Exploring Principles-of-Art Features For Image Emotion Recognition

Sicheng Zhao[†], Yue Gao[‡], Xiaolei Jiang[†], Hongxun Yao[†], Tat-Seng Chua[‡], Xiaoshuai Sun[†]

[†]School of Computer Science and Technology, Harbin Institute of Technology, China.
[‡]School of Computing, National University of Singapore, Singapore.
{zsc, xljiang, h.yao, xiaoshuaisun}@hit.edu.cn; {dcsgaoy,dcscts}@nus.edu.sg

ABSTRACT

Emotions can be evoked in humans by images. Most previous works on image emotion analysis mainly used the elements-of-art-based low-level visual features. However, these features are vulnerable and not invariant to the different arrangements of elements. In this paper, we investigate the concept of principles-of-art and its influence on image emotions. Principles-of-art-based emotion features (PAEF) are extracted to classify and score image emotions for understanding the relationship between artistic principles and emotions. PAEF are the unified combination of representation features derived from different principles, including *balance*, *emphasis*, *harmony*, *variety*, *gradation*, and *movement*. Experiments on the International Affective Picture System (IAPS), a set of artistic photography and a set of peer rated abstract paintings, demonstrate the superiority of PAEF for affective image classification and regression (with about 5% improvement on classification accuracy and 0.2 decrease in mean squared error), as compared to the state-of-the-art approaches. We then utilize PAEF to analyze the emotions of master paintings, with promising results.

Categories and Subject Descriptors

H.3.1 [**Information storage and retrieval**]: Content Analysis and Indexing; I.4.7 [**Image processing and computer vision**]: Feature Measurement; J.5 [**Computer Applications**]: Arts and Humanities

General Terms

Algorithms, Human Factors, Experimentation, Performance

Keywords

Image Emotion; Affective Image Classification; Image Features; Art Theory; Principles of Art

1. INTRODUCTION

Humans are able to perceive and understand images only at high level semantics (including cognitive level and affective level [10]), rather than at low level visual features. Most previous works on image content analysis focus on understanding the cognitive aspects of images, such as object detection and recognition. Little research effort has been dedicated to the understanding of images at the affective level, due to the subjective evaluation on emotions and the "affective gap", which can be defined as "the lack of coincidence between the measurable signal properties, commonly referred to as features, and the expected affective state in which the user is brought by perceiving the signal" ([10], p. 91). However, with the increasing use of digital photography technology by the public and users' high requirement for image understanding, the analysis of image content at higher semantic levels, in particular the affective level, is becoming increasingly important.

For affective level analysis, how to extract emotion related features is the key problem. Most existing works target low level visual features based on the elements-of-art, such as *color*, *texture*, *shape*, *line*, *etc*. Obviously, these features are not invariant to their different arrangements and their link to emotions is weak, while different element arrangements share different meanings and evoke different emotions. Therefore, elements must be carefully arranged and orchestrated into meaningful regions and images to describe specific semantics and emotions. The rules, tools or guidelines of arranging and orchestrating the elements-of-art in an artwork are known as the principles-of-art, which consider various artistic aspects including *balance*, *emphasis*, *harmony*, *variety*, *gradation*, *movement*, *rhythm*, and *proportion* [6, 12]. Different combinations of these principles can evoke different emotions. For example, symmetric and harmonious images tend to express positive emotions, while images with strong color contrast may evoke negative emotions [31] (see Section 5.2). Further, the artistic principles are more interpretable by humans than elements [5].

Inspired by these observations, we propose to study, formulate, and implement the principles-of-art systematically, based on the related art theory and computer vision research. After quantizing each principle, we combine them together to construct image emotion features. Different from previous low level visual features, PAEF take the arrangements and orchestrations of different elements into account, and it can be used to classify and score image emotions evoked in humans. The framework of our proposed method is shown in Figure 1. We then apply the proposed PAEF to predict the emotions implied in famous artworks to capture the masters' emotional status.

The rest of this paper is organized as follows. Section 2 introduces related work on affective content analysis, aesthetics, composition and photo quality assessment. We summarize the elements-of-art-based low level emotion features (EAEF) and their limitations in emotion prediction as a preliminary in Section 3. The proposed PAEF are described in Section 4. Experimental evaluation, analysis and applications are presented in Section 5, followed by conclusion and future work in Section 6.

MM '14, November 03 - 07 2014, Orlando, FL, USA.

http://dx.doi.org/10.1145/2647868.2654930.

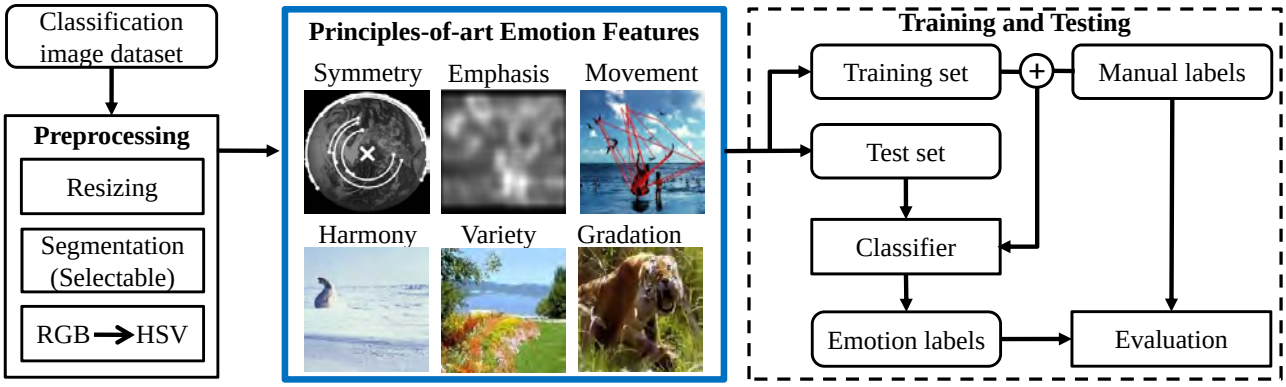


Figure 1: The framework of our proposed method. The main contributions, principles-of-art-based emotion features (PAEF), lie in the central feature extraction part in blue solid rectangle.

Table 1: Related works on affective content analysis

Category	Classification		Publications
Data type	Still images		[26, 38, 42, 24, 20, 15, 29, 33]
	Dynamic videos		[10, 17, 41, 43, 39, 3, 13]
Emotion model	Categorical		[26, 38, 42, 43, 24, 20, 15, 17, 39, 13]
	Dimensional		[10, 24, 29, 41, 3, 33]
Features	Generality	Generic	[42, 29, 4]
		Specific	[10, 26, 38, 43, 24, 20, 15, 17, 41, 39, 3, 13, 33]
	Level	Low	[10, 26, 38, 42, 24, 20, 15, 29, 17, 41, 39, 3]
		Mid	[26, 13, 33]
		High	[26, 4, 43]
	Art theory	Elements	[26, 38, 42, 24, 20, 15, 33]
		Principles	[26]

2. RELATED WORK

Affective content analysis. Some research efforts have been made recently to improve the accuracy of affective understanding in images and videos. Table 1 presents the related works, which can be divided into different types, according to the analyzed multimedia type, the adopted emotion model and the extracted features.

