Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)

# Affective Computing and Applications of Image Emotion Perceptions

**Sicheng Zhao, Hongxun Yao**
School of Computer Science and Technology,
Harbin Institute of Technology, Harbin, China.
zsc@hit.edu.cn, h.yao@hit.edu.cn

## Introduction

Images can convey rich semantics and evoke strong emotions in viewers. The research of my PhD thesis focuses on image emotion computing (IEC), which aims to predict the emotion perceptions of given images. The development of IEC is greatly constrained by two main challenges: affective gap and subjective evaluation (Zhao et al. 2014a). Previous works mainly focused on finding features that can express emotions better to bridge the affective gap, such as elements-of-art based features (Machajdik and Hanbury 2010) and shape features (Lu et al. 2012).

According to the emotion representation models, including categorical emotion states (CES) and dimensional emotion space (DES) (Zhao et al. 2014a), three different tasks are traditionally performed on IEC: affective image classification, regression and retrieval. The state-of-the-art methods on the three above tasks are image-centric, focusing on the dominant emotions for the majority of viewers.

For my PhD thesis, I plan to answer the following questions: (1) Compared to the low-level elements-of-art based features, can we find some higher level features that are more interpretable and have stronger link to emotions? (2) Are the emotions that are evoked in viewers by an image subjective and different? If they are, how can we tackle the user-centric emotion prediction? (3) For image-centric emotion computing, can we predict the emotion distribution instead of the dominant emotion category?

## Principles-of-Art Based Emotion Features

The artistic 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* (Collingwood 1958; Hobbs, Salome, and Vieth 1995). We systematically study and formulize the former 6 artistic principles, without considering *rhythm* and *proportion*, as they are ambiguously defined. For each principle, we explain the concepts and meanings and translate these concepts into mathematical formulae for quantization measurement.

Take *emphasis* for example. *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, which is usually used to direct and focus viewers' attention to the most important area or centers of interests of a design. We adopt Itten's color contrasts and the rate of focused attention (RFA) to measure *emphasis*. This part has been finished in (Zhao et al. 2014a). We demonstrate that the principles-of-art based emotion features (PAEF) can model emotions better and are more interpretable to humans.

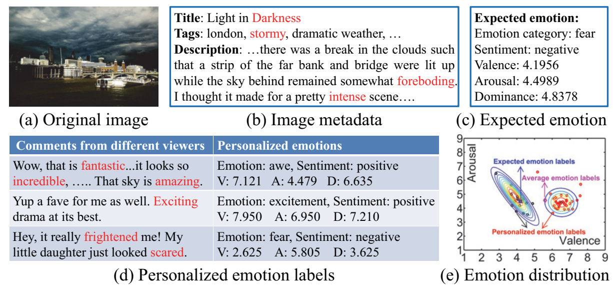


Figure 1: Illustration of personalized image emotion perceptions in social networks. The emotions are obtained using the keywords in red. The contour lines of (e) are the assigned emotion distributions by expectation-maximization (EM) algorithm based on Gaussian Mixture Models (GMM).

## Personalized Emotion Perception Prediction

The images in Abstract dataset (Machajdik and Hanbury 2010) were labeled by 14 people on average. 81% images are assigned with 5 to 8 emotions. So the perceived emotions of different viewers may vary.

To further demonstrate this observation, we set up a large-scale dataset, named Image-Emotion-Social-Net dataset, with over 1 million images downloaded from Flickr. To get the personalized emotion labels, firstly we use traditional lexicon-based methods as in (Jia et al. 2012; Yang et al. 2014) to obtain the text segmentation results of the title, tags and descriptions from uploaders for expected emotions and the comments from viewers for actual emotions. Then we compute the average value of valence, arousal and dominance of the segmentation results as ground truth for dimensional emotion representation based on recently published VAD norms of 13,915 English lemmas (Warriner, Kuperman, and Brysbaert 2013). After refinement, we obtain 1,434,080 emotion labels on 1,012,901 images uploaded by 11,347 users and commented by 106,688 users. The average STDs of VAD forlabeled by more than 10 users are 1.44, 0.75 and 0.87 (VAD range over the interval (1,9)).

Similar to (Joshi et al. 2011; Peng et al. 2015), we can conclude that the emotions that are evoked in viewers by an image are subjective and different, as shown in Figure 1. Therefore, predicting the personalized emotion perceptions for each viewer is more reasonable and important. In such cases, the emotion prediction tasks become user-centric.

Intuitively, four types of factors can influence the emotion perception and can be exploited for emotion prediction: visual content, social context, temporal evolution, and location influence. We propose rolling multi-task hypergraph learning to jointly combine these factors for personalized emotion perception prediction. This part is being in progress.

