# INTERPRETABLE AESTHETIC FEATURES FOR AFFECTIVE IMAGE CLASSIFICATION

*Xiaohui Wang*[⋆]*, Jia Jia*[⋆]*, Jiaming Yin*[†] *and Lianhong Cai*[⋆]

[⋆] Key Laboratory of Pervasive Computing, Ministry of Education
Tsinghua National Laboratory for Information Science and Technology (TNList)
Department of Computer Science and Technology, Tsinghua University
[†]University of Science and Technology Beijing

## ABSTRACT

Images can not only display contents themselves, but also convey emotions, e.g., excitement, sadness. Affective image classification is useful and hot in many fields such as computer vision and multimedia. Current researches usually consider the relationship model between images and emotions as a black box. They extract the traditional discursive visual features such as SIFT and wavelet textures, and use them directly upon various classification algorithms. However, these visual features are not interpretable, and people cannot know why such a set of features induce a particular emotion. And due to the highly subjective nature of images, the classification accuracies on these visual features are not satisfactory for a long time. We propose the interpretable aesthetic features to describe images inspired by art theories, which are intuitive, discriminative and easily understandable. Affective image classification based on these features can achieve higher accuracy, compared with the state-of-the-art. Specifically, the features can also intuitively explain why an image tends to convey a certain emotion. We also develop an emotion guided image gallery to demonstrate the proposed feature collection.

***Index Terms***— image features, affective classification, interpretability, art theory

## 1. INTRODUCTION

With the rapid development of the Internet, people now more incline to convey and share emotions using images, and pay more attention to emotions behind the visual contents. So understanding images on the affective level is charming and raises more and more concerns of researchers. This is known as affective image classification or more generally affective computing, and is a rather challenging problem as emotions are very subjective due to the subtle and elusory connections with various and complex visual features.

In image processing, visual features play a crucial role in describing images. Researchers proposed lots of low-level visual features such as SIFT and wavelet textures, and used traditional learning tools to bridge the gap between low-level visual features and high-level semantics, e.g., emotions [1, 2]. J. Machajdik summarized more than 100 visual features to perform affective image classification [1]. However, the low-level nature of features means they are generally not interpretable, and people cannot know why such a set of features induce a particular emotion. The features also lack the necessary connections with the artistical feelings, which seriously limits the accuracy on some highly subjective datasets, as demonstrated in Sec. 3.

In recent years, image attributes are proposed to convey semantic information, defined as *"properties observable in images that have human-designated names (e.g., 'four-legged') and they are valuable as a new semantic cue in various problems"* [3]. For the task of affective image classification, the attributes can be seen as a middle layer between low-level visual features and high-level emotional semantics. They are more understandable than visual features and easily interpretable to people. However, attribute categories are usually defined too specifically, e.g., "house", "four-legged". They may be sufficient for some specific classification problems. But for the highly subjective affective classification, the attributes are powerless as emotions are rather subjective and complex. And these attributes usually have nothing to do with affective categories. So we still need a kind of image features that have both great interpretability and strong power of emotional description.

In this paper, we focus on mining the interpretable visual features directly affecting human emotional perception from the view point of art theories. Artists often jointly use figure-ground relationships, color patterns, shapes and their diverse combinations to express emotions in their art creations. Inspired by art theories [4, 5, 6], we propose a set of features (figure-ground relationships, color patterns, shapes and compositions) reflecting how images are related to emotions. These features are more semantic than traditional low-level visual features, more general than image attributes, and more relevant to human emotions. For these features, we design an automatic interpreter telling why an image belongs to a certain affective category. We apply our new features on the affective image classification. Compared with the state-of-the-art [1], the classification performance is greatly improved.

## 2. INTERPRETABLE AESTHETIC FEATURES

Colors and shapes determine what we see, and these elements along with their compositions induce intuitive emotional feelings, as discovered by the famous art theorist and perceptual psychologist Arnheim [4]. In this work, we design a set of features to reveal artistical color patterns, shapes and their relations, as summarized in Table 1. We manually construct a database to assist the interpretation and validate the feature extraction algorithms.

### 2.1. Figure-ground Relationship

The figure-ground relationship refers to the cognitive feasibility to distinguish the foreground (figure) and the background (ground), which is an important concept in art design [4]. To convey a specific emotion, the figure and the ground are harmoniously combined and cooperate with each other. Two types of figure-ground relationships are commonly used, called *figure-ground separation* and *figure-ground harmony* respectively.