Generally, there are two typical models to represent emotions: categorical emotion states (CES), and dimensional emotion space (DES). CES methods [26, 38, 42, 24, 20, 15, 17, 39, 13] consider emotions to be one of a few basic categories, such as *fear*, *contentment*, *sadness*, *etc*. DES methods mostly employ the 3-D valence-arousal-control emotion space [32], 3-D natural-temporal-energetic connotative space [3], 3-D activity-weight-heat emotion factors [33], and 2-D valence-arousal emotion space [10, 24, 29, 41] for affective representation and modeling. CES in the classification task is easier for users to understand and label, while DES in the regression task is more flexible and richer in the descriptive power. Similar to [26, 42], we adopt CES to classify emotions into eight categories defined in a rigorous psychological study [27], including *anger*, *disgust*, *fear*, *sadness* as negative emotions, and *amusement*, *awe*, *contentment*, *excitement* as positive emotions. We also use valence-arousal DES to predict the scores of image emotions as in [24].

From a feature's view point, most methods extract low level visual and audio features. Lu *et al*. [24] investigated the computability of emotion through *shape* features. Machajdik and Hanbury [26] exploited theoretical and empirical concepts from psychology and art theory to extract image features that are specific to the domain of artworks. In their method, *color* and *texture* are used as low level features, *composition* are used as mid level features, while image *semantic content* including human faces and skin are used as high level features. Besides color features, Jia *et al*. [15] also extracted social correlation features for social network images. Solli and Lenz [33] classified emotions using emotion-histogram features and bag-of-emotion features derived for patches surrounding each interest point. Irie *et al*. [13] extracted mid level features based on affective audio-visual words and proposed a latent topic driving model for video classification task. Borth *et al*. proposed to infer emotions based on the understanding of visual concepts [4]. A large-scale visual sentiment ontology composed of adjective noun pairs (ANPs) is constructed and SentiBank is proposed to detect the presence of ANPs. Popular features in previous works on image emotion analysis are mainly based on elements-of-art, such as *color*, *texture*, *shape*, *etc*. Machajdik and Hanbury [26] extracted composition features, some of which can be considered as principles. However, there is still no systematic study on the use of principles-of-art for image emotion analysis.

Aesthetics, composition and photo quality assessment. Aesthetics, composition in images and the quality of photos are strongly related to humans' emotions. Joshi *et al*. [16] and Datta *et al*. [7] discussed key aspects of the problem of computational inference of aesthetics and emotions from natural images. Liu *et al*. [21] evaluated the composition aesthetics of a given image based on measuring composition guidelines and changed the relative position of salient regions using a compound operator of crop-and-retarget. Aesthetics and interestingness are predicted through high level describable attributes, including compositional, content and sky-illumination attributes [8]. Compositional features are also exploited for scene recognition [30] and category-level image classification [37]. Based on professional photography techniques, Luo and Tang [25] extracted the subject region from a photo and formulated a number of high-level semantic features based on this subject and background division. Sun *et al*. [35] presented a computational visual attention model to assess photos by using the rate of focused attention. In this work, we expand related research on computer vision and multimedia to measure the artistic principles for affective image classification and score prediction.

3. ELEMENTS OF ART: A PRELIMINARY

Low level features extracted for emotion recognition are mostly based on the elements-of-art (EAEF), including *color*, *value*, *line*, *texture*, *shape*, *form* and *space* [12], as shown in Figure 2. In this

section, we briefly introduce EAEF and their limitations in image emotion prediction.

3.1 Elements-of-art-based Low Level Features

Color. An element of art which has three properties: hue, intensity, and value, representing the name, brightness and lightness or darkness of a color. Color is often used effectively by artists to induce emotional effects, such as *saturation*, *brightness*, *hue*, and *colorfulness* [26, 38, 20, 15].

Value. An element of art that describes the lightness or darkness of a color. Value is usually found to be an important element in works of art. This is true with drawings, prints, photographs, most sculpture, and architecture. The description of *lightness* or *darkness* is often used as value features [26, 38].

Line. An element of art which is a continuous mark made on some surface by a moving point. There are mainly two types of lines, emphasizing lines and de-emphasizing lines. Emphasizing lines, better known as contour lines, show and outline the edges or contours of an object. When artists stress contours or outlines in their work, the pieces are usually described as lines. Not all artists emphasize lines in their works. Some even try to hide the outline of objects in their works. De-emphasizing lines are used to describe works that do not stress the contours or outlines of objects.

Lines can be used to suggest movement in some direction. They are also used in certain ways to give people different feelings. For example, horizontal lines suggest calmness and usually make people feel relaxed, vertical lines suggest strength and stability, diagonal lines suggest tension, and curved lines suggest flowing movement [26]. Usually, the *amounts* and *lengths* of *static* and *dynamic* lines are calculated by Hough transform to describe lines [26].

Texture. An element of art which is used to describe the surface quality of one object. It refers to how things feel, or look as if they might feel if you were able to touch it. Some artists paint carefully to give their paintings a smooth appeal, while others use a lot of paint to produce a rough texture. The most frequently used texture features are *wavelet-based* features, *Tamura* features, *gray-level co-occurrence matrix* [26, 20] and *LBP* features.

Shape and *Form*. Shape is flat and has only 2 dimensions, height and width. The descriptions of *roundness*, *angularity*, *simplicity*, and *complexity* are used as shape features [24]. Form is 3 dimensional with height, width and depth, thus having mass and volume.

Space. An element of art which refers to the distance or area between, around, above, below or within things.

3.2 Limitations of EAEF

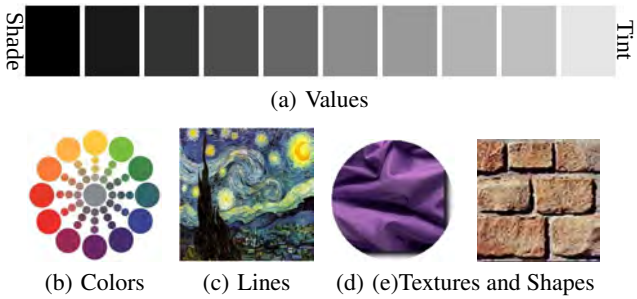
These elements-of-art based low level visual features are easy to extract based on current computer vision and multimedia research. However, there are several disadvantages using them to model image emotions:

(1) **Weak link to emotions** [1, 26]. EAEF suffer from the greatest "affective gap" and are vulnerable and not invariant to the different arrangements of elements, resulting in the poor performance on image emotion recognition. These low level features cannot represent high level emotions well.

(2) **Not interpretable by humans** [1]. As EAEF are extracted from low level view point. Humans cannot understand the meanings of these features and why such a set of features induce a particular emotion.

4. PROPOSED EMOTION FEATURES

In this section, we systematically study and formulize 6 artistic principles. For each principle, we first explain the concepts and meanings, under the art theory in [6, 12], and then translate these concepts into mathematical formulae for quantization measurement. As rhythm and proportion are ambiguously defined, we do not take them into account here.



