Proceedings of the International Conference on Computer and Information Science and Technology
Ottawa, Ontario, Canada, May 11 – 12, 2015
Paper No. 131

Emotion Discovery by Analysing Facial Images

Yingxiong Li
Department of Computing
The Hong Kong Polytechnic University
Hong Kong, China

Vincent Ng
Department of Computing
The Hong Kong Polytechnic University
Hong Kong, China
cstyng@comp.polyu.edu.hk

Abstract- Recently, detection of emotion via facial recognition has been becoming as an active research topic. We are interested to analyze the emotional state, happy or not happy through the identification of the happy status in a photo retrieved from social media networks. In our work, three innovative methods are proposed. The first one used the slopes of control points for comparison; while the other two adopted the artificial neural network (ANN). For these two ANN methods, one used control point coordinates as input, and the other one used the black pixel percentage as input. Before processing an image with these three methods, we apply some pre-processing steps to achieve better results. First, with a given image, a binary image would be obtained for detecting the face correspondingly. Then, eyes and lip are identified. These features are normalized and boundary curves of the respective regions are estimated. In order to obtain the data for comparison or input to ANN, our methods would locate the control points and partition the facial image. Experiments have been conducted to compare the recognition results of these methods with different images datasets. The results show that the slope comparison and the control points ANN methods have better performance when compared to the black pixel percentage ANN method.

***Keywords*:** facial recognition, happy, smiling social media

1. Introduction

Human face plays an important role in social media interactions for expressing different personal emotion, and recognition of facial emotion has been acknowledged as an active research topic recently. Its application can be applied to many areas, like smart home, intelligent robots, metal disease study and criminal detection. The challenges include face recognition, data representation, emotion detection scheme, etc. In the area of face recognition, there already have a large number of methods existed, and most of them have good performance. Hence, our work focuses on the emotion detection and assumes that the head image of a person has been successfully captured.

For data representation, control points located in the boundary curve can effectively represent the relative position and the outline shape of eye or lip. Also, after performing image binarization, black pixel percentages in different parts of the face in a binary image can reflect the shape and shadow for eye or lip. One idea is to use these two features for analyzing the emotion from the facial image, which has not been investigated before.

In order to improve the emotion detection scheme, artificial neural network can be used to classify the emotion from the image. In our work, control points coordinates and black pixel percentages will be used as the input to the network. In addition, a dissimilarity measure for calculating and comparing the relative position for the control points will be proposed. This method can also be used to recognize the input data. Therefore, totally there are three recognition methods developed: using control points as input to neural network, using black pixel percentage numbers as input to neural network and comparing the slope of lines connecting the control points.

The next section discusses some of previous works in this area. Section 3 will then discuss the main idea of our work and it is followed by Section 4 on the design methodology. Section 5 presents experimental results and Section 6 concludes our work afterwards.

2. Background Review

Currently, there exist a number of methods to analyze the emotion from a facial image. In this section, some of the existing algorithms will be summarized and presented. Kharat and his group have a "Humanoid Robots" that can conduct intellectual conversation with humans (Kharat et. al. 2008). In their project, Statistical Parameters and Discrete Cosine Transform (DCT) are used for feature extraction. Generalized Feed Forward Neural Network (GFFNN) and Multilayer Perceptron (MLP) are employed and their performance is compared. Totally six emotions namely happy, surprise, disgust, angry, fear and sad are recognized. This project considers a wide range of emotions and claims to achieve a 100% recognition rate, but the size of the data set is limited (210 photos), and the data are not classified to compare the result.

In 2012, Kudiri and others proposed a study in extraction of feature based on several positions of facial muscles (Kudiri et.al. 2012). Relative sub-image based features are proposed to improve detection performance. They use support vector machine to implement an automated emotion detection system for classification of facial emotions. Their work required only a relatively low computational effort, but does not classify the experiment data for comparison. One of the early works of artificial neural network dealing with human emotions is in (Kharat et. al. 2012). On the other hand, researchers started to adopt genetic algorithm to classify lip features using lip features such as (Rizon et.al 2007). For their work, only Japanese are considered. They proposed three extraction methods of lip feature and compared their respective performances to determine the feature of the lips. Different emotions from the subject result in unique characteristic of lips. Genetic algorithm was applied to optimize the lip features ellipse characteristics. They suggested two equations of ellipse fitness, as well as the parameters that can define the emotion. However, this project only considered the lip features, and the testing data were limited to Japanese people.

