. As shown in Table 2, the 73-dimensional “color+3-con” feature improved upon the “color” feature slightly. Compared to the 332-dimensional features, “color+3-con” achieved a competitive and even better classification results

TABLE 2. CLASSIFICATION RESULTS OF HIGH VALENCE VS. LOW VALENCE

Features	roundness	angularity	simplicity	3-constructs	color	color+3-con	C.T.C.F.+shape	C.T.C.F.+3-con
Dimension	1	1	1	3	70	73	332	116
Accuracy(%)	51.08	54.75	57.36	58.08	64.42	64.97	64.86	65.5

using low-dimensional features (73 dimensions). We also noticed that the best classification results were achieved by “C.T.C.F.+3-con”, which clearly demonstrated the capacity of roundness, angularity, and simplicity of evoking the positivity or negativity of human emotion.

3.3. Visual Characteristics and Emotion

To examine the ability of visual characteristics of complex scenes to evoke different dimensions of emotion, we classified the calmness of emotions, *i.e.*, high vs. low arousal following an approach similar with what was used with positivity of emotions. First, a gap of 2.7 was adopted to divide image collections into two groups, high arousal (arousal > 6.4) and low arousal (arousal < 3.7). Next, training, testing, and validation sets were generated, where 7,000 images were used for training, 1,000 for validation, and 2,000 for testing. Finally, the SVM classifier was trained and the best set of parameters was selected according to their performance on the validation set.

We compared classification results in valence and arousal using color, texture, composition, shape, and the three constructs. Because dominance is often confusing to human subjects, we did not classify that. The results are presented in Table 3. As shown in the Tables 2 and 3, color features performed the best among the five feature groups for distinguishing high-valence images from low-valence ones. Texture and shape features performed better at classifying images that aroused calm emotions and excited emotions. The three constructs showed consistent predictability for both of the classification tasks using merely the three numbers as the predictors.

TABLE 3. EMOTION CLASSIFICATION RESULTS

Features	three constructs	color	texture	composition	shape
Dimensions	3	70	26	13	219
Accuracy for valence (%)	58.08	64.42	61.47	62.58	60.19
Accuracy for arousal (%)	56.1	58	59.55	56.15	58.7

4. Discussions

The research approach and findings from this study may have lasting impact on studies in psychology, visual arts, and computer science with respect to the three visual characteristics investigated and other visual characteristics that evoke human emotion. In previous psychological and visual arts studies, many hypotheses were made regarding the visual characteristics and their capacity to evoke emotion. From the perspective of color vision, Changizi *et al.* indicated that different colors led to different emotional states [30]. Emotions

of various gender, age, and ethnic groups were investigated by Rodgers *et al.* through examining the frames in news photographs on three dimensions of emotion [31]. From the perspective of roundness and angularity, Aronoff *et al.* showed that increased roundedness leads to more warmth, and increased linearity, diagonality, and angularity of forms lead to feeling threatened [32]. Aronoff *et al.* later confirmed that geometric properties of visual displays conveyed emotions such as anger and happiness [2]. Similar hypotheses were made and demonstrated by Bar *et al.* [4] showing that curved contours led to positive feelings and sharp transitions in contour triggered a negative bias. Meanwhile, Reber *et al.* found that beauty was reflected through the fluency of perceivers in processing an object, and the more fluently the perceiver interpreted the object, the more positive was the response [3]. That paper reviewed factors that may have had an impact on aesthetic responses, including figural goodness, figure-ground contrast, stimulus repetition, and symmetry, and confirmed the findings by monitoring the influences introduced by changes of those factors. Visual arts studies also indicated that humans visually preferred simplicity. Any stimulus pattern was always perceived in the most simplistic structural setting.

In these studies, conventional approaches, such as human subject studies and interviews, were adopted, which limited the generalizability of the study in complex scenes. The ecologically valid EmoSet, the crowdsourcing approach for collecting visual stimuli, and the computational approach to studying the visual characteristics proposed by our work could be adopted in future studies of this kind. The findings presented in this paper may advance emotion-related studies. Possible future research questions may include: 1) In which scenarios do roundness, angularity, and simplicity have the strongest capacity of evoking human emotion; and 2) How does demographic information affect aroused emotion when humans encounter complex scenes.