(a) Values

(b) Colors (c) Lines (d) (e)Textures and Shapes

Figure 2: Low-level representation features of emotions based on elements-of-art.

4.1 Balance

Balance (symmetry) refers to the feeling of equilibrium or stability of an art work. The artists arrange balance to set the dynamics of a composition. There are three types of balances: symmetrical, asymmetrical and radial. Symmetrical balance is the most visually stable, and characterized by an exact or nearly exact compositional design on both sides of the horizontal, vertical or any axis of the picture plane. If the two halves of an image are identical or very similar, it is symmetrical balance. Asymmetry uses compositional elements that are offset from each other, creating a visually unstable balance. Asymmetrical balance is the most dynamic because it creates a more complex design construction. Radial balance refers to balance within a circular shape or object, offering stability and a point of focus at the center of the composition [6, 12].

Since the asymmetrical balance is difficult to measure mathematically, in this paper we only consider symmetry, including bilateral symmetry, rotational symmetry [22] and radial symmetry [23, 28]. Symmetry can be seen as the reverse measurement of asymmetry.

To detect bilateral and rotational symmetry, we use the symmetry detection method in [22], which is based on matching symmetrical pairs of feature points. The method for determining feature points should be rotationally invariant, so SIFT is a good choice, although scale-invariance is not necessary. Each feature can be represented by a point vector describing its location in x, y coordinates, its orientation and (optionally) scale. Every pair of feature points is a potential candidate for a symmetrical pair. In the case of bilateral symmetry, each pair of matching points defines a potential axis of symmetry passing perpendicularly through the mid-point of the line joining these two points. Unlike bilateral symmetry detection, detecting rotational symmetry does not require the development of additional feature descriptors. It can be simply detected by matching the features against each other. Given a pair of non-parallel feature point vectors, there exists a point about which feature vector can be rotated to precisely align with another feature vector. The Hough transform [2] is used to find dominant symmetry axes or centers. Each potential symmetrical pair casts a vote in Hough space weighted by their symmetry magnitude. The rotational symmetry magnitude may be set to unity, while the bilateral symmetry magnitude may involve the discrepancy between the orientation of one feature point and the mirrored orientation of another feature point. Finally the symmetries exhibited by all individual pairs in a voting space are accumulated to determine the dominant symmetries present in the image. The result is blurred with a Gaussian and

Figure 3: Symmetrical gray scale images. The first row is images in bilateral symmetry with symmetry axis and symmetrical feature points. The second row is images in rotational symmetry with symmetry center and symmetrical feature points.

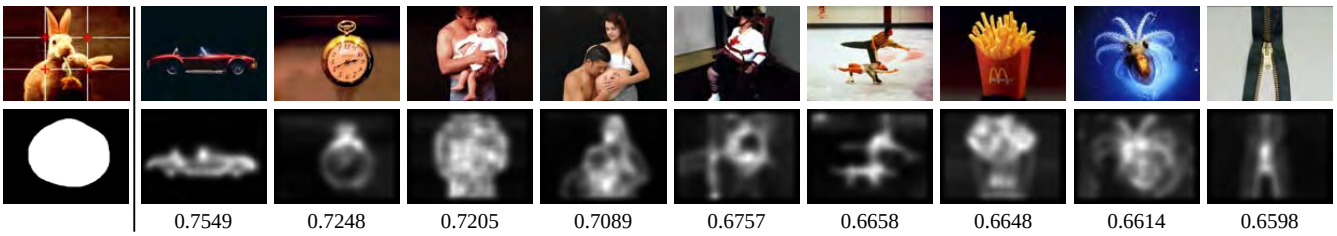


Figure 4: Images with *high* RFA based on statistic subject mask in [35]. The first column is "Rule of the third" composition and this mask. The three rows on the right of the black line are related images, saliency maps and RFA scores.

the maxima are identified as dominant axes of bilateral symmetry or centres of rotational symmetry. We compute symmetry number, radius, angle and strength of the maximum symmetry for bilateral symmetry, symmetry number, center and strength of the maximum symmetry for rotational symmetry, as shown in Figure 3.

Based on the symmetry detection method in [23], we compute the distribution of symmetry map after radial symmetry transformation for radial symmetry (see Section 4.4).

4.2 Emphasis

Emphasis, also known as contrast, is used to stress the difference of certain elements. It can be accomplished by using sudden and abrupt changes in elements. Emphasis is usually used to direct and focus viewers' attention to the most important area or centers of interests of a design, because it catches your attention [12]. We adopt Itten's color contrasts [14] and Sun's rate of focused attention (RFA) [35] to measure the principle of emphasis.

Itten defined and identified strategies for successful color combinations [14]. Seven methodologies were devised to coordinate colors using hue's contrasting properties. Itten contrasts include contrast of saturation, light and dark, extension, complements, hue, warm and cold and the simultaneous contrast. We calculate six color contrasts by the mathematical expressions in [26] and represent the contrast of extension as the standard deviation of the pixel amount of 11 basic colors as in Section 4.4.

RFA was proposed to measure the focus rate of an image when people watch it [35]. FRA is defined as the attention focus on some predefined aesthetic templates or some statistical distributions according to image's saliency map. Here we adopt Sun's response map method [34] to estimate saliency. Besides the statistic subject mask coincidence with "Rule of the third" composition method, defined in [35], we use another two diagonal aesthetic templates [21]. A 3 dimensional RFA vector is obtained by computing,

$$RFA(i) = \frac{\sum_{x=1}^{Wid}\sum_{y=1}^{Hei} Saliency(x,y)Mask_i(x,y)}{\sum_{x=1}^{Wid}\sum_{y=1}^{Hei} Saliency(x,y)}, \quad (1)$$

where Wid and Hei denote the width and height of image I, while $Saliency(x,y)$ and $Mask_i(x,y)$ are the saliency value and mask value at pixel (x,y), respectively. In Eq. (1), $i = 1, 2, 3$, representing different aesthetic templates. Illustrations of different masks are shown in Figure 4, 5, and 6, together with related images, saliency maps and RFA scores.

4.3 Harmony

Harmony, also known as unity, refers to a way of combining similar elements (such as *line*, *shape*, *color*, *texture*) in an artwork to accent their similarities. It can be accomplished by using repetition and gradual changes when the components of an image are perceived as harmonious. Pieces that are in harmony give the work a sense of completion and have an overall uniform appearance [12].