We adopt the salient region detection technique to extract the figure from an image [7, 8]. Then a feature vector is computed as the statistics of differences between the figure and the ground, including differences of areas, color and texture complexities. These features describe the contrast between the figure and the ground. The bigger differences an image has, the closer it is to the *figure-ground separation* style.

### 2.2. Color Pattern

The color pattern describes the pattern of overall color distributions of an image, which has direct impact on emotions [4, 6]. *Color combinations* are templates of colors to describe the main color composition. The 5-color theme is commonly used in art design [6]. To make the extracted the color combination consistent with visual perception, we adopt an optimization balancing the area, contrast and position rules inspired by psychological studies [9, 10]. *Saturation and its contrast* describe the brilliant degree of colors and the differences in an image (e.g., high saturation makes people feel fresh). *Brightness and its contrast* illustrate the black-while degree and the differences (e.g., low brightness makes people feel negative and deep). *Warm or cool color* is defined based on human feelings stimulated by colors (e.g., warm colors like red, yellow can arouse excitement, and cool colors like blue, green and purple make people calm). *Clear or dull color* depicts the clear or dull feeling and is determined by both saturation and brightness of colors.

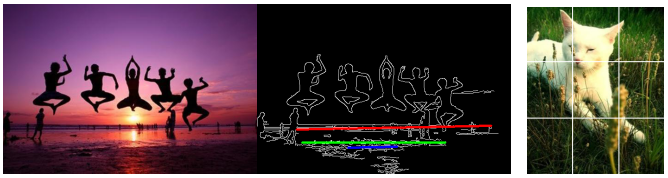
### 2.3. Shape

Different shapes usually relate to different thoughts and feelings [4]. For example, squares make people feel regular and intensive; circular shapes arouse smooth and relaxed feelings, and V or S shapes relate to unstable or lively feelings respectively. Here the shapes do not strictly follow the geometric rules, but a broad and fuzzy notion. We extract the salient region and compute the shape similarities with these reference shapes by shape context [11].

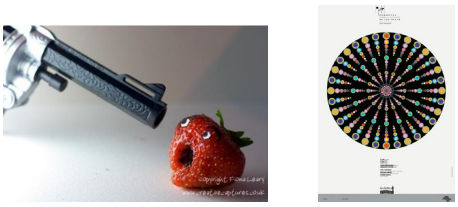
### 2.4. Composition

Composition is the distribution and mutual combination of image elements and prominent lines, which is essential in image creation such as photographic images and paintings [4, 12]. Symmetrical composition relates to order, and unbalanced composition tends to express unusual feelings. Such relations are rather complex [1], and we analyze some simple and intuitional ones having more directly impacts on the feelings.

We extract multiple salient regions as *image elements* [7] and detect *prominent lines* [12]. The details of computation can be found in Table 1. The scores are defined as the negative exponent of the distances, and normalized to $[0, 1]$. Some examples are shown in Fig. 1.



(a) Prominent line detection (b) Rule of Thirds: 0.86



(c) Diagonal dominance: 0.96 (d) Symmetry: 0.89 (e) Visual balance: 0.88, 0.99, 0.68

**Fig. 1**. Examples of composition. (b)-(e) are composition scores.

### 2.5. Feature database

We construct a database including 500 images belonging to seven categories (photograph, poster, furniture, landscape, clothes, shows and paintings). This is a high category coverage from the aspect of art design. Each image is labeled with the features in Table 1 by five students from Academy of Art&Design. For convenience, each feature is associated with discrete levels, e.g., brightness is divided into high, middle and low levels. For each image and each feature, we manually label its level and may drop a feature for an image if no level is suitable. The database supports two purposes.

**Table 1**. Summary of interpretable aesthetic features. The column '#' indicates the dimension of each feature.

| Type | Name | # | Short description |
|---|---|---|---|
| Figure-ground relationship | Area difference | 1 | area ratio difference between the figure and the ground |
| | Color difference | 1 | Euclidean distance of mean colors between the figure and the ground |
| | Texture complexity | 2 | density of Canny edges in the figure and the ground respectively |
| Color Pattern | Five-color combination | 15 | five dominant colors in the HSV color space |
| | Saturation and its contrast | 2 | mean saturation and average contrast of saturation |
| | Brightness and its contrast | 2 | mean brightness and average contrast of brightness |
| | Warm or cool color | 1 | ratio of cool colors with hue ([0-360]) in the HSV space between 30 and 110 |
| | Clear or dull color | 1 | ratio of colors with brightness ([0-1]) greater than 0.7 |
| Shape | Shape match | 14 | shape context scores between the largest salient region and reference shapes: the square, rectangles with two different orientations, triangles with five different shapes, the five-pointed star, the rhombus, the trapezoid, the circular, the line and the free-form curve |
| Composition | Rule of Thirds (RT) | 1 | distance between salient regions and power points in terms of rule of thirds |
| | Diagonal dominance | 1 | distance between prominent lines and the two diagonals |
| | Symmetry | 1 | sum of brightness differences between each pixel and its symmetric pixel about the central line |
| | Visual balance | 3 | distances of the barycenter of the biggest salient region from the image midpoint, the vertical central line and the horizontal central line, respectively |