The findings in this paper may motivate potential applications of identifying image affect in computer vision and multimedia systems. For instance, in image retrieval systems, the constructs of simplicity and roundness could be incorporated into the ranking algorithm to help arouse positive feelings of users, as they use image search engines. The construct of angularity could be utilized to help protect children from viewing pictures that stimulate anger, fear, or disgust, or contain violence. Similarly, image editing softwares could take the findings of our work into account when making design suggestions to photographers or professionals. Meanwhile, findings from this study may contribute to relevant research studies in computer science, such as automatic predictions of image memorability, interestingness, and popularity.

5. Conclusions and Future Work

This paper investigated three visual characteristics of complex scenes that evoked human emotion utilizing a large collection of ecologically valid image stimuli. Three new constructs were developed that mapped the visual content to the scales of roundness, angularity, and simplicity. Results of correlation analyses, between each construct and each dimension of emotional responses, showed that some of the correlations were statistically significant, e.g., simplicity and valence, angularity and valence. And classification results demonstrated the capacity of the three constructs in classifying both dimensions of emotion. Interestingly, by combining with color features, the three constructs showed comparable classification accuracy on distinguishing positive emotions from negative ones to a set of 200 texture, composition, facial, and shape features.

As future work, the proposed approach could be applied to examine other visual characteristics that evoke human emotion in complex scenes. We expect that our efforts may contribute to research regarding visual characteristics of complex scenes and human emotion from perspectives of visual arts, psychology, and computer science.