Inspired by Kass' idea of smoothed filters for local histogram [18], we compute the harmony intensity of each pixel on its hue and gradient direction in a neighborhood. We divide the circular hue or gradient direction equally into eight parts, which are separated into two adjacent groups $c = \{i_1, i_2, \ldots, i_k | 0 \le i_j \le 7, j = 1, 2, \ldots, k\}$ and $I \backslash c$ (see Figure 7(a)), where $i_{k+1} \equiv i_k + 1 (mod 8)$, $I = \{0, 1, \ldots, 7\}$. The harmony intensity at pixel $p(x,y)$ is defined as

$$H(x,y) = \min_c e^{-|h^m(c) - h^m(I\backslash c)|} |i^m(c) - i^m(I \setminus c)|, \quad (2)$$

where

$$\begin{aligned} h^m(c) &= \max_{i \in c} h^i(c) \\ i^m(c) &= \arg\max_{i \in c} h^i(c), \end{aligned} \quad (3)$$

where $h_i(c)$ is the hue or gradient direction in groups c. The harmony intensity of the whole image is the sum of all pixels' harmony intensity, that is

$$H = \sum_{(x,y)} H(x,y). \quad (4)$$

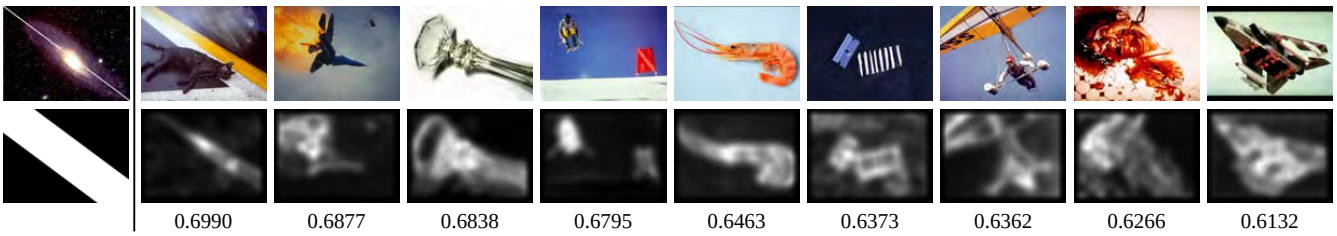


Figure 5: Images with *high* RFA based on diagonal mask [21], shown in the first column. The three rows on the right of the black line are related images, saliency maps and RFA scores.

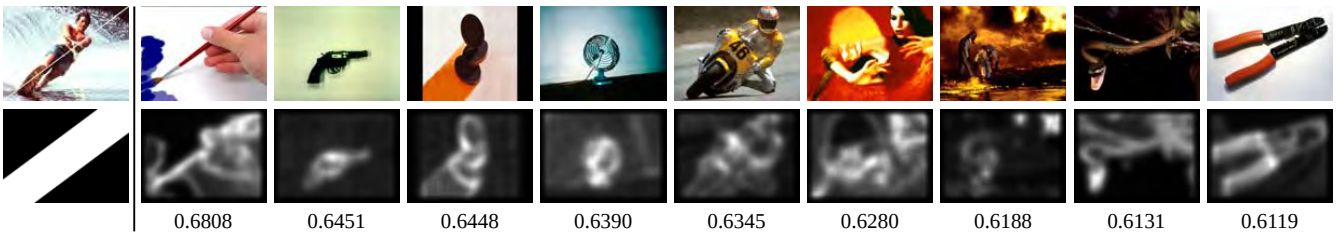


Figure 6: Images with *high* RFA based on back diagonal mask [21], shown in the first column. The three rows on the right of the black line are related images, saliency maps and RFA scores.

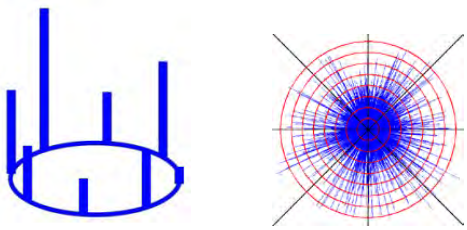


Figure 7: (a) Local histogram on eight equal parts. (b) The gradient distribution of '1300.jpg' in IAPS on red channel.

Figure 10: Images of different texture gradations, but with similar content meanings and emotions.

4.4 Variety

Variety is used to create complicated images by combining different elements. A picture made up of many different *hues*, *values*, *lines*, *textures*, or *shapes* would be described as a complex picture, which increases humans' visual interestingness [12].

However, harmony and variety are not opposites. A careful blend of the two principles is essential to the success of almost any work of art. Artists who focus only on harmony and ignore variety might find it easier to achieve balance and unity; but the visual interest in the piece could be lost. On the other hand, artists who focus only on variety and not harmony would make their works too complex; and consequently, the overall unity of the piece could be lost, which makes viewers confused [12].

Each color has a special meaning and is used in certain ways by artists. We count how many basic colors (black, blue, brown, green, gray, orange, pink, purple, red, white, and yellow) are present and the pixel amount of each color using the algorithm proposed by Weijer *et al*. [36]. Image examples of different color variety are shown in Figure 8.

Gradient depicts the changes of values and directions of pixels in an image. We calculate the distribution of gradient statistically (Figure 7(b)). For directions, we count the number of pixels in the eight regions equally divided of the circle. For lengths, we divide the relative maximum length (RML) into 8 parts equally, by computing RML as,

$$RML = \mu + 5\sigma, \tag{5}$$

where μ and σ are respectively the mean and standard deviation of the gradient matrix.

4.5 Gradation

Gradation refers to a way of combining elements by using a series of gradual changes. For example, gradation may be a gradual change of a dark value to a light value [12].

We adopt the concepts of pixel-wise *windowed total variation* and *windowed inherent variation* proposed by Xu *et al*. [40] and their combination to measure gradation for each pixel. The *windowed total variation* for pixel $p(x, y)$ in image I is defined as

$$D_x(p) = \sum_q g_{p,q}|(\partial_x I)_q|,\ D_y(p) = \sum_q g_{p,q}|(\partial_y I)_q|, \tag{6}$$

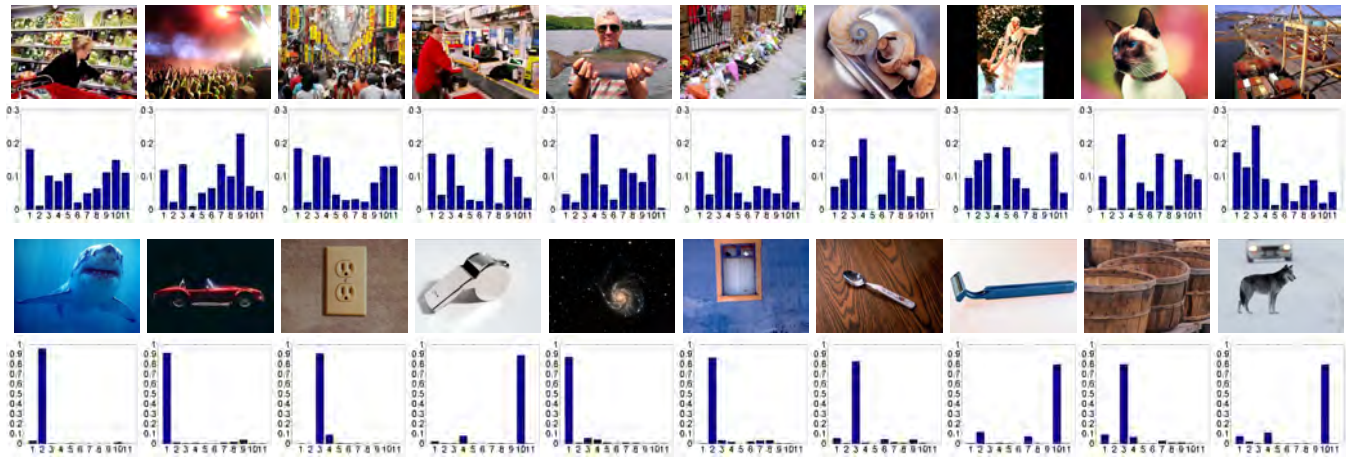


Figure 8: Images of different color variety. The first row shows images of *high* color variety with their color distributions in terms of 11 basic colors shown in row 2. The third and fourth rows respectively show images of *low* color variety and related distributions of the 11 basic colors.