- **Interpretation assistant.** As the extracted features are continuous, we need the thresholds for verbalization (e.g., high, low). The user-labeled levels help to find suitable thresholds (having the highest accuracy).
- **Extraction algorithms tuning.** Some of feature extraction algorithms (saliency/prominent line detection, etc.) need some magic parameters to work. Along with previous thresholds, they turn to be an optimization problem over the labeled database. We simply adopt a genetic algorithm.

## 3. AFFECTIVE IMAGE CLASSIFICATION

To validate the effectiveness of our features, we use them for affective image classification. We adopt eight emotional categories, *amusement, anger, awe, contentment, disgust, excitement, fear* and *sad* [1] on two public data sets: (1) *ARTphoto*: a set of 806 artistic photos and (2) *ABSTRACT*: a set of 228 abstract paintings, which are the same as those in [1].

The experimental setup is also the same as [1], with similar generic-based feature selection and K-fold cross validation ($K = 5$). See [9, 11, 12] for the parameter settings of the aesthetic features extraction. Fig. 2 shows the averaged true positive rate per class of the two data sets by our method and [1], indicating a significant accuracy improvement. Note that our performance on the ABSTRACT dataset is much better, 18% higher averagely and 22% higher for the *fear* category. Our features are mainly aesthetics based and have much stronger artistic evidences and nuanced connections with human feelings, as revealed by various art theories. So they are more appropriate to depict emotions, especially on highly subjective abstract paintings.

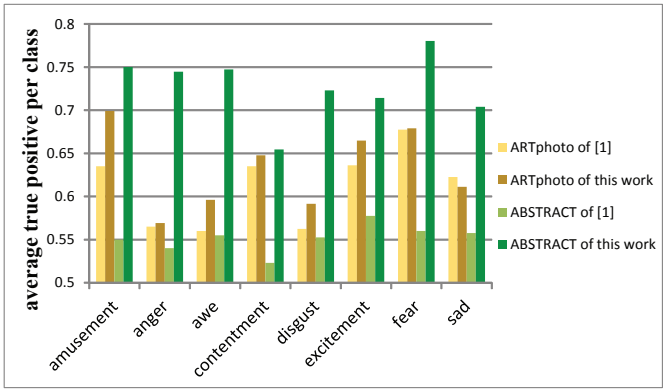



**Fig. 2**. Classification performance of the two data sets in this work for each class compared with the best features in [1].

Besides, compared with the traditional visual features such as wavelet textures, our features are more suitable for interpretation, as every dimension in the feature space has an explicit meaning. For example, in the *amusement* category, most images have high brightness, warm colors, smooth and soft shapes like circular and balanced composition. In contrast, in the *fear* category, most images have low brightness and saturation, cool colors, sharp shapes and cluttered composition. Table 2 gives some examples. The interpretations are automatically given by choosing the most significant feature responses, and adopting the thresholds as in Sec. 2.5.

## 4. IMAGE GALLERY

We develop an [1]emotion guided image gallery as shown in Fig. 3. The gallery contains about 10,000 images organized according to features in Table 1. For each image and each feature, the algorithm automatically calculates its level ( see Sec. 2.5). For example, the shape types include square, circle, X-shaped, S-shaped, V-shaped, L-shaped, etc. We assign an image to a feature (or a feature level) if the response is strong enough (according to the user labels as in Sec. 2.5). And the users can explore by individual aesthetic types. Because these aspects directly relate to emotions, the gallery is called *emotion guided*. Experiments show that the gallery can provide inspirations to creators of paintings and graphic design. If combined with various image search techniques, it can be much more useful to art design.

