A FRAMEWORK OF CHANGING IMAGE EMOTION USING EMOTION PREDICTION

Kuan-Chuan Peng, Kolbeinn Karlsson, Tsuhan Chen — *Dong-Qing Zhang, Heather Yu*

Cornell University
School of Electrical and Computer Engineering
Ithaca, NY 14850

Huawei
Futurewei Technologies
400 Crossings Blvd, Suite 102, Bridgewater, NJ 08807

ABSTRACT

Most works about affective image classification in computer vision treat each emotion category independently and predict hard labels, ignoring the correlation between emotion categories. In this work, inspired by psychological theories, we adopt a dimensional emotion model to model the correlation among certain emotion categories. We also propose a framework of changing image emotion by using our emotion predictor. Easily extendable to other feature transformations, our framework changes image emotion by color histogram specification, relaxing the limitation of the previous method that each emotion is associated with a monotonic palette. Effective and comparable to the previous work of changing image emotion shown by user study, our proposed framework provides users with more flexible control in changing image emotion compared with the previous work.

***Index Terms*—** Emotion modification, dimensional emotion model, emotion prediction

(a) (b)

Fig. 1. An example of changing image emotion by our proposed framework. (a) and (b) are images before and after emotion adjustment. In this case, the goal of emotion adjustment is adding joy to the input image. 14 out of 15 subjects in our user study agree that (b) better represents joy compared with (a).

1. INTRODUCTION

What do you feel after looking at an image? The answer to this question varies from person to person depending on not only the content of the image but also their personal experiences. For example, a picture showing a juicy beef burger may elate some fast food lovers, but some people may be irritated due to health reasons. Since an image in general evokes people's emotions differently, it is more appropriate to describe the emotion associated with an image in real numbers rather than hard labels, which motivates us to predict emotion in real numbers.

In computer vision, abstract concepts like affective image classification [1, 2, 3] and aesthetic quality estimation [4] attract researchers' attention recently. Even these two abstract concepts are related, they are not equivalent. For instance, the emotion joy can be possibly evoked by either aesthetically ideal images or noisy images. Furthermore, aesthetic quality is a one-dimensional attribute, but emotions are not [5].

In recent literature about affective image classification in computer vision, researchers conduct their experiments on various kinds of images. Solli and Lenz [2] focuses on Internet images, while Wang et al. [3] pay more attention to abstract paintings and artistic pictures. Machajdik and Hanbury [1] perform affective image classification on realistic as well as artistic images. To our surprise, none of these previous works made probabilistic or soft-label emotion prediction. In their works, different emotion categories are treated independently in 1-vs-all setting of multi-class classification, which is inconsistent with the fact that some emotion categories are closely related. For example, joy and sadness have strong negative correlation. To model the correlation of emotion categories, we use dimensional emotion model based on psychological studies [6] for emotion prediction in this work.

Besides the issue of ignoring the correlation between emotion categories, many emotion-related researches use certain image databases (such as emodb [2], GAPED [7], and IAPS [8]) which suffer from a few drawbacks: 1: Assigning hard labels to images, these databases ignore the fact that people do not necessarily have consensus in terms of emotions. Even with similar kind of emotion, the degree of emotion may vary (like joy vs. ecstacy). 2: The emotion categories of these databases are chosen in an ad-hoc way without solid foun-

(a) joy (b) sadness (c) non fear (d) fear (e) non disgust (f) disgust

Fig. 2. Some example images from Huawei3 with their categories.

Category	Dimension	Description
Edge	512	cascaded edge histograms in the most / least salient regions
Texture	27	features from gray-level co-occurrence matrix and Tamura features
Color	80	cascaded CIECAM02 color histograms in the most / least salient regions
Saliency	4	the differences of areas / color / edge features in the most / least salient regions
Composition	8	rule of third, diagonal dominance, symmetry, and visual balance
Shape	128	features of the fit ellipses of the segments from color segmentation

Table 1. The feature set we use to train the emotion predictor.

dation of psychological theories. 3: The number of images in each emotion category is not equal in these databases, so the unbalanced database may cause bias in experimental results. To solve the issues mentioned above, we build a new database, Huawei3, for emotion prediction. We provide more information about Huawei3 in Sec. 2.

Inspired by Wang et al. [9], we also propose a new framework of changing image emotion by leveraging the results of emotion prediction and adjusting the color tone of the image with histogram specification. In Wang's method [9], each emotion keyword is mapped to a particular palette and they adjust the color tone of an image according to the palette. Our method, on the other hand, is able to generate images with different color tones given the same emotion keyword. Fig. 1 shows an example of changing image emotion with our framework.

