An E-learning System With Multifacial Emotion Recognition Using Supervised Machine Learning

Ashwin T S, Jijo Jose, Raghu G, G Ram Mohana Reddy
Department of Information Technology
National Institute of Technology Karnataka
Surathkal, Mangalore, India
{ashwindixit9, jijojosenitk@gmail.com, raghugolla.22@gmail.com, profgrmreddy}@gmail.com

***Abstract*— E-Learning systems based on Affective computing are popularly used for emotional/behavioral analysis of the users. Emotions expressed by the user is depicted by detecting the facial expression of the user and accordingly the teaching strategies will be changed. The present eLearning systems mainly focus on the single user face detection. Hence, in this paper, we propose multiuser face detection based eLearning system using support vector machine based supervised machine learning technique. Experimental results demonstrate that the proposed system provides the accuracy of 89% to 100% w.r.t different datasets (LFW, FDDB, and YFD). Further, to improve the speed of emotional feature processing, we used GPU along with the CPU and thereby achieve a speedup factor of 2.**

General Terms — CPU, GPU

Keywords—Affective Computing, eLearning, Facial expression recognition, Machine Learning, Active Appearance Model, Local Binary Patterns

I. INTRODUCTION

E-Learning plays a vital role in modern education system where the students learn and interact with the computer or the digital technology to gain required knowledge. In the traditional learning due to face to face interaction, teachers were able to interact with the students according to the facial expressions. This became one of the major demerits of the eLearning system. Affective computing thus helps to overcome this major demerit of traditional classroom based teaching-learning process and thereby improving the eLearning system.

Affective computing is an emerging area in computer science and information technology in which emotional, behaviour and stress analytics are carried out by multimodal features such as facial, gesture, posture, speech etc. There are many methods for detecting the emotion of the users. Each emotion expressed by the user is directly dependent on the teaching strategies. Hence, the accurate detection of the emotion plays a vital role in analysing the efficiency of eLearning system. The accurate detection of emotions is done by using computer vision and the classification by local binary patterns and machine learning techniques [1].

There are many frameworks available in the computer vision for matching a statistical model of object shape and appearance to a new image. One among them is active appearance model (AAM) [2] which are built during the training phase. A set of images, together with coordinates of landmarks that appear in all of the images, is provided to the training supervisor. Active Shape Model (ASM) [2] is also a statistical model in which image deforms itself to map to an existing data and Active Contour Model (ACM) [2] trace the shape of an object outline from a possible noisy 2D image.

Affective computing tools and machine learning techniques are popularly used to detect the emotions accurately not only for single user face in a single frame but also for the multi user faces in a single frame. Further to speed up the processing, we can use both Central Processing Unit (CPU) and Graphics Processing Unit (GPU).

The existing eLearning systems concentrate more on one to one interaction between the student and the eLearning tool, hence the emotion recognition is done for the single user face in a single frame. Our proposed method detects the multi user face emotion recognition for a group of eLearning students (in a single frame). We also tested multi user face emotion recognition on four different set of images namely original/normal, aligned, funnelled and deep funnelled. Once the detection of emotion is done then we recommend this analysed group emotion to modify the teaching strategy accordingly. The use of GPU along with the CPU significantly increases the speed of processing.

The main/key contributions in this paper are,

- To the best of our knowledge, this is the first work on multi user face detection for eLearning systems.
- Further, the processing speed of the eLearning system is augmented by using both CPU and GPU.

The rest of the paper is organised as follows. Section II outlines the background and related work, Section III describes the proposed methodology in detail, Section IV deals with experimental results and analysis, then finally we conclude with future enhancements in Section V.

II. BACKGROUND AND RELATED WORK

Facial expression recognition is the key feature of emotional analytics. These emotions are classified into various categories and sub categories [3]. The summary of the major classification in emotion is as shown in the Table I.