Figure 9: Eye scan path for measuring the principle of movement.

where $q \in R(p)$, $R(q)$ is a rectangular region centered at p, $D_x(p)$ and $D_y(p)$ are windowed total variations in the x and y directions for pixel p, which count the absolute spatial difference within the window $R(q)$. $g_{p,q}$ is a weighting function

$$g_{p,q} = exp\left(-\frac{(x_p - x_q)^2 + (y_p - y_q)^2}{2\sigma^2}\right), \quad (7)$$

where σ controls the spatial scale of the window.

The *windowed inherent variation* for pixel $p(x, y)$ in image I is defined as

$$L_x(p) = |\sum_q g_{p,q}(\partial_x I)_q|,\ L_y(p) = |\sum_q g_{p,q}(\partial_y I)_q|. \quad (8)$$

Different from $D_x(p)$ and $D_y(p)$, $L_x(p)$ and $L_y(p)$ capture the overall spatial variation, without incorporating modules.

It has been proven that in the *relative total variation* (RTV) defined in Equ. (9), opposite gradients in a window cancel out each other (Figure 10), regardless whether the pattern is isotropic or not. We adopt the sum of RTV, the sum of *windowed total variation* and the sum of *windowed inherent variation* to measure an image's relative gradation and absolute gradation, respectively.

$$RG = \sum_p RTV(p) = \sum_p \left(\frac{D_x(p)}{L_x(p) + \varepsilon} + \frac{D_y(p)}{L_y(p) + \varepsilon}\right), \quad (9)$$

$$AGT_x = \sum_p D_x(p), AGT_y = \sum_p D_y(p), \quad (10)$$

$$AGI_x = \sum_p L_x(p), AGI_y = \sum_p L_y(p). \quad (11)$$

4.6 Movement

Movement is used to create the look and feel of action. It guides and moves the viewers' eye throughout the work of art. Movement is achieved through placement of elements so that the eye follows a certain path, like the curve of a line, the contours of shapes, or the repetition of certain colors, textures, or shapes [12].

Based on Super Gaussian Component analysis, Sun *et al*. [34] obtained a response map by filtering the original image and adopted the winner-takes-all (WTA) principle to select and locate the simulated fixation point and estimate a saliency map. We calculate the distribution of eye scan path obtained using Sun's method (see Figure 9).

4.7 Application to Emotion Classification and Score Prediction

From the above six subsections describing the measurements for each principle, we can see that: (1) PAEF are more interpretable and semantic than EAEF; and are easier for humans to understand. For example, humans can understand *symmetry* and *variety* better than *texture* and *line*. (2) PAEF take the arrangements and orchestrations of elements into consideration and are more relevant to image emotions and more robust in image emotion prediction, as demonstrated in Sections 5.2 and 5.3.

We then apply the proposed PAEF to image emotion classification and score prediction. Firstly, we combine the representation of the six principles into one feature vector consistently. The dimensions of these principles are 60, 18, 2, 60, 9 and 16 respectively. The measurements for each principle are summarized in Table 2. Secondly, we adopt Support Vector Machine (SVM) and Support Vector Regression (SVR) both with radial basis function (RBF) kernel to classify categorial emotions and predict dimensional emotion

Table 2: Summary of the measurements for principles of art. '#' indicates the dimension of each measurement.

Principles	Measurement	#	Short Description
Balance	*Bilateral symmetry*	12	Symmetry number, Maximum symmetry radius, angle and strength
	Rotational symmetry	12	Symmetry number, Maximum symmetry center (x and y), strength
	Radial symmetry	36	Distribution of symmetry map after radial symmetry transformation
Emphasis	*Itten color contrast*	15	Average contrast of saturation, contrast of light and dark, contrast of extension, contrast of complements, contrast of hue, contrast of warm and cold, simultaneous contrast
	RFA	3	Rate of focused attention based on saliency map and subject mask
Harmony	*Rangeability of hue and gradient direction*	2	The first and second maximums of local maximum hues and gradient directions in relative histograms of an image patch, and their differences; the combination of all patches of an image
Variety	*Color names*	12	Color types of black, blue, brown, gray, green, orange, pink, purple, red, white, yellow and each color's amount
	Distribution of gradient	48	The distribution of gradient on eight scales of direction and eight scales of length
Gradation	*Absolute and relative variation*	9	Pixel-wise windowed total variation, windowed inherent variation in x and y direction respectively, and relative total variation
Movement	*Gaze scan path*	16	The distribution of gaze vector

scores, respectively. We use the LIBSVM[1] to conduct the emotion classification and score prediction task.

5. EXPERIMENTS

To evaluate the effectiveness of the proposed PAEF, we carried out two experiments, affective image classification and emotion score prediction. PAEF were then applied to predict emotions of masterpieces.

5.1 Datasets

IAPS dataset. The International Affective Picture System (IAPS) is a standard emotion evoking image set in psychology [19]. It consists of 1,182 documentary-style natural color images depicting complex scenes, such as portraits, babies, animals, landscapes, *etc*. Each image is associated with an empirically derived mean and standard deviation of valance, arousal, and dominance ratings in a 9-point rating scale. In this rating setting, rating score 9 represents a high rating on each dimension (i.e. high pleasure, high arousal, high dominance), and 1 represents a low rating on each dimension (low pleasure, low arousal, low dominance). This dataset and related emotion ratings were used for DES modeling. Similar to [24], we only modelled on the valence and arousal dimension, without considering the dominance dimension for its relatively small contributing scope on emotions [11].

Subset A of the IAPS dataset (**IAPSa**). Mikels *et al*. [27] selected 395 pictures from **IAPS** and categorized them into eight discreet categories: *Anger*, *Disgust*, *Fear*, *Sadness*, *Amusement*, *Awe*, *Contentment*, and *Excitement*.

Artistic dataset (**ArtPhoto**). This dataset consists of 806 artistic photographs from a photo sharing site searched by emotion categories [26].

Abstract dataset (**Abstract**). This dataset includes 228 peer rated abstract paintings without contextual content [26].

The latter three datasets (IAPSa, ArtPhoto, Abstract for short) were used for CES modeling. The summary of these datasets is listed in Table 3.

5.2 Affective Image Classification

We compared our emotion classification method with Wang *et al*. [38], Machajdik *et al*. [26] and Yanulevskaya *et al*. [42]. We adopted a "one category against all" strategy for experimental setup. The data was separated into a training set and a test set using K-fold Cross Validation (K=5) for 10 runs. Similar to [26], we optimized for the *true positive rate per class* averaged over the positive and negative classes, to overcome the limit of unbalanced data distribution of each category. We utilized PCA to perform dimensionality reduction on the feature vectors. Figures 11 to 13 illustrate the comparison of average classification performance and standard deviation between the proposed method and those of Machajdik *et al*. [26], Wang *et al*. [38] and Yanulevskaya *et al*. [42] on the IAPSa dataset, the Abstract dataset and the Artistic dataset, respectively.