We make the following contributions: 1: We build a new image database, Huawei3, which solves the issues of previous databases mentioned in the third paragraph and models correlation between emotion categories in the dimensional emotion model for emotion prediction. 2: We propose a new framework in Sec. 4 for changing the emotions associated with images by changing the color tone with histogram specification. Easily extendable to other feature transformations, the proposed framework relaxes the limitation of the Wang's method [9].

2. THE HUAWEI3 DATABASE

Huawei3 contains 6 emotion categories forming 3 dimensions in emotion space. There are 500 images in each category. Each image provides its binary label in one of the three dimensions. The details about the selection of emotion categories, image collection, and labeling procedure are described in the following subsections. Fig. 2 shows example images from Huawei3.

2.1. Emotion category

Despite extensive psycological research and debates, there is still no consensus on how to model emotions [10]. One popular class of emotion models is the dimensional emotion model, originated by Wundt, the father of modern psychology, who described emotions in three dimensions [11]. We adopt dimensional emotion models because they are consistent with real-valued emotion prediction. Inspired by Plutchik's wheel [6], one of the dimensional emotion models, we define three dimensions d_i ($i \in \{1, 2, 3\}$) in emotion space where each dimension represents a basic emotions in Plutchik's wheel. In our emotion model, the three dimensions are joy–sadness (d_1), fear–non fear (d_2), and disgust–non disgust (d_3). We treat the six emotions forming the three dimensions as the emotion categories in Huawei3.

2.2. Image collection and labeling

By entering the six category keywords as the searching keywords, we collect the images of Huawei3 from Flickr. In practice, we use synonyms and antonyms of joy, fear, and disgust as searching keywords. Instead of downloading the raw searching results directly like emodb [2] without verification, we check every image in Huawei3 to prevent erroneous images. We label each image with 0 (sadness, fear, or disgust) or 1 (joy, non fear, or non disgust) in the corresponding dimension. For example, an image in the fear category will be labeled 0 in fear–non fear dimension. We collect 500 images for each of the six emotion categories, so Huawei3 consists of 3000 images, comparable to previous databases [7, 8]. Every image in Huawei3 is resized to approximately VGA resolution with the original aspect ratio intact.

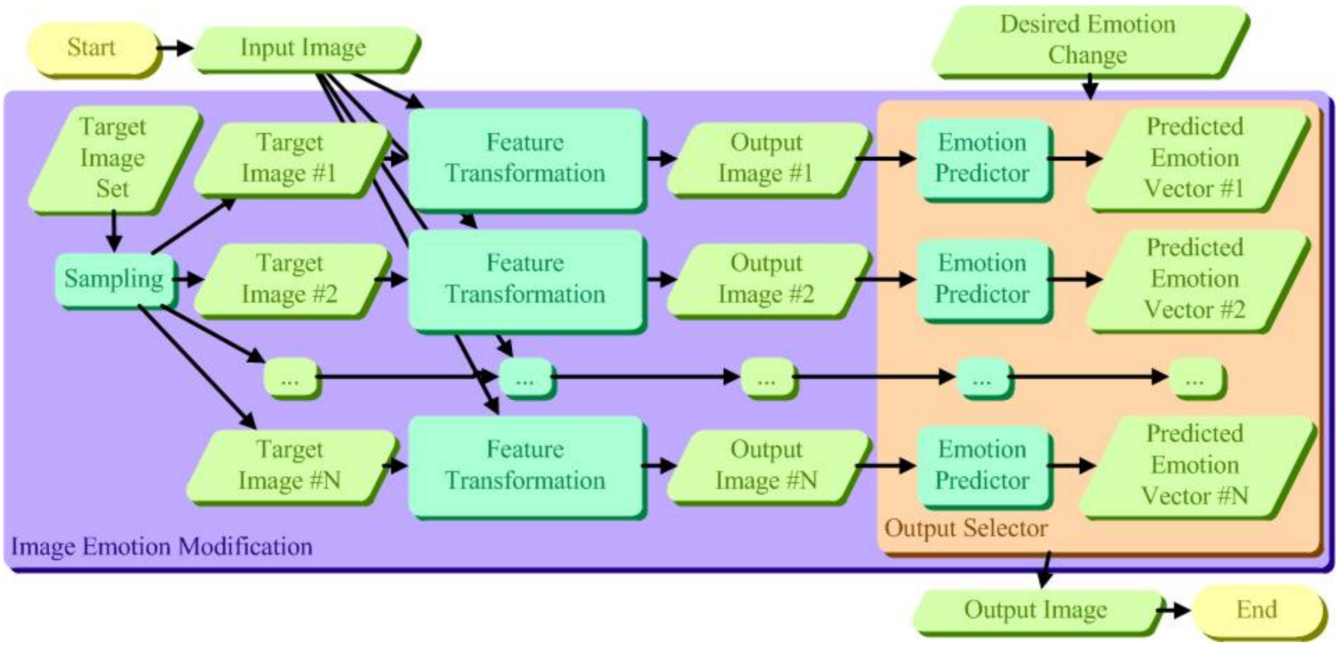

Fig. 3. The flowchart of changing image emotion.