Jingjing Chen et al. [4] proposed facial expression recognition system which uses Euclidean distance in 6D space to recognize the facial expression and a classification tree that

978-1-4673-9509-0/15 $31.00 © 2015 IEEE
DOI 10.1109/T4E.2015.21

Authors	Methodology and Approch		Limitations
	Based On	Details	
Russell et al. [8]	Reaction	Emotions are seen as combinations of arousal (high activation/low activation) and valence (positive/negative).	Not suitable for teaching stratergy
Ortony et al. [10]	5 Basic Emotion	Authors have proposed OCC model based on 5 basic (anger, fear, happiness, joy, love) and 14 secondary emotions.	Surprise Element will be missed for teaching statergy
Csikszentmihalyi [5]	Zone	Author had identified a zone, If a task is not challenging enough, boredom sets in, while too great a challenge results in anxiety, and both cases result in task, and thus learning, avoidance.	Intermediate emotions are missing
Pekrun et al. [4]	Academic Emotions	Authors have examined the impact of the so-called academic emotions (four positive: joy, hope, pride, relief and five negative: boredom, anger, anxiety, shame, hopelessness)	emotions like relief and shame are not necessary for teaching stratergy
Damasio et al. [9]	Four Basic Emotions	Authors have distinguished between primary (anger, fear,happiness, and sadness) and secondary emotions.	Surprise emotion is also required for teaching stratergy
Plutchik et al. [12]	8 Basic emotion	Authors created a wheel of emotions which consisted of 8 basic emotions arranged as four pairs of opposites (joysadness, trust-distrust, fear-anger, surpriseanticipation), and 8 advanced emotions each composed of 2 basic ones.	Few emotions are seldom used
Kort B et al.	6*6 Emotion Access	Authors have suggested 6x6 possible emotion axes (anxiety-confidence, ennui-fascination, frustration-neuphoria, dispirited-enthusiasm, terror-excitement, humiliated-proud) that may arise in the course of learning raging from negative (rank -1.0) to positive (rank +1.0) valence.	Too many intermediate emotions are not necessary for Teaching Stratergy
Juslin et al. [13]	Tone or Intensity of Voice	Authors have shown how different emotions can change the tone, articulation and intensity of the voice. mood and accuracy of the device to extract.	All emotions are not clearly identified
Fiedakis et al. [12]	10 Basic emotion Model	Authors have made an attempt to classify the aforementioned fundamental models and theories of basic emotion, which resulted in ten basic emotions: anger, happiness, fear, sadness, surprise, disgust and love, anticipation, joy and trust.	Too many intermediate emotions are not necessary for Elearing model

tracks the emotional state. These emotional states are categorized into four categories i.e., Self-confidence, puzzle or bewilderment, depression or despair and surprise. Accordingly teaching strategies are recommended based on the emotional states. The major disadvantages of the above methodology is the facial expression recognition which is not accurate in the Euclidian distance methods.

Zhu Aiqin et al. [5] proposed a similar facial expression recognition system which uses an algorithm based on rough set theory and the classification is done based on the template based classifications. Other classification techniques are also used like rule based, ANN based, HMM based, Bayesian and SVM based classifiers. Then the change of emotion state is tracked and teaching strategies are changed accordingly. The major limitation is that the classification tree fails to recognise all possible required emotions.

Jacob Whitehill et al. [1] developed an automatic recognition of student engagement from facial expression using binary classification and machine learning techniques. S L Happy et al. [6] proposed an automated facial landmark detection technique which is similar to the state-of-art-landmark detection methods.

The other eLearning systems based on affective computing were also based on similar existing methodologies [7] [8] [10] [11] [12] [13]. Automatic recognition system works fine with the feature extraction but it fails to recognize multiple faces in a single frame.

This motivates us to propose a multi user single frame based face detection system for eLearning environment and there by improving the processing speed by using both GPU and CPU.

III. PROPOSED METHODOLOGY

The proposed methodology consists of two modules. The Affective computing module and the traditional eLearning module. The block diagram of the proposed eLearning system is as shown in the Figure 1. The affective computing module consists of two phases: Training phase and execution phase. Traditional eLearning system consists of recommendation phase.

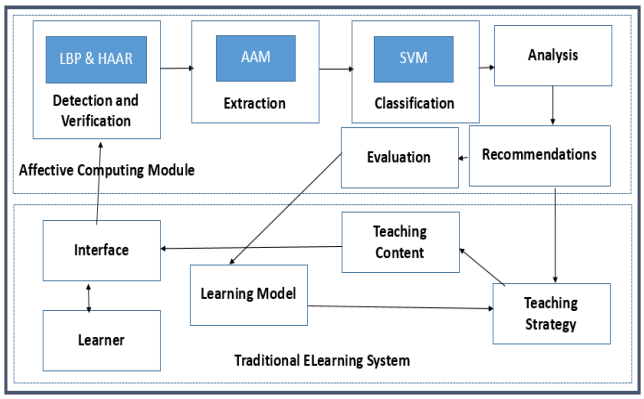

Fig. 1. Proposed Elearning System

A. Training Phase

Training phase consists of six major steps. The block diagram of the training phase is shown in Figure 2.