## Emotion Distribution Prediction

By the statistical analysis on the images viewed by a large population, we observe that though subjective and different, the personalized emotion perceptions follow certain distributions (see Figure 1(e)). Predicting the emotion probability distribution instead of a single dominant emotion for an image is accordant with the subjective evaluation, which reveals the difference of emotional reactions between users. Generally, the distribution prediction task can be formulized as a regression problem. For different emotion representation models, the distribution prediction varies slightly.

For CES, the task aims to predict the discrete probability of different emotion categories, the sum of which equals to 1. We propose to use shared sparse learning to predict the discrete probability distribute of image emotion. This work has been finished (Zhao et al. 2015).

For DES, the task usually transfers to predict the parameters of specified continuous probability distribution, the form of which should be firstly decided, such as Gaussian distribution and exponential distribution. From the statistics analysis with one example in Figure 1(e), we observe that (1) the perceived dimensional emotions can be clearly grouped into two clusters, corresponding to the positive and negative sentiments; (2) the VA emotion labels can be well modeled by a mixture of two bidimensional Gaussian distributions. Based on these observations, we define the distribution of VA emotion labels as a GMM by $p(\mathbf{x};\boldsymbol{\theta}) = \sum_{l=1}^{L} \pi_l \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)$, where $\mathbf{x} = (v, a)$ is pair-wise VA emotion labels, $\boldsymbol{\mu}_l$ and $\boldsymbol{\Sigma}_l$ are the mean vector and covariance matrix of the $l$th Gaussian component, while $\pi_l$ is the mixing coefficient, which satisfies $\pi_l \geq 0$ and $\sum_{l=1}^{L} \pi_l = 1$.

The EM algorithm is used to estimate the parameters of GMM. Specifically, the initializations are obtained by firstly partitioning the VA labels into two clusters based on whether valence is greater than 5 and then computing the mean vector $\boldsymbol{\mu}_l$ and covariance matrix $\boldsymbol{\Sigma}_l$ of each cluster. The mixing coefficients are set as the proportions of related VA labels in each cluster to the total labels. In experiment, the EM algorithm is converged in 6.28 steps on average. Now the task turns to predict $\boldsymbol{\theta}$. An preliminary version has been finished (Zhao, Yao, and Jiang 2015).

## Emotion Based Applications

We design some interesting applications based on image emotions. The first is affective image retrieval, which aims to retrieve images with similar emotions to the given image. We propose to use multi-graph learning as a feature fusion method to efficiently explore the complementation of different features (Zhao et al. 2014c). The second application is emotion based image musicalization (Zhao et al. 2014b), which aims to make images vivid when people are watching them. The music with approximate emotions to the image emotions is selected to musicalize these images. For future works, we plan to modify the images without changing the high level content to transfer the original evoked emotions.

This work was supported by the National Natural Science Foundation of China (No. 61472103 and No. 61133003).

## References


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

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

Jia, J.; Wu, S.; Wang, X.; Hu, P.; Cai, L.; and Tang, J. 2012. Can we understand van gogh's mood? learning to infer affects from images in social networks. In *ACM MM*, 857–860.

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

Lu, X.; Suryanarayan, P.; Adams Jr, R. B.; Li, J.; Newman, M. G.; and Wang, J. Z. 2012. On shape and the computability of emotions. In *ACM MM*, 229–238.

Machajdik, J., and Hanbury, A. 2010. Affective image classification using features inspired by psychology and art theory. In *ACM MM*, 83–92.

Peng, K.-C.; Sadovnik, A.; Gallagher, A.; and Chen, T. 2015. A mixed bag of emotions: Model, predict, and transfer emotion distributions. In *IEEE CVPR*, 860–868.

Warriner, A. B.; Kuperman, V.; and Brysbaert, M. 2013. Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behavior Research Methods* 45(4):1191–1207.

Yang, Y.; Jia, J.; Zhang, S.; Wu, B.; Chen, Q.; Li, J.; Xing, C.; and Tang, J. 2014. How do your friends on social media disclose your emotions? In *AAAI*, 306–312.

Zhao, S.; Gao, Y.; Jiang, X.; Yao, H.; Chua, T.-S.; and Sun, X. 2014a. Exploring principles-of-art features for image emotion recognition. In *ACM MM*, 47–56.

Zhao, S.; Yao, H.; Wang, F.; Jiang, X.; and Zhang, W. 2014b. Emotion based image musicalization. In *IEEE ICMEW*, 1–6.

Zhao, S.; Yao, H.; Yang, Y.; and Zhang, Y. 2014c. Affective image retrieval via multi-graph learning. In *ACM MM*, 1025–1028.

Zhao, S.; Yao, H.; Jiang, X.; and Sun, X. 2015. Predicting discrete probability distribution of image emotions. In *IEEE ICIP*.

Zhao, S.; Yao, H.; and Jiang, X. 2015. Predicting continuous probability distribution of image emotions in valence-arousal space. In *ACM MM*, 879–882.