In (Ghandi et. al. 2010), the authors modified the particle swarm optimization (PSO) to a guided particle swarm optimization (GPSO) algorithm for detection of facial emotions. Action Units (AUs), which are a set of points located on the face of a person, are used by the GPSO. They also define the swarm of particles to ensure that within the neighbourhood of each unit a particle would have a component. The GPSO is implemented and was tested to detect six emotions. However, application of the system was limited to pre-recorded video clips and the calculation is complicated.

All previously discussed works do not consider classifying the testing data for comparing their recognition methods. We note that different methods may have different performance in different segments of people such as gender, age and skin colour.

3. Overall System Design

Our work focuses on analyzing and recognizing the emotion for a person from his/her facial image. A facial emotion detection system will be developed to analyze and recognize the emotions from facial images. We have adopted the approach done by an online program developer of a human emotion detection system to (codeproject 2010) as shown in Fig 1. In the system, it first takes an image. Then it performs skin colour segmentation to detect human skin colour. With the detection of a human face, the system continues to separate the eyes and lip from the face. Next, Bezier curves for eyes and lips will be calculated. These curves will be used to compare with those stored in a pre-processed database for finding the nearest Bezier curves which represent the emotion of a face. We adopt the main structure, as well as some basic image processing and the user interface in the system to complete our development.

Our design improves the existing workflow of the image process pipeline, the binary conversion method, the method of obtaining the eye/lip, the curve drawing, the control point location, but using our own emotion detection approach. In addition, we develop the modules of obtaining the black pixel percentage numbers, and three emotion detection methods.

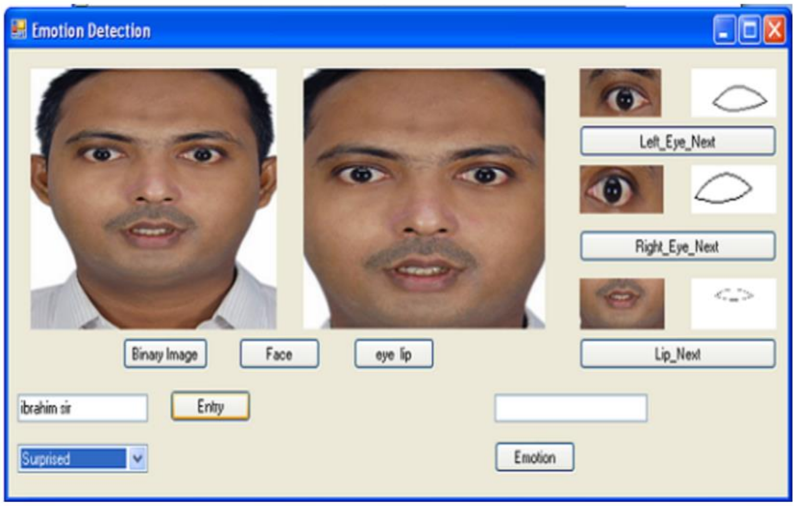

Fig. 1. Emotion Detection System

The pipeline of the operations of our design is shown in the flowchart as in Fig 2. An image will first be retrieved by the system to generate a binary image. After that, the corresponding face will be extracted from the image, and the system will locate 3 features: the left eye, right eye and lip, respectively. For each feature, the system will draw an outline curve and calculate the black pixel percentages. Then, the control points are calculated. Finally, the system can run the Control Points Slope Comparison, Control Points Artificial Neural Network and Black Pixel Percentage Artificial Neural Network to analyze and detect the emotion. For the whole process flow, the red boxes are related to the image processing, the green boxes are related to mathematical calculation, and the black boxes are related to the detection methods.

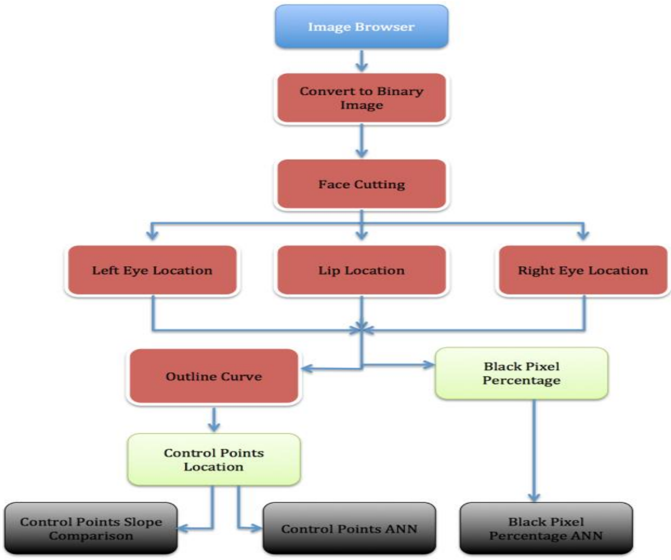

Fig. 2. Operation Pipeline

4. Emotion Detection Methods

In this section, we present the three methods for analyzing and recognizing the emotion from facial image and they are:

- Control Points Slope Comparison
- Control Points Artificial Neural Network
- Black Pixel Artificial Neural Network

4. 1. Control Points Slope Comparison

As discussed in Section 1, control points located in the Bézier curve can be used to detect the emotion. Each eye or lip curve will have 6 control points. The first and last control points are the leftmost and rightmost control point. There are two control points on the upper part and two are on the lower part. With these control points, we can calculate the similarity between a sample with a standard set for guessing a specific emotion (e.g. happy). For example, the similarity between the sample and the happy set would be

$$\sum_{i=1}^{i=6} |\mathrm{sam}_{\mathrm{lei}} - \mathrm{hap}_{\mathrm{lei}}| + \sum_{i=1}^{i=6} |\mathrm{sam}_{\mathrm{rei}} - \mathrm{hap}_{\mathrm{rei}}| + \sum_{i=1}^{i=6} |\mathrm{sam}_{\mathrm{lipi}} - \mathrm{hap}_{\mathrm{lipi}}|$$

However, this method has not considered that the relative positions of the control points which may remain the same while the absolute distance between the control points may vary. Very often, we observed that the emotion is only related to the relative position of the control points, and all other factors should be eliminated.

Here, we propose to use the relative positions of the control points, through calculating the slopes of lines connecting the control points and then works out the dissimilarity between these slopes and a standard set of slopes for comparison. As shown in Fig 3, for each outline curve of eye/lip, six control points are located, and six lines connecting the control points are drawn. We will calculate the slopes for each line, and totally 18 (6*3) slopes will be obtained. The dissimilarity is calculated by summing up the absolute distance between each slope in the sample and in the standard set. A standard set of slopes of happy and a standard set of slopes of sad are computed. Next, the emotion detection can be based on the dissimilarity calculation.

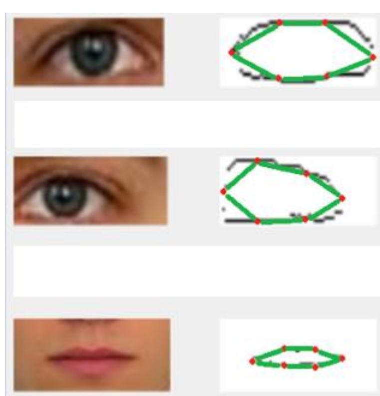

Fig. 3. Slope Calculations with Control Points

4. 2. Control Points Artificial Neural Network

This method uses the artificial neural network to determine the emotion of a person. The inputs to the artificial neural network are the control points generated by the first method. Then, a number of image samples with designated emotion will be used as the training data. For each sample, the 6*3 = 18 control points' x and y coordinates will be used as the input for the neural network. During the training procedure, the samples will be used recursively for adjusting the neuron parameters until the error between the intermediate output and the designated output is small enough. After the neural network is trained and a sample image is processed, the control points extracted from the sample will be input to the network. Then, the emotion can be determined based on the output of the network.

4. 3. Black Pixel Artificial Neural Network

This method also uses the artificial neural network to analyze and detect the emotion. Unlike the previous two methods uses a list of percentage numbers as the input. To get the input to the artificial neural network, for each sub-image of left eye, right eye and lip, we normalize each image to a bitmap of 160 pixels * 80 pixels. Each of the images will be divided into 8 regions, 4 columns times 2 rows as shown in Figure 4.

For each region (40 pixels * 40 pixels) in the image, the percentage of black pixels will be calculated. After going through all the three images, 24 (8*3) percentage numbers will be generated. Each of them indicates the black pixel percentage in a particular region. These 24 percentage numbers will then be used as the input of a sample to the neural network. A number of image samples with designated emotion will be used as the training data. During the training procedure, the samples will be used recursively for adjusting the neuron parameters until the error between the intermediate output and the designated output is small enough. After the neural network is trained and a sample image is processed, the black pixel percentages extracted from the sample will be input to the network. Then, the emotion can be determined based on the output of the network.

