Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 125 (2018) 2–10

Procedia Computer Science

www.elsevier.com/locate/procedia

6th International Conference on Smart Computing and Communications, ICSCC 2017, 7-8 December 2017, Kurukshetra, India

An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment

D. Yang[a], Abeer Alsadoon[a], P.W.C. Prasad[*a], A. K. Singh[b], A. Elchouemi[c]

[a]*School of Computing and Mathematics, Charles Sturt University, Sydney, Australia*
[b]*Department of Computer Applications, National Institute of Technology, Haryana, India*
[c]*Walden University, USA*

Abstract

The purpose of this study is to introduce a method based on facial recognition to identify students' understanding of the entire distance learning process. This study proposes a learning emotion recognition model, which consists of three stages: Feature extraction, subset feature and emotion classifier. A Haar Cascades method is used to detect the input image, a face, as the basis for the extraction of eyes and mouth, and then through the Sobel edge detection to obtain the characteristic value. Through Neural Network classifier training, six kinds of different emotional categories are obtained. Experiments using JAFF database show that the proposed method has high classification performance. Experimental results show that the model proposed in this paper is consistent with the expressions from the learning situation of students in virtual learning environments. This paper demonstrates that emotion recognition based on facial expressions is feasible in distance education, permitting identification of a student's learning status in real time. Therefore, it can help teachers to change teaching strategies in virtual learning environments according to the student's emotions.

Peer-review under responsibility of the scientific committee of the 6th International Conference on Smart Computing and Communications.

Keyword: facial expression, facial recognition, emotion recognition, distance education

* Corresponding author
E-mail address: CWithana@studygroup.com

1877-0509 © 2018 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the scientific committee of the 6th International Conference on Smart Computing and Communications
10.1016/j.procs.2017.12.003

1. Introduction

In the field of education, existing virtual learning environments successfully simulate interaction at a cognitive level during traditional teaching processes [1]. In the process of human-computer interaction, teachers naturally expect a facial recognition system to have the ability to detect, analyze and process emotions in order to get good teaching effect, such as perception, understanding and expressing emotions [2] If in virtual learning environments, students show different expressions in videos, teachers can identify whether the students understand their teaching content according to students' different expressions, and can adjust their teaching programs.

In virtual environments, students' emotional data can be obtained from a series of video frames. Therefore, to achieve real time understanding of students' emotions, accuracy and time needs to be balanced. To improve accuracy, more video frames are required, increasing computation time. On the other hand, in pursuit of efficiency, we would reduce the accuracy which means collection less feature data. Experimental results from current solutions show that in some cases accuracy is higher, but time spent is also high, such as Salman et al [3] who used Decision Trees to identify Facial Expressions. Other solutions have different limitations; for example, Chu et al [4] provided low-cost facial expressions on a mobile platform by just extracting lip features, spending less time but achieving less accuracy.

To find the best solution for emotion recognition in learning virtual environment, both accuracy and efficiency needs to be achieved. Optimization of recognition of emotional changes in the facial features of online learners, will allow teachers to adjust teaching strategies and methods, give real-time feedback to students, and achieve the best teaching quality.

The goal is to find the best solution for emotion recognition based on facial recognition in virtual learning environments, in real time. To achieve this goal, it is necessary to improve the accuracy and efficiency of facial recognition systems.

2. Literature Review

The goal of emotion recognition is to gather data and analyse feelings of subjects, to make appropriate responses possible. Such data may be obtained from different physical features such as face voice, body movements, and other biological physical signals. However, learners' emotions are expressed first through facial expressions which can be divided into six kinds of categories: sadness, happiness, surprise, fear, anger and disgust. Fig 1 lists the emotion recognition process.

The facial recognition process consists of three main stages: acquisition, feature extraction, and emotion classification. This paper is organised along the three stages.

Facial acquisition

Facial localization is the process of determining whether a face is included in an image, and its location and size [5]. Due to the use of different features, the facial localization has a variety of methods.