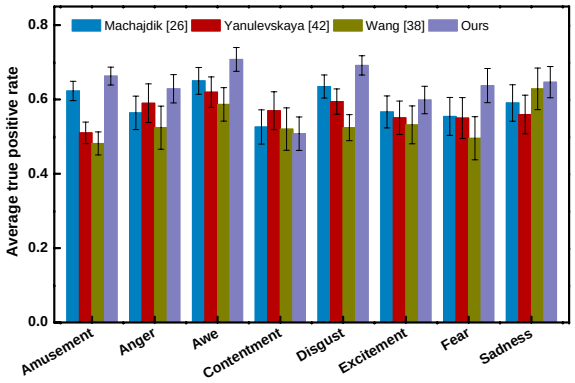


Figure 11: Classification performance on the IAPSa dataset compared to Machajdik *et al* [26], Yanulevskaya *et al* [42] and Wang *et al* [38].

From the results, it is clear that our method outperforms the state-of-the-art methods, achieving an improvement of about 5% on classification accuracy on average. This improvement arises because the state-of-the-art methods only consider the value of different low-level visual features, without considering the relationships of elements, while our proposed PAEF takes the elements' arrangements and orchestrations into account. The classification improvement demonstrates that principles-of-art are important in expressing image emotions. From the results of standard deviation, we can conclude that the proposed features are more robust for affective image classification than the use of low-level visual features.

[1]http://www.csie.ntu.edu.tw/~cjlin/libsvm/

Table 3: Summary of the three datasets with discrete emotion categories for affective image classification.

Dataset	Amusement	Anger	Awe	Contentment	Disgust	Excitement	Fear	Sadness	Sum
IAPSa	37	8	54	63	74	55	42	62	395
ArtPhoto	101	77	102	70	70	105	115	166	806
Abstract	25	3	15	63	18	36	36	32	228
Combined	163	88	171	196	162	196	193	260	1429

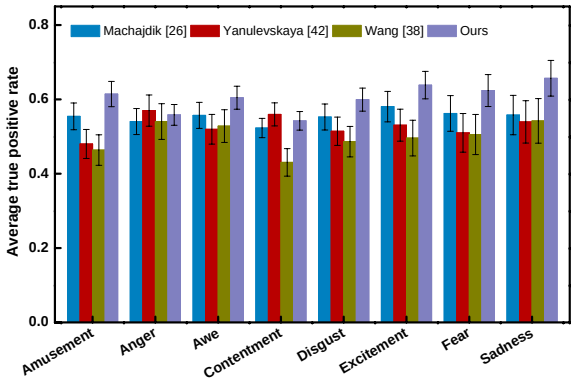


Figure 12: Classification performance on the Abstract dataset compared to Machajdik *et al* [26], Yanulevskaya *et al* [42] and Wang *et al* [38].

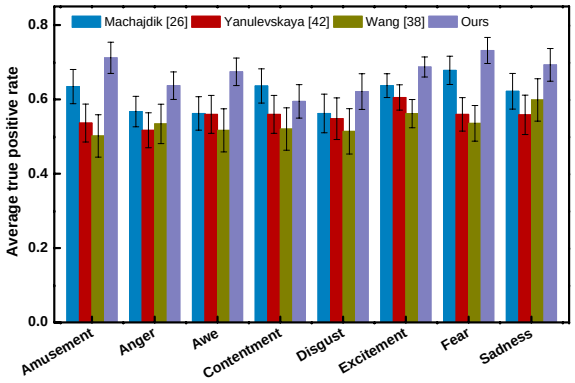
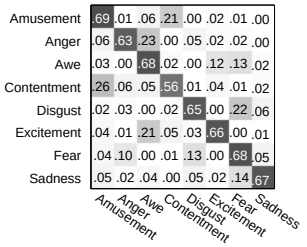


Figure 13: Classification performance on the ArtPhoto dataset compared to Machajdik *et al* [26], Yanulevskaya *et al* [42] and Wang *et al* [38].

Table 4: Measurements ranking list for the contribution to affective image classification.

	IAPSa	Abstract	ArtPhoto
Amusement	jbeghacdkifl	egdhkbfacjli	gedjkhfcaibl
Anger	bgkdfacjlehi	efgjklabcidh	befgjklacdih
Awe	bjefgldchkai	edkclbghafji	cbkhfldegaji
Contentment	bfhjegkdilac	ebfkhacljdig	fbcekhagdjli
Disgust	ecdajhfbkigl	eklghbcadfji	gelbadhcfkji
Excitement	fghdbcjkleia	gjcdhkafiebl	cbejdgkahifl
Fear	dcghkaejfbli	bgdhakcejfil	cgdkhejlabif
Sadness	fbecljhkagdi	efkbcdgahjli	fbhjdkglciea



(a)

(b)

Figure 14: (a) The average confusion matrix of classification results on the three dataset. (b) The image named '2070.jpg' in the IAPS dataset.

Comparing different datasets, we can also observe that the classification accuracy on the Abstract and ArtPhoto datasets is better than that on the IAPSa dataset. This is because in the IAPSa dataset, the emotions are usually evoked by certain objects in the images, while in the other two datasets, the images are taken by artists who understand and utilize the principles-of-art better.

The 8-class confusion matrix of our final results is shown in Fig. 14(a). Some pair-wise emotions are difficult to classify, such as *amusement* and *contentment*, *fear* and *disgust*. This is easy to understand, because one image can evoke different emotions. For example, for the image shown in Fig. 14(b), some people may feel *amusement* while others may feel *contentment*.

In order to evaluate the effectiveness of the measurements for each principle and its contribution for affective image classification, we built classifiers for each measurement. We sorted the measurements based on the classification accuracy in a descending order with the results in Table 4. The letters from 'a' to 'l' represent the measurements of *Bilateral symmetry*, *Rotational symmetry*, *Radial symmetry*, *Itten color contrast*, *RFA*, *Rangeability of hue and gradient direction*, *Color names*, *Distribution of gradient*, *Absolute variation*, *Relative variation*, *Relative total variation* and *Gaze scan path*, respectively. Readers can refer to Table 2 for the detailed meanings of each measurement.

From the results and the visulization results for different principles, we draw the following conclusions: (1) The best features for affective image classification are dependent on the emotion category, which means that different combinations of principles express different emotions. (2) The best features for affective image classification are dependent on the dataset, this is because the three datasets vary greatly from each other. Hence, based on the above two observations, we use all the principle features instead of selecting optimal feature combinations for different datasets and different emotions. (3) In terms of roles of different principles, symmetry (balance) and harmony tend to express positive emotions more often, while emphasis (contrast) and variety play an important role in classifying all the 8 categories of emotions. (4) Relative variation performs better than absolute variation, the eye scan path (movement) mainly focuses on the emphasis area, while RFA is extremely effective for emotion classification in the Abstract dataset.

Table 5: Comparison of MSE (standard deviation) for VA dimensions in the IAPS dataset.