References

- [1] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system: Affective ratings of pictures and instruction manual," in *Technical Report A-8, University of Florida, Gainesville, FL*, 2008.
- [2] J. Aronoff, "How we recognize angry and happy emotion in people, places, and things," *Cross-Cultural Research*, vol. 40, no. 1, pp. 83–105, 2006.
- [3] R. Reber, N. Schwarz, and P. Winkielman, "Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience?" *Personality and Social Psychology Review*, vol. 8, no. 4, pp. 364–382, 2004.
- [4] M. Bar and M. Neta, "Humans prefer curved visual objects," *Psychological Science*, vol. 17, no. 8, pp. 645–648, 2006.
- [5] R. Arnheim, *Art and Visual Perception: A Psychology of the Creative Eye*. University of California Press; Second Edition, Fiftieth Anniversary Printing edition, 1974.
- [6] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Studying aesthetics in photographic images using a computational approach," in *European Conference on Computer Vision (ECCV)*, 2006, pp. 288–301.
- [7] R. Datta, J. Li, and J. Z. Wang, "Algorithmic inferencing of aesthetics and emotion in natural image: An exposition," in *International Conference on Image Processing (ICIP)*, 2008, pp. 105–108.
- [8] J. Machajdik and A. Hanbury, "Affective image classification using features inspired by psychology and art theory," in *ACM International Conference on Multimedia (MM)*, 2010, pp. 83–92.
- [9] G. Csurka, S. Skaff, L. Marchesotti, and C. Saunders, "Building look & feel concept models from color combinations," in *The Visual Computer*, vol. 27, no. 12, 2011, pp. 1039–1053.
- [10] H. Zhang, E. Augilius, T. Honkela, J. Laaksonen, H. Gamper, and H. Alene, "Analyzing emotional semantics of abstract art using low-level image features," in *Advances in Intelligent Data Analysis*, 2011, pp. 413–423.
- [11] D. Joshi, R. Datta, E. Fedorovskaya, Q. T. Luong, J. Z. Wang, J. Li, and J. Luo, "Aesthetics and emotions in images," *IEEE Signal Processing Magazine*, vol. 28, no. 5, pp. 94–115, 2011.
- [12] X. Lu, P. Suryanarayan, R. B. Adams Jr, J. Li, M. G. Newman, and J. Z. Wang, "On shape and the computability of emotions," in *ACM International Conference on Multimedia (MM)*, 2012, pp. 229–238.
- [13] S. Zhao, Y. Gao, X. Jiang, H. Yao, T.-S. Chua, and X. Sun, "Exploring principles-of-art features for image emotion recognition," in *ACM International Conference on Multimedia (MM)*, 2014, pp. 47–56.
- [14] M. Nishiyama, T. Okabe, I. Sato, and Y. Sato, "Aesthetic quality classification of photographs based on color harmony," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 33–40.
- [15] P. O'Donovan, A. Agarwala, and A. Hertzmann, "Color compatibility from large datasets," *ACM Transactions on Graphics (TOG)*, vol. 30, no. 4, pp. 63:1–63:12, 2011.
- [16] L. Marchesotti, F. Perronnin, D. Larlus, and G. Csurka, "Assessing the aesthetic quality of photographs using generic image descriptors," in *IEEE International Conference on Computer Vision (ICCV)*, 2011, pp. 1784–1791.
- [17] N. Murray, L. Marchesotti, and F. Perronnin, "AVA: A large-scale database for aesthetic visual analysis," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 2408–2415.
- [18] U. Hess and O. Gryc, "How shape influence social judgement," *Social Cognition*, vol. 31, no. 1, pp. 72–80, 2013.
- [19] J. R. Averill, "A semantic atlas of emotional concepts," *JSAS: Catalog of Selected Documents in Psychology*, vol. 5, no. 330, pp. 1–64, 1975.
- [20] D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang, "Large-scale visual sentiment ontology and detectors using adjective noun pairs," in *ACM International Conference on Multimedia (MM)*, 2013, pp. 223–232.
- [21] C. Osgood, G. Suci, and P. Tannenbaum, *The measurement of meaning*. Urbana, USA: University of Illinois Press, 1957.
- [22] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "Emotion, motivation, and anxiety: Brain mechanisms and psychophysiology," *Biological Psychiatry*, vol. 44, no. 12, pp. 1248–1263, 1998.
- [23] K. A. Lindquist, T. D. Wager, H. Kober, E. Bliss-Moreau, and L. F. Barrett, "The brain basis of emotion: A meta-analytic review," *Behavioral and Brain Sciences*, vol. 173, no. 4, pp. 1–86, 2011.
- [24] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "Emotion, Attention, and the Startle Reflex," *Psychological Review*, vol. 97, no. 3, pp. 377–395, 1990.
- [25] P. J. Lang, "Behavioral treatment and bio-behavioral assessment: Computer applications," in *In J. Sidowski, J. Johnson, & T. Williams (Eds.), Technology in mental health care delivery systems*. Norwood, NJ: Ablex, 1980, pp. 119–137.
- [26] J. Mikel, B. L. Fredrickson, G. R. Larkin, C. M. Lindberg, S. J. Maglio, and P. A. Reuter-Lorenz, "Emotional category data on images from the international affective picture system," *Behavior Research Methods*, vol. 37, no. 4, pp. 626–630, 2005.
- [27] M. Mills, Introduction to the measurement of roundness, *Taylor-Hobson Precision*, online.
- [28] J. Li, "Agglomerative connectivity constrained clustering for image segmentation," *Statistical Analysis and Data Mining*, vol. 4, no. 1, pp. 84–99, 2011.
- [29] V. Pratt, "Direct least-squares fitting of algebraic surfaces," *SIG-GRAPH Comput. Graph.*, vol. 21, no. 4, pp. 145–152, 1987.
- [30] M. A. Changizi, Q. Zhang, and S. Shimojo, "Bare skin, blood and the evolution of primate colour vision," *Biology Letters*, vol. 2, no. 2, pp. 217–221, 2006.
- [31] S. Rodgers, L. J. Kenix, and E. Thorson, "Stereotypical portrayals of emotionality in news photos," *Mass Communication and Society*, vol. 10, no. 1, pp. 119–138, 2007.
- [32] J. Aronoff, B. A. Woike, and L. M. Hyman, "Which are the stimuli in facial displays of anger and happiness? Configurational bases of emotion recognition," *Journal of Personality and Social Psychology*, vol. 62, no. 6, 1992.