The face edge can be used to complete the facial localization [6]. In addition, facial texture information is often used for facial localization; Dai et al [7] proved facial localization using facial grey scale. Skin colour has special properties; it is in a relatively independent position in the colour space, so skin colour information is often used to carry out facial localization [8].

2.1 Facial feature extraction

Facial feature extraction methods are mainly dependent on the classification method and the application environment. Different classification methods need different characteristics, and the different methods can be applied to different environments. Thus, Durmusoglu [9] used the 19 possible 153 key landmarks and distance between the candidates.

2.2 Emotion expression classification

The main differences of classifier selection and design are whether time information is used. The classification method without using temporal information can be called spatial domain method. Artificial neural network is a typical spatial method. The whole image is used as input of neural network or by image processing, such as image Gabor filtering, or through the feature representation of image processing, such as PCA and ICA [10]. Due to the use of feature vector space method, the general classification method can be used as a spatial method. Gosavi et al [11] used principal component analysis for facial expression recognition. Kiran et al [12] and Quraishi et al [13] used support vector machine (SVM) as the classifier.

2.3 Current emotion detection system

Quraishi et al. [14] provided emotion detection that indicates the need for a desired direction of the facial image. Table 1 shows the details information of current emotion detection system.

Table 1 Current emotion detection system

Method	Author	Tested Images	Correct Recognition	Wrong Recognition	Performance (%)	Processing Time(sec)
A frame-work for the Recognition of Human Emotion using Soft Comput-ing models	Quraishi, l.,M., Choudhury, P.,J., De, M., & Chakraborty, P.	20	14	6	70	105

The whole flow chart is shown in Fig. 1. Eyes have been used to extract the origin of facial emotional characteristics. A method of filtering and edge detection for feature extraction has been proposed. Then, the processed image has been used to identify some optimal parameters by genetic algorithm (GA). A new fitness function for extracting eye parameters by GA has been proposed. Because the emotion detection algorithm can be used as an expert system for continuous processing of the available data, the proposed method of emotion estimation is considered to be suitable for the personalized face. From the obtained training data set, the back-propagation neural network is used to identify the emotion of the person.

3. Proposed Work

The best solution offered a solution for the identification of emotion using facial features. However, at the pre-processing stage, the solution is only to crop the eye region and grab the characteristic value, which causes the results of emotion recognition to be lower. Therefore, to improve accuracy, this process needs to be transformed.. Fig.2 shows the whole proposal solution procedure. In order to improve the accuracy, mouth features are added using Haar Cascades.

We use the Haar Cascades method to identify whether a face exists in the images, and if the face does not exist, then return to the start and input the image frames. If the face exists, eyes and mouth need to be located and eye and mouth regions need to be cropped. Filter and edge detection is carried out using Sobel edge detection method, followed by feature extraction. We train the feature extraction using the neural network method.

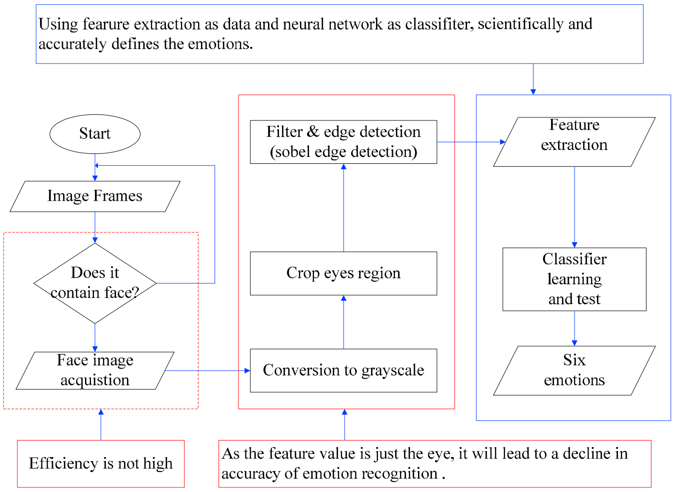


Fig..1. Current best solution

We have used a database of 20 images of 6 emotion classes for training of the neural network. In the first phase the inputs are propagated through the layers of processing elements, generating an output pattern in response to the input pattern presented. In the second phase, the errors calculated in the output layer are then back propagated to the hidden layers where the synaptic weights are updated to reduce the error. This learning process is repeated until the output error value, for all patterns in the training set, is below a specified value.