3. REAL-VALUED EMOTION PREDICTION

Inspired by previous works in affective image classification [1, 2, 3], we create a 759-dimensional feature set consisting of the features from 6 categories listed in Table 1. For each image, the corresponding feature vector is computed. Each dimension of the feature vector is properly normalized to the range [0, 1]. Randomly splitting each emotion category of Huawei3 into training and testing set, we train an emotion predictor EP_i in each d_i using the training set associated with d_i. By using standard support vector regression (SVR) provided by LIBSVM [12], each EP_i predicts s_i, the regression value of the corresponding emotion in d_i. Specifically, s_1, s_2, and s_3 represent the regression values of joy, non fear, and non disgust respectively. In each d_i, higher regression value represents more positive emotion. The parameters of SVR are learned by performing 10-fold cross validation on the training set. The final real-valued emotion predictor EP is formed by cascading all EP_is such that EP will take an image as input and output a vector $\vec{e} = (s_1, s_2, s_3)$ in emotion space.

The mean squared errors of the prediction of our model on the testing set in d_1, d_2, and d_3 are 0.209, 0.111, and 0.215 respectively. If we place a threshold at 0.5 for each predicted value and treat each dimension as a binary classification problem, the accuracy of the prediction of our model on the testing set in d_1, d_2, and d_3 are 0.703, 0.850, and 0.663 respectively, which is comparable to the results of previous works in affective image classification [1, 2, 3].

4. CHANGING IMAGE EMOTION

The entire framework of changing image emotion is summarized in Fig. 3. Given an input image and the desired change of $\vec{e}$, the framework transforms features of the input image with the guidance of a target image sampled from a predefined target image set which consists of 250 unlabeled images collected from the Internet. A total of N target images are randomly sampled from the target image set and N corresponding output images are generated by the feature transformations. Using EP to predict the $\vec{e}$ vectors of these N output images, the framework will output the image with the change of emotion closest to what the user specified. To reach a balance between output variety and computational efficiency, we empirically set $N = 20$.

In our experiment, the default sampling method is random sampling, and the feature transformation we use is a color tone adjustment by applying histogram specification to CIE XYZ channels independently. The distance metric used to compare the change of emotions in output selector can be fully customized, for example, L2-norm. In our experiment, our output selector chooses the result with the highest or lowest s_i in the specified dimension. Comparing with Wang's method of changing emotions [9], our proposed framework has the following advantages: 1: Wang's method did color transformation by associating each emotion keyword with a palette, so the output images will have similar color tones given the same emotion keyword. Our framework can generate output images with different color tones given the same input because of the target image set and the sampling. 2:

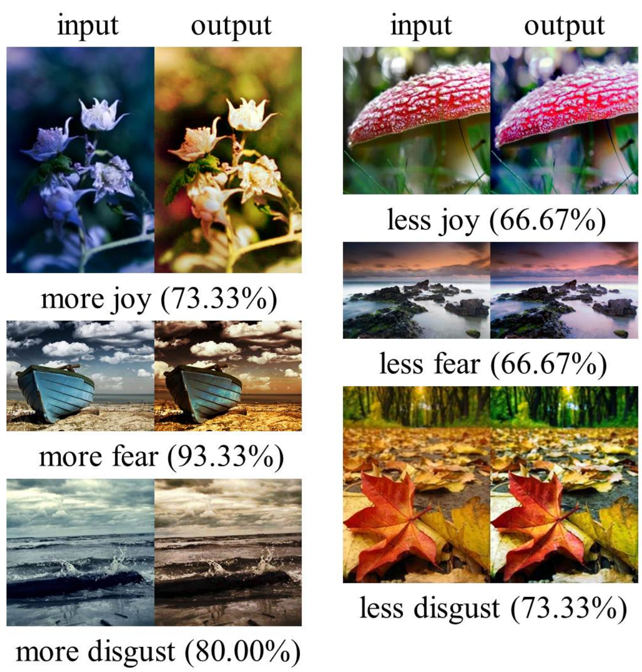

Fig. 4. Some examples of changing image emotion by using our framework. The criterium of output selector and the percentage out of 15 subjects agreeing with the corresponding emotion change are also shown under each pair of images. In these examples, more than 50% of the subjects agree with our emotion change.