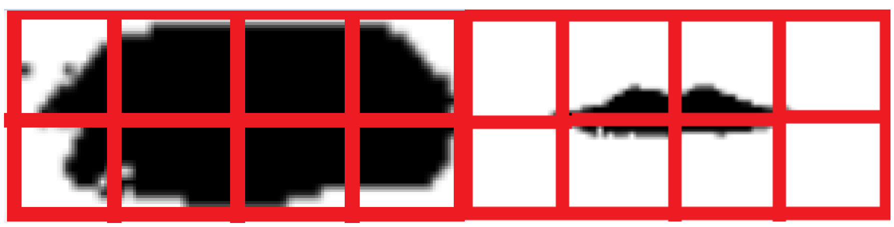

Fig. 4. Image Region Division

5. Experiments

In our experiment, image data are collected from different sources. Some of them are from Facebook (21 photos), some of them are from Google (42 photos), some of them are taken from other students in the University (40 photos) and some of them are photos of first author (113 photos). The photos taken by our research team have been used as the training set while other photos will be used as the testing set (as shown in Fig 5). This was because these photos have better light condition, clear face and more smooth hue, and produce more a reliable inputting result for comparison and training. The photos from other sources generally have worse quality in light condition and hue, but reflect the common scenario. Therefore, they are used as the testing set.

For each sample in the training set, the program will get the binary image, detect the face, locate the eyes and lip, normalize the eyes and lip, get the boundary curves for eyes and lip, locate the control points and divide the image. The emotion (happy or sad) will be decided manually and input into the database with the coordinates of control points and the black pixel percentage numbers.

During the training process, some of the photos may be not clear enough to be detected by the program, and these photos will be ignored by the program. The images are divided into a variety of groups, according to different classification characteristics. There are groups of adult and children (Figure 6), male and female as well as white people and colour people. Below are some samples from the data. For each group of the images, the program will be run once, and the accuracy of the three methods will be recorded. For each image inside a group, the program will get the binary image, detect the face, locate the eyes and lip, normalize the eyes and lip, get the boundary curves for eyes and lip, locate the control points and divide the image. After a series of image processing, the program will connect to the database get the pre-stored data for comparison (Control Points Slope Comparison) or training (artificial neural network).

Fig. 5. Testing Set

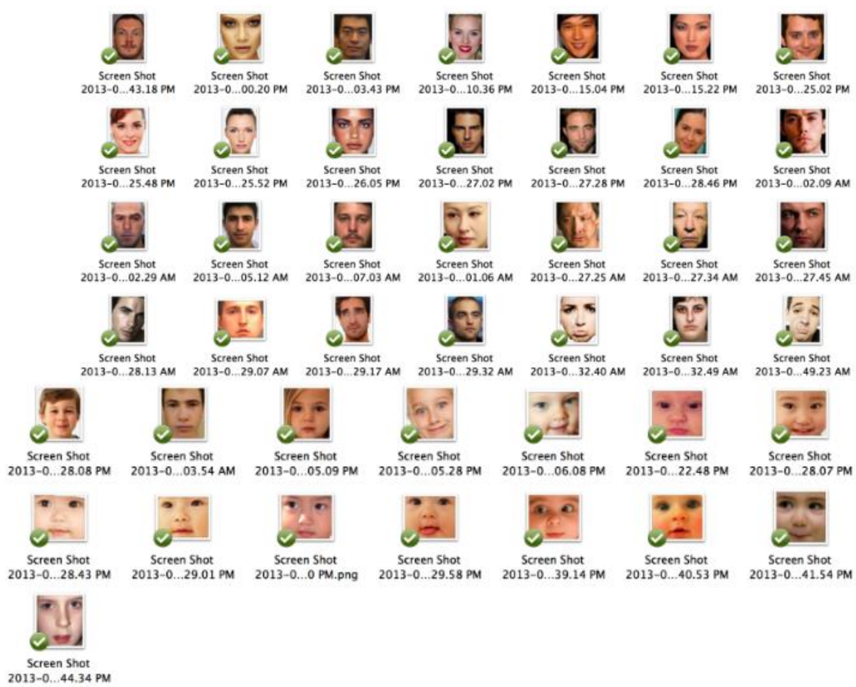

Fig. 6. Images of adults and young people

Generally, Control Points Artificial Neural Network (CPANN) and Control Points Slope Comparison (CPSC) performed better than the Black Pixel Percentage Artificial Neural Network (BPPANN) as shown in Table 1, 2, 3 and 4. CPANN performed better in all the groups except children while CPSC performs best in the children group. This is probably because children's facial features are relatively small and it brings difficulty to the control point location, while the relative position of control points will not be affected. The black pixel percentage method performs the worst among the three methods. This may be because this method does not consider the relative position of the eye/lip inside the image box. Due to the translation of the eye/lip inside the image box, the black pixel percentage will be largely influenced, and this will result in bad recognition obstacle for artificial neural network.