A neural network based on the optimal range of human eyes is proposed. The optimized values of the previously obtained data are used as inputs to the network. The network is a 1 input neuron structure model, the 1 hidden layer has 20 neurons and 7 output neurons. The input is the output of the estimate is mutually exclusive of the binary bits representing an emotion.

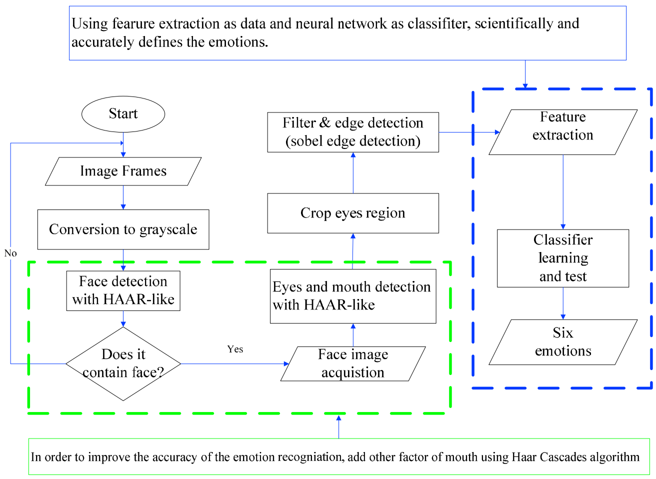


Fig. 2. Proposed solution

3.1 Eyes and mouth detection using Haar Cascades

Sablik, Velten, and Kummert [15] use object cascade classifier Haar feature detection based on Paul Viola and Michael Jonesin their paper "Rapid Object Detection using a Boosted Cascade of Simple Features" in the effective object detection method. It is a machine learning method, which is based on a large number of positive and negative image training cascade functions. Then it is used to detect objects in other images.

For the proposal solution, firstly we train input images, get the face features and store these in face.xml; secondly, we detect eyes and mouth based on facial features. The details below show the algorithm for eyes and mouth detection.

Algorithm : Eyes and mouth detection
INPUT: Test image samples images (I) = {I1, I2, …,I_n}
OUTPUT: Detected eyes and mouth

BEGIN
Step 1: Input the samples images
(I) = { I1, I2, …, I_n }
Step 2: Read the image and store to the frame
F = { I1, I2, …, I_n }
Step 3: Read the contents of the xml file for face detection and store in the memory
R = face.xml //read face.xml, it stores in OpenCV lib.
Step 4:
i= {1,2,3,…,n}
fi ={f1, f2,…f_n}
START loop
F_0 = 1
i = 0
While F_i > F(target) and i < n(Stages)
i = i + 1
Train classifier for stage(i)
Initialize weights
Normalize weights
Pick the (next) best weak classifier
Update weights
Evaluate F_i
if F_i > f
go to Normalize Weights
Combine weak classifiers to form the strong classifier stage
compute F_i
Step 6: End of the algorithm

3.2 Filter & edge detection

Mean and median filters have been applied to make the image smoother by removing unwanted noise that has already been provided. It returns the edge at the point where the gradient of the input images is at the maximum. The details below show the algorithm 2 of eyes and mouth edge.

4. Results and Discussion

This section tests the implementation of the proposed emotion recognition system based on facial recognition. This proposed solution uses a python 2.8 platform with OpenCV 2.0 lib. Due to OpenCV 2.0 with many functions, it is easy to deal with image transactions. The application imports the OpenCV library to solve the computer vision process. OpenCV library method is mainly used to detect the object's joint data and tracking objects.