Our feature transformation block is easily extendable to feature transformations other than color (e.g. edge-histogram specification [13]), while Wang's method needs the predefined mapping between emotion keywords and features. 3: Wang's method can only change image emotion by specifying one emotion keyword. Our framework allows the user to specify the desired change of emotion in every dimension in our dimensional emotion model, offering more sophisticated control.

5. EXPERIMENTAL RESULTS

To test the efficacy of our proposed framework, we randomly select 20 images as input and change their emotion content using our framework. For each image, our output selector chooses the output image based on one of the following six criteria: the image represents joy/fear/disgust the most/least. We assign one of the six criteria to each input image randomly under the constraint that at least three input images are assigned to each criterium. We apply our framework to generate an output image for each input image according to the assigned selecting criterium, and put these 20 input/output image pairs on Amazon Mechanical Turk (AMT) to perform a user study. For each pair of images, we show the subjects both images simultaneously and ask them to choose the one that best corresponds to the given emotion keyword (one of joy/fear/disgust consistent with the selecting criterium of the pair). Without letting the subjects know that our framework increases or decreases emotions, we collect responses from 15 different subjects on AMT for each HIT consisting of one pair of images. We offer 2 cents to reward the subject's completion of each HIT.

The experimental setting of our user study is inspired by that of Wang's method [9]. However, Wang's user study only compares two images processed by their algorithm and Photoshop artists without comparing with the original image. We believe that our user study which directly compares with the original images is a more convincing way to show the efficacy of our proposed framework. Figure 4 shows some examples of changing image emotion generated by our framework. The criterium of output selector and the percentage out of 15 subjects agreeing with the corresponding emotion changing are also shown under each pair of images. Out of 20 pairs of images, there are 13 pairs where more than 50% of the subjects agree with our emotion changing. 66.67% of all the individual responses are consistent with the corresponding selecting criterium of our framework, which is comparable to the performance of Wang's method [9].

6. CONCLUSION

In this paper we present a new image database, Huawei3, based on psychological theories, fixing some issues of previous emotion databases. We use a dimensional emotion model to capture the correlation between emotion categories and build a real-valued emotion predictor from Huawei3. We also propose a novel framework for changing image emotion content and show that it is effective and comparable to previous work by the results of our user study.

Proposing a novel framework of changing emotion, we show that our framework is effective and comparable to the previous work from the results of user study. Moreover, our framework outshines the previous work in output variety, flexibility of output selecting criteria, and extensibility to other feature transformations.

7. FUTURE WORK

Even though we show that 66.67% of the user evaluations from AMT agree with our framework's changing emotion, there are still 33.33% of the user evaluations expressing different opinions, which is expected because people in general do not have consensus on emotions evoked by an image. Therefore, we plan to use machine learning to personalize the major components in our proposed framework, including the target image set, sampling strategy, emotion predictor, and the criterium of output selector to change image emotion according to individual feelings.

8. REFERENCES

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

[2] M. Solli and R. Lenz, "Emotion related structures in large image databases," in *International Conference on Image and Video Retrieval*. ACM, 2010, pp. 398–405.

[3] X. Wang, J. Jia, J. Yin, and L. Cai, "Interpretable aesthetic features for affective image classification," in *International Conference on Image Processing*. IEEE, 2013, pp. 3230–3234.

[4] D. Joshi et al., "Aesthetics and emotions in images: a computational perspective.," *IEEE Signal Processing Magazine*, vol. 28, no. 5, pp. 94–115, 2011.

[5] J. R. J. Fontaine, K. R. Scherer, E. B. Roesch, and P. C. Ellsworth, "The world of emotions is not two-dimensional.," *Psychological Science*, vol. 18, no. 2, pp. 1050–1057, 2007.

[6] R. Plutchik, "A general psychocvolutionary theory of emotion," *Emotion: Theory, Research, and Experience*, vol. 1, pp. 3–31, 1980.

[7] E. S. Dan-Glauser and K. R. Scherer, "The geneva affective picture database (GAPED): a new 730-picture database focusing on valence and normative significance.," *Behavior Research Methods*, vol. 43, no. 2, pp. 468–477, 2011.

[8] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system (IAPS): affective ratings of pictures and instruction manual. technical report a-8.," 2008.

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

[10] A. Ortony and T. J. Turner, "What's basic about basic emotions?," *Psychological Review*, vol. 97, no. 3, pp. 315–331, 1990.

[11] W. M. Wundt, "Outlines of psychology," *Classics in the History of Psychology*, 1897.

[12] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.

[13] M. Mignotte, "An energy-based model for the image edge-histogram specification problem.," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 379–386, 2012.