Table. 1 Adult (sample size 86)

	Happy (CPANN)	**Sad (CPANN)**	**Happy (CPSC)**	**Sad (CPSC)**	**Happy (BPPANN)**	**Sad (BPPANN)**
Total	39	47	39	47	39	47
Successfully recognized	30	37	27	34	20	28
Fail to recognize	9	10	12	13	19	19
Precision	75%	80%	67.5%	73.9%	51.3%	59.6%
Recall	76.9%	78%	69.2%	72.3%	51.3%	59.6%

Table. 2 Children (sample size 15)

	Happy (CPANN)	**Sad (CPANN)**	**Happy (CPSC)**	**Sad (CPSC)**	**Happy (BPPANN)**	**Sad (BPPANN)**
Total	7	8	7	8	7	8
Successfully recognized	5	5	5	7	3	2
Fail to recognize	2	3	2	1	4	6
Precision	62.5%	71.4%	83.3%	77.7%	33.3%	33.3%
Recall	71.4%	62.5%	71.4%	87.5%	42.8%	25%

Table. 3 Male (sample size 49)

	Happy (CPANN)	**Sad (CPANN)**	**Happy (CPSC)**	**Sad (CPSC)**	**Happy (BPPANN)**	**Sad (BPPANN)**
Total	20	29	20	29	20	29
Successfully recognized	16	20	16	18	12	15
Fail to recognize	4	9	4	11	8	14
Precision	64%	83.3%	59.3%	81.8%	46.2%	65.2%
Recall	80%	68.9%	80%	62.1%	60%	51.7%

Table. 4 Female (sample size 48)

	Happy (CPANN)	**Sad (CPANN)**	**Happy (CPSC)**	**Sad (CPSC)**	**Happy (BPPANN)**	**Sad (BPPANN)**
Total	26	22	26	22	26	22
Successfully recognized	18	14	18	14	15	12
Fail to recognize	8	14	8	14	8	14
Precision	69.2%	63.6%	69.2%	63.6%	60%	52.2%
Recall	69.2%	63.6%	69.2%	63.6%	57.7%	54.5%

When compared to (codeproject 2010), the former two methods achieve a relative higher recognition rate. Since the (codeproject 2010) is developed by the programmer from south Asia, it can only recognize people in darker skin color. Therefore, only the data provided by that project (40 photos) are used to test that system. It achieves the correct rate of 61.3%.

6. Conclusion

In our work, three facial image data processing and analyzing methods have been presented with some experimental results. Generally, they have achieved a relatively high recognition rate, and the computational effort is low compared to other methods. However, the size of the data sample is still quite limited, and the image-processing phase will be influenced by the quality of the images. The recognition rate may be improved by increasing the number of control points or black pixel sections in the image. In addition, our work can be extended to customize emotion recognition. Each person can have his/her own set of happy/sad standard data for comparison or training. Then, the system can treat each person individually.

References

Bashir Mohammed Ghandi, R. Nagarajan and Hazry Desa, Real-Time System for Facial Emotion Detection Using GPSO Algorithm, 2010 IEEE Symposium on Industrial Electronics and Applications (ISIEA 2010), October 3-5, 2010, Penang, Malaysia, pp 40-45.

G. U. Kharat and S.V.Dudul, Neural Network Classifier for Human Emotion Recognition from Facial Expressions Using Discrete Cosine Transform, First International Conference on Emerging Trends in Engineering and Technology , 16 – 18 July 2008, pp 653-658.

Govind U.Kharat and Sanjay V Emotion, Recognition from Facial Expression Using Neural Networks, International Journal of Soft Computing and Engineering (IJSCE), ISSN: 2231-2307, Volume-2, Issue-1, March 2012, pp 224-227.

Krishna Mohan Kudiri, Abas Md Said and M Yunus Nayan, Emotion Detection Using Sub-image Based Features through Human Facial Expressions, 2012 International Conference on Computer & Information Science (ICCIS), pp 332-335.

Mohamed Rizon, M. Karthigayan, Sazali Yaacob and R. Nagarajan, Japanese Face Emotion Classification Using Lip Features, Geometric Modelling and Imaging (GMAI'07).

Web sites:

Web-1: http://www.codeproject.com/Articles/110805/Human-Emotion-Detection-from-Image, 2010

Web-2: http://www.codeproject.com/Articles/16447/Neural-Networks-on-C, 2010