For testing the images, this paper adopts JAFFE database. The database contains 213 images (each image: 256 * 256 pixel) of Japanese women’s faces, with each image defined in the original expression. There are 10 people in the expression library. The JAFFE database shows frontal faces, and the original image to re adjust the trim, making the eyes in the database image in roughly the same location; facial dimensions are basically the same. Because this expression database is completely open, and the expression calibration is standardised, most of the articles used to study expression recognition are used to train and test.

```
Algorithm 2: eyes and mouth edge
INPUT: Detected image(I)
OUTPUT: Edged image(I)
BEGIN
      X={1,2,3,...n}
      C is that point the image in the frame
      S is that sotre the image
      X=0, C = 0, S=0
Step 1: Read the detected image from the system memory
      R = Detected image(I)
Step 2: Detect in the image
   START loop
      For X = 1 to n
         Read each pixel in In
         Point the image in the frame
            C = Sobel (In) // import sobel function in OpenCV
lib
            Store the image in the system directory
                  S = Store(C)  // import store function in
                                // OpenCV lib
    End of For loop
Step 3: End of the algorithm 2
```

Table 2 shows that the results from the proposed solution and the current solution, identifying differences in accuracy and efficiency. This test uses three groups of data, with each group of data used in six expressions (happiness, sadness, surprise, anger, disgust, fear), the last group of data shows only happy and sad expressions. There is a total of 20 images; accuracy was expressed as percentage; efficiency is expressed in seconds. Detailed information is shown in Table 4.

From table 2, we get the average accuracy for six kinds of emotions, and details are shown in table 3. Table 3 demonstrates that the recognition rate of happiness was significantly higher than the others, followed by the emotion of surprise, while the lowest is sadness. The characteristic value of happiness is obvious and the characteristic value of other facial expressions is prominent, but the characteristic value of sadness is not particularly prominent, which often contains other facial expression characteristic values, leading to a low recognition rate.

From table 2, we get the average processing time (sec) of each emotion, and details are shown in table 4. Overall, whether the best solution or the proposed solution, the processing time spent is relatively long; this is a certain distance from the commercial use. In terms of the average processing time, the proposed solution is more efficient than the best solution. The longest time is needed for surprise recognition, followed by fear recognition, while the shortest is happiness recognition, and anger and disgust recognition are basically the same time.

Overall, because the mouth was added as an identification elements using Harr-like method, accuracy has improved. From the results, it proves that the proposed solution is better than the current solution.

Table 2 Compare results of proposed solution and current solution

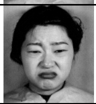
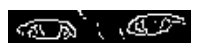
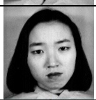
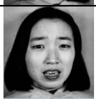
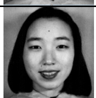
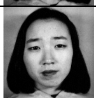
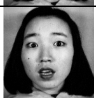
results							
Sample Images	**Proposed Solution (Haar Cascades) Eyes and Mouth**				**Current Solution (Sobel Edge Detection Eyes)**		
	Output	**Detected Emotion**	**Accuracy (%)**	**Processing Time(sec)**	**Output**	**Accuracy (%)**	**Processing Time(sec)**
Group 1							
		Sad	80.5	138		79	142
		Disgust	89.2	113		89	115
		Fear	84.3	158		83	164
		Anger	87	113.5		82	118
		Disgust	81	115.5		80	119
		Fear	79	155.5		75	161
		Happy	90	115		89	122
		Sad	75	131		72	143
		Surprise	89.8	165		87	172
		Happy	93	117		90	124
		Sad	78	132		73	150

Table 3 Average accuracy(%) of each emotion

	Sad	**Surprise**	**Happy**	**Anger**	**Disgust**	**Fear**
Proposed solution	78.54	93.26	95.25	91.22	84.32	82.58
Current solution	76	87.72	94	87.66	82.76	79.73

Fig..3. Average accuracy(%) of each emotion

Table 4 Average processing time(sec) of each emotion

	Sad	**Surprise**	**Happy**	**Anger**	**Disgust**	**Fear**
Proposed solution	130.25	163.66	114.5	112.16	114.16	154.16
Current solution	143.62	170.5	121	116	117	159