	Machajdik [26]	**PAEF**	**Combination**
Valence	1.49(0.21)	1.31(0.15)	1.27(0.13)
Arousal	1.06(0.13)	0.85(0.10)	0.82(0.09)

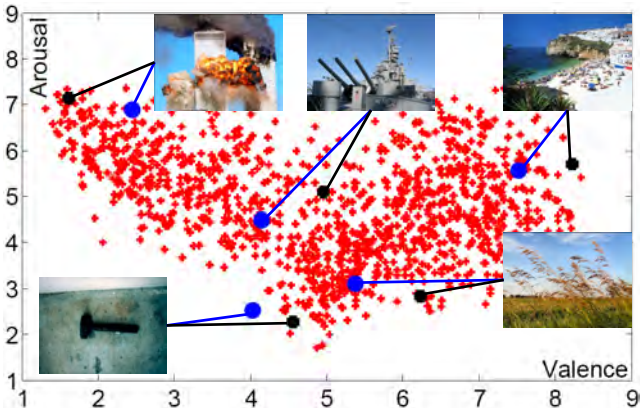


Figure 15: Emotion prediction results of our method. The black plus signs and blue circles represent the ground truth and our predicted values of image emotions, respectively.

5.3 Emotion Score Prediction

We used SVR with RBF kernel to model the VA dimensions on the IAPS dataset, and computed the mean squared error (MSE) of each dimension as the evaluation measurement. The lower the MSE is, the better the regression is. We compared our method with Machajdik's features [26] and the combination, using 5-cross validation for 10 runs. From Table 5, we can see that: (1) both valence and arousal are more accurately modeled by our principles-of-art features than Machajdik's features; (2) both our principles-of-art features and Machajdik features predict arousal better than valence; and (3) there is little improvement (3.05% and 3.53% decrease in MSE for valence and arousal) by combining them together, indicating that the principle features provide a strong enough ability in understanding image emotions. Some regression results are given in Fig. 15, which demonstrates the effectiveness of our image emotion prediction method.

We also conducted the VA emotion regression task using each of the six principles. From the MSE results in Table 6, we find that *variety*, *emphasis*, *gradation* and *balance* have higher correlations with valence, while *emphasis*, *variety*, *harmony* and *movement* are more correlated with arousal.

5.4 Inferring Masters' Moods

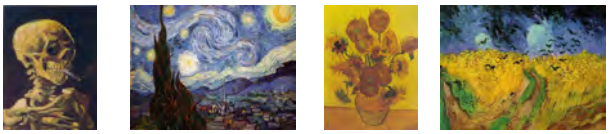
Masters have strong abilities to capture scenes or subjects into artworks which evoke strong emotional responses [42, 15]. Inferring the emotions implied in the masterpieces can immensely help in understanding the essential moods that the masters intended to express at that time.

Here we gathered 1,029 paintings and 158 watercolors of Vincent van Gogh, a famous Post-Impressionist painter, to infer his moods at different life periods, including early years (1881-1883), Nuenen (1883-1886), Antwerp (1883-1886), Paris (1886-1888), Arles (1888-1889), Saint-Remy (1889-1890) and Auvers (1890).

We used PAEF to predict the implied image emotions from van Gogh's artworks based on the training results of the three different datasets, IAPSa, Abstract and ArtPhoto, respectively. Some representational paintings and our predicted emotions are shown in Figure 16. We can observe that the training result in ArtPhoto dataset performs best. So we used this training result to predict all the artworks. The prediction result is shown in Table 7, from which we can see the distribution of the number of his paintings and watercolors. Note that one image can evoke different emotions. For each life period of van Gogh in Table 7, the first and second rows of each entry represent the numbers of paintings and watercolors, respectively. Take the painting "Wheat Field with Crows" (van Gogh's last painting) as example, the comments from *vangoghgallery* (www.vangoghgallery.com) are heavy, lonely, gloom, and melancholy, and our prediction emotion is *sadness*, clearly indicting the emotional status of his final days.

Table 6: MSE of each principle for VA dimensions in the IAPS dataset.

	Ban	Emp	Har	Var	Gra	Mov
Valence	1.85	1.72	2.16	1.67	1.78	2.37
Arousal	1.52	0.98	1.12	1.07	1.61	1.15



Awe Excitement	Disgust	Sadness	Contentment
Awe	Awe Excitement	Neutral	Excitement
Fear	Awe Contentment	Contentment	Sadness

Figure 16: van Gogh's masterpieces, and our predicted emotions. The paintings are "Skull with burning cigarette", "Starry night", "Still life vase with fourteen sunflowers" and "Wheat field with crows" from left to right. The three rows of predicted emotions below the paintings are based on the training results in IAPSa, Abstract and ArtPhoto datasets, respectively.

Table 7: Emotion prediction result of van Gogh's artworks.

Period	**Am**	**An**	**Aw**	**Co**	**Di**	**Ex**	**Fe**	**Sa**	**Ne**	**Sum**
Early	0	0	0	1	4	3	9	22	8	35
	0	0	0	1	7	3	15	22	45	88
Nuenen	0	0	1	3	41	26	73	75	25	200
	0	0	0	0	3	1	3	10	9	24
Antwerp	0	0	0	0	3	0	5	2	1	7
	0	0	0	0	0	0	0	0	0	0
Paris	11	1	3	7	35	47	44	45	53	224
	0	0	0	0	1	0	3	2	4	10
Arles	11	5	1	19	45	69	52	49	87	304
	1	0	0	0	6	7	3	1	3	21
SaintRemy	8	12	3	11	36	36	22	11	57	177
	1	0	0	0	1	6	1	1	2	11
Auvers	6	0	5	3	7	22	6	8	29	82
	0	0	0	0	0	0	1	2	2	4

5.5 Discussion

From the classification results in Section 5.2 and the regression results in Section 5.3, we can conclude that PAEF can indeed help to improve the performance of image emotion recognition. The results demonstrate that the principles-of-art features can model image emotions better and are more robust in image emotion recognition than the elements-of-art features. PAEF are especially helpful and accurate to handle the abstract and artistic images, the emotions of which are mainly determined by the composition.

However, as our method does not consider the semantics of images, it does not work so well for the images whose emotions are dominated by some specific objects, concepts or scenes; and the emotion recognition performance is relatively low for these images. For example, in one image containing snakes, the emotion of *fear* may directly be evoked by the presence of snakes. In such cases, our method may fail. Combining the visual concept detection method, such as SentiBank [4], may help to tackle this problem and further improve the emotion recognition performance.

6. CONCLUSION AND FUTURE WORK

In this paper, we proposed to extract emotion features based on principles-of-art (PAEF) for image emotion classification and scoring task. Different from previous works that mainly extract low level visual features based on elements-of-art, we drew inspirations from the concept of principles-of-art for higher level understanding of images. Experimental results on affective image classification and regression have demonstrated that the performance of the proposed features is superior over the state-of-the-art approaches. The application of PAEF in emotion prediction of masterpieces is also interesting and has much potential for future research. PAEF can also be used to develop other emotion based applications, such as image musicalization [44] and affective image retrieval [45].