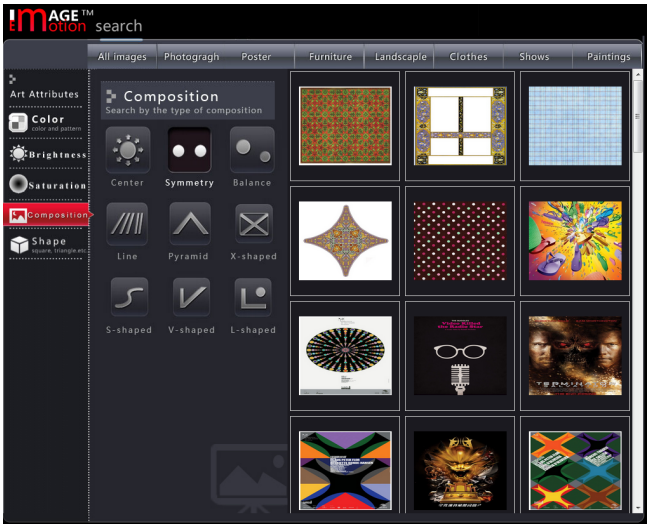



**Fig. 3**. Emotion guided image gallery. The frame corresponds to the symmetry composition.

## 5. CONCLUSIONS

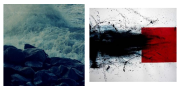
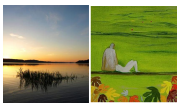
In this paper, we propose the interpretable aesthetic features, which are more interpretable and more suitable for affective prediction. As confirmed by our experiments, we can achieve significantly higher prediction accuracy on the public datasets. And the interpretability also makes our features more practical. Currently, our interpretation is simple and direct, and more sophisticated or even NLP related schemas can be used to better improve the interpretation quality. Another direct extension is to pay more attention to personalized affective prediction, as it is sometimes a highly subjective issue. We may use some types of transfer learning to make a general model better satisfying individual requirements.

[1]http://hcsi.cs.tsinghua.edu.cn/Demo/imageSearch/main.php

## 6. ACKNOWLEDGEMENT

This work is supported by the National Basic Research Program (973 Program) of China (2011CB302201), partially supported by the National High Technology Research and Development Program ("863"Program) of China (2012AA011602), and partially funded by Microsoft Research Asia-Tsinghua University Joint Laboratory.

**Table 2**. Image classification and interpretations

| Images & Category | Interpretation |
|---|---|
|  amusement | high brightness, middle saturation, warm color, circular shape (smooth and soft), very significant visual balance |
|  anger | low brightness, curve shape (exaggerated) |
|  awe | low saturation, high figure-ground color difference, trapezoid shape (regular), middle visual balance |
|  contentment | middle brightness, low saturation contrast, middle visual balance |
|  disgust | middle brightness, cool color, high texture complexity (make people feel uncomfortable) |
|  excitement | high brightness, very high saturation, low saturation contrast |
|  fear | low saturation, cool color, dull color, low color difference between figure and ground, cluttered composition |
|  sad | middle brightness, low saturation, low saturation contrast, cool color, square shape (regular), line shape (regular), very high RT, symmetry |

## 7. REFERENCES

[1] J. Machajdik and A. Hanbury, "Affective image classification using features inspired by psychology and art theory," in *Proceedings of the international conference on Multimedia*. ACM, 2010, pp. 83–92.

[2] Q. F. Wu, C. L. Zhou, and C. N Wang, "Content-based affective image classification and retrieval using support vector machines," , no. 3784, pp. 239–247, 2005.

[3] D. Parikh and K. Grauman, "Relative attributes," in *IEEE International Conference on Computer Vision*. IEEE, 2011, pp. 503–510.

[4] R. Arnheim, *Art and visual perception: A psychology of the creative eye*, Univ of California Press, 1954.

[5] J. Itten, *The art of color: the subjective experience and objective rationale of color*, Wiley, 1974.

[6] S. Kobayashi, *Art of Color Combinations*, K. International, 1995.

[7] M. M. Cheng, G. X. Zhang, N. J. Mitra, X. L. Huang, and S. M. Hu, "Global contrast based salient region detection," in *IEEE Conference on Computer Vision and Pattern Recognition*. 2011, pp. 409–416, IEEE.

[8] S. M. Hu, T. Chen, K. Xu, M. M. Cheng, and R. R. Martin, "Internet visual media processing: a survey with graphics and vision applications," *The Visual Computer*, vol. 29, no. 5, pp. 393–405, 2013.

[9] X. H. Wang, J. Jia, and L. H. Cai, "Affective image adjustment with a single word," *The Visual Computer*, 2012.

[10] X. H. Wang, J. Jia, H. Y. Liao, and L. H. Cai, "Affective image colorization," *Journal of Computer Science and Technology*, vol. 27, no. 6, pp. 1119–1128, 2012.

[11] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, 2002.

[12] L. G. Liu, R. J. Chen, L. Wolf, and D. Cohen-Or, "Optimizing photo composition," *Computer Graphics Forum*, vol. 29, no. 2, pp. 469–478, 2010.