5. Conclusion

This paper provided a proposed model to solve the problems of emotion recognition based on facial recognition in virtual learning environments, and the efficiency and accuracy are considered at the same time. Using HAAR Cascades to detect eyes and mouth and identify all kinds of emotion through the neural network method, the combination of efficiency and accuracy is achieved. It can be applied to real distance education. The application of emotion recognition in virtual learning environments is a much-researched topic. In addition to the change of uncertainty factors makes teachers and students face pattern is more complex, so the emotion recognition in the online learning network application mode is a very challenging topic. Accordingly, it is proposed that the subject requires further research from the following aspects:

1) Due to the fact that this research does not involve the illumination and pose of the image, it is uncertain how much these factors influence facial expressions and thus the final emotion recognition. All of these issues need to be explored in future research and validated by experiments.

2) To make the theory and technology of emotion recognition fully meet the practical requirement, there are suggesting that the comprehensive application of image processing, pattern recognition, computer vision and neural networks, psychology, cognitive science, and integrates with other biometric authentication methods and methods of human-computer interaction perception based on the in-depth and meticulous research work.

REFERENCES

[1] Das (2015) "Virtual Classroom For Effective Learning in IT Industry", International Conference on Information Technology IEEE. 221-226 .

[2] Farsi, Munro, & Thobaiti (2015) "The Effects of Teaching Primary School Children the Islamic Prayer in a Virtual Environment", Science and Information Conference. 765-769 .

[3] Salman, Madaini, & Kissi (2016) " Facial Expression Recognition using Decision Trees". IEEE International Conference on Computer Graphics, Imaging and Visualization (CGiV), 120-130.

[4] Chu, Chen, & Hsihe (2015) " Low-cost Facial Expression on Mobile Platform", IEEE International Conference on Machine Learning and Cybernetics (ICMLC), pp 586-590,.

[5] Yang, Kriegman, & Ahuja (2002) "Detecting Faces in Images: A Survey", IEEE Transactions on Pattern Analysis and Machine Intelligence. 24(1) 34 – 58.

[6] Yow, & Cipolla (1997) "Feature-based Human Face Detection", Image and Vision Computing, 15(9), 713-735.

[7] Dai, & Nakano (1996) "Face-texture Model Based on SGLD and its Application in Face Detection in a Color Scene", Pattern Recognition. 29(6), 1007-1017.

[8] Yang, & Waibel (2015) "A Real-time Face Tracker", Proceedings of 3rd Workshop Applications of Computer Vision, 142-147.

[9] Durmusoglu, & Kahraman (2006) "Facial Expression Recognition Using Geometric Feautres", IEEE. DOI: 10.1109/IWSSIP.2016.7502700, 23-26.

[10] Rowley, Baluja, and Kanade (2015) "Neural Network-based Face Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence. 20(1), 23-38.

[11] Gosavi, Khot (2013) "Facial Expression Recognition Using Principal Component Analysis", International Journal of Soft Computing and Engineering (IJSCE). 3(4), 258-262.

[12] Kiran, Kushal (, 2016) "Facial Expression Classification using Support Vector Machine Based on Bidirectional Local Binary Pattern Histogram Feature Descriptor", 17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). 115-120.

[13] Vosian, Precup, Dragan (2016) "Facial Expression Recognition System Based on a Face Statistical Model and Support Vector Machines", IEEE 11th International Symposium on Applied Computational Intelligence and Informatics (SACI), 63-68 .

[14] Quraishi, Choudhury, Dept. & Chakraborty (2012) "A Framework for the Recognition of Human Emotion using Soft Computing Models", International Journal of Computer Applications. 40(17), 50-55.

[15] Sablik, Velten, Kummert (2013), "Automatic Privacy Protection Based on Haar Cascades and Spectral Domain Watermarking", VDE VERLAG GMBH, ISBN: 978-3-8007-3543-3, pp 199-204.