For further studies, we will continue our efforts to quantize the principles using more effective measurements and to improve the efficiency for real time implementation. Applying high level content detection and recognition methods may improve the performance of emotion recognition. In addition, we will consider using social network (e.g., Flickr) data, combining the descriptions and images to jointly learn the expected emotion of specified image based on visual-textual-social features [9] and analyzing the comments to distinguish expected emotion and actual emotion. How to analyze videos using visual features together with acoustic signals from an emotional perspective is also worth studying.

7. ACKNOWLEDGEMENTS

This work was supported by National Natural Science Foundation of China (No. 61071180) and Key Program (No. 61133003), and partially supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office.

Sicheng Zhao was also supported by the Ph.D. Short-Term Overseas Visiting Scholar Program of Harbin Institute of Technology.

8. REFERENCES

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

[2] D. H. Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111–122, 1981.

[3] S. Benini, L. Canini, and R. Leonardi. A connotative space for supporting movie affective recommendation. *IEEE Transactions on Multimedia*, 13(6):1356–1370, 2011.

[4] 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 MM*, 2013.

[5] S. Calahan. Storytelling through lighting: a computer graphics perspective. *SIGGRAPH course notes*, 96, 1996.

[6] R. G. Collingwood. *The principles of art*, volume 62. Oxford University Press, USA, 1958.

[7] R. Datta, J. Li, and J. Z. Wang. Algorithmic inferencing of aesthetics and emotion in natural images: An exposition. In *ICIP*, 2008.

[8] S. Dhar, V. Ordonez, and T. Berg. High level describable attributes for predicting aesthetics and interestingness. In *CVPR*, 2011.

[9] Y. Gao, M. Wang, Z.-J. Zha, J. Shen, X. Li, and X. Wu. Visual-textual joint relevance learning for tag-based social image search. *IEEE Transactions on Image Processing*, 22(1):363–376, 2013.

[10] A. Hanjalic. Extracting moods from pictures and sounds: Towards truly personalized tv. *IEEE Signal Processing Magazine*, 23(2):90–100, 2006.

[11] A. Hanjalic and L.-Q. Xu. Affective video content representation and modeling. *IEEE Transactions on Multimedia*, 7(1):143–154, 2005.

[12] J. Hobbs, R. Salome, and K. Vieth. *The visual experience*. Davis Publications, 1995.

[13] G. Irie, T. Satou, A. Kojima, T. Yamasaki, and K. Aizawa. Affective audio-visual words and latent topic driving model for realizing movie affective scene classification. *IEEE Transactions on Multimedia*, 12(6):523–535, 2010.

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

[15] J. Jia, S. Wu, X. Wang, P. Hu, L. Cai, and J. Tang. Can we understand van gogh's mood? learning to infer affects from images in social networks. In *ACM MM*, 2012.

[16] D. Joshi, R. Datta, E. Fedorovskaya, Q. Luong, J. Z. Wang, J. Li, and J. Luo. Aesthetics and emotions in images. *IEEE Signal Processing Magazine*, 28(5):94–115, 2011.

[17] H. Kang. Affective content detection using hmms. In *ACM MM*, 2003.

[18] M. Kass and J. Solomon. Smoothed local histogram filters. *ACM Transactions on Graphics*, 29(4):100, 2010.

[19] P. Lang, M. Bradley, B. Cuthbert, et al. *International affective picture system (IAPS): Affective ratings of pictures and instruction manual*. NIMH, Center for the Study of Emotion & Attention, 2005.

[20] B. Li, W. Xiong, W. Hu, and X. Ding. Context-aware affective images classification based on bilayer sparse representation. In *ACM MM*, 2012.

[21] L. Liu, R. Chen, L. Wolf, and D. Cohen-Or. Optimizing photo composition. In *Computer Graphics Forum*, 2010.

[22] G. Loy and J. Eklundh. Detecting symmetry and symmetric constellations of features. In *ICCV*, 2006.

[23] G. Loy and A. Zelinsky. Fast radial symmetry for detecting points of interest. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):959–973, 2003.

[24] 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 MM*, 2012.

[25] Y. Luo and X. Tang. Photo and video quality evaluation: Focusing on the subject. In *ECCV*, 2008.

[26] J. Machajdik and A. Hanbury. Affective image classification using features inspired by psychology and art theory. In *ACM MM*, 2010.

[27] J. Mikels, B. Fredrickson, G. Larkin, C. Lindberg, S. Maglio, and P. Reuter-Lorenz. Emotional category data on images from the international affective picture system. *Behavior research methods*, 37(4):626–630, 2005.

[28] J. Ni, M. Singh, and C. Bahlmann. Fast radial symmetry detection under affine transformations. In *CVPR*, 2012.

[29] M. A. Nicolaou, H. Gunes, and M. Pantic. A multi-layer hybrid framework for dimensional emotion classification. In *ACM MM*, 2011.

[30] M. Redi and B. Merialdo. Enhancing semantic features with compositional analysis for scene recognition. In *ECCV Workshop*, 2012.

[31] J. Ruskan. *Emotion and Art: Mastering the Challenges of the Artist's Path*. R. Wyler & Co., 2012.

[32] H. Schlosberg. Three dimensions of emotion. *Psychological review*, 61(2):81, 1954.

[33] M. Solli and R. Lenz. Color based bags-of-emotions. In *CAIP*, 2009.

[34] X. Sun, H. Yao, and R. Ji. What are we looking for: Towards statistical modeling of saccadic eye movements and visual saliency. In *CVPR*, 2012.

[35] X. Sun, H. Yao, R. Ji, and S. Liu. Photo assessment based on computational visual attention model. In *ACM MM*, 2009.

[36] J. Van De W., C. Schmid, and J. Verbeek. Learning color names from real-world images. In *CVPR*, 2007.

[37] J. C. van Gemert. Exploiting photographic style for category-level image classification by generalizing the spatial pyramid. In *ICMR*, 2011.

[38] W. Wang, Y. Yu, and S. Jiang. Image retrieval by emotional semantics: A study of emotional space and feature extraction. In *IEEE SMC*, 2006.

[39] X. Xiang and M. Kankanhalli. Affect-based adaptive presentation of home videos. In *ACM MM*, 2011.

[40] L. Xu, Q. Yan, Y. Xia, and J. Jia. Structure extraction from texture via relative total variation. *ACM Transactions on Graphics*, 31(6):139, 2012.

[41] M. Xu, J. S. Jin, S. Luo, and L. Duan. Hierarchical movie affective content analysis based on arousal and valence features. In *ACM MM*, 2008.

[42] V. Yanulevskaya, J. Van Gemert, K. Roth, A. Herbold, N. Sebe, and J. Geusebroek. Emotional valence categorization using holistic image features. In *ICIP*, 2008.

[43] S. Zhao, H. Yao, X. Sun, P. Xu, X. Liu, and R. Ji. Video indexing and recommendation based on affective analysis of viewers. In *ACM MM*, 2011.

[44] S. Zhao, H. Yao, F. Wang, X. Jiang, and W. Zhang. Emotion based image musicalization. In *ICMEW*, 2014.

[45] S. Zhao, H. Yao, Y. Yang, and Y. Zhang. Affective image retrieval via multi-graph learning. In *ACM MM*, 2014.