Preprocessing: The given images are converted to gray scale image, the resolution is diminished to 1024*768. For

some images cropping of the facial images is also done to increase the accuracy of feature extraction.

Face Detection using CPU: Haar feature based classifier is used to detect the frontal face in the data set. The detectors parameters are varied such that the dimension of the object to detected (face) will have minimal effect on the detection accuracy.

Feature Extraction Using AAM [2]: The landmark points or the feature points are identified using a AAM (Active Appearance Model) which gather 100 such points of which the some key distances and measures such as the distance between the eye lids, mouths, width of mouth etc. are extracted.

SVM Training: The SVM is trained using a combination of Radial Basis Function Kernel and polynomial kernel with the input as the feature points. Each set of feature points is mapped to a particular emotional expression out of 7 major emotions. The trained SVM data is stored in a serialized format.

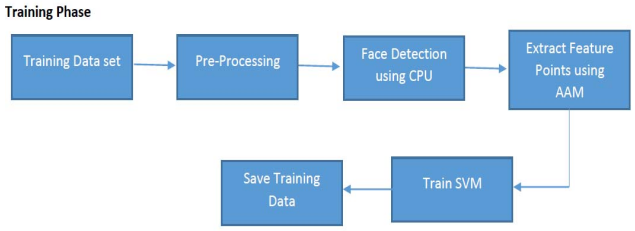

Fig. 2. Block Diagram of Training Phase in Emotional Recognition

B. *Execution Phase:*

Face Detection using GPU: A GPU based object detector based on Local Binary Patterns is used in this step. Though the accuracy of this detection is not impressive but much faster than other detectors. It can be used to localize the search or minimize the computational load of detecting the object in the upcoming phase.

Face Verification using CPU: Similar to the detection stage in training phase, this module uses a CPU based accurate detector based on Haar Cascades, but it is slow. Hence the regions so detected as face by the previous stage is applied to execution stage where it cross-validates using a more accurate detector. Since the area to be scanned is smaller when compared to the original image, hence this verification phase executes faster.

Emotion Classification: After facial feature point detection by AAM, the desired measures are applied to the SVM classifier which is initialized with already trained data. This stage classifies the detected face image in to one of the seven emotions. The accuracy of classification can be improved by increasing the size of training data set. The block diagram of the execution phase is shown in Figure3.

C. *Recommendation Phase*

Based on the emotional behavior of the users, for each frame the effective emotion of the group of people within the frame is calculated by assigning individual values to each emotions and averaging them out. A graph is plotted using these values and area of the graph is calculated. Certain threshold is fixed such that if the area of the graph is not above this predetermined threshold, the teaching is ineffective. Then depending on the value obtained as the area, recommendations can be given to the next phase of teaching strategies and are modified accordingly.

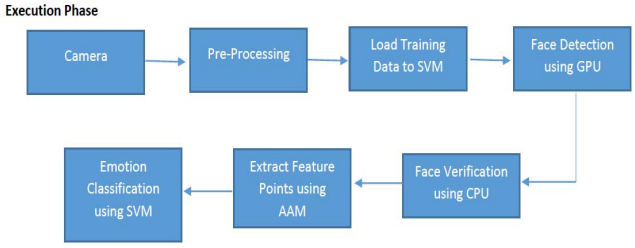

Fig. 3. Block Diagram of Execution Phase in Emotional Recognition

IV. EXPERIMENTAL RESULTS AND ANALYSIS

Experimental Setup

A. *System Configuration:*

- Processor: Intel Core I5
- RAM: 4GB
- OS: Windows 7
- GPU: Nvidia GeForce GT 830 M

B. *Datasets Used:*

- *Training Phase Dataset:* Yale Face Database (YFD) [16] has been considered as a training dataset. Before using YFD as a training dataset, some pre-processing steps like cropping and thresholding are performed so as to increase the accuracy of feature extraction. But thresholding does not significantly increase the accuracy.
- *Execution Phase Dataset:* Once the training is done in Yale Face Database, the proposed method is tested for the multiple face detection using Face Detection Dataset and Benchmark (FDDB) [17] and Labelled Faces in the Wild (LFW) [18] face databases. The FDDB database is a dataset which consists of multiple faces in a single frame. Although the training is done using Yale Face Database, which trains for single face, those features are used to detect multiple face emotions present in the FDDB and LFW face databases.

The FDDB dataset is designed for studying the problem of unconstrained face detection. This dataset is taken from faces of wild dataset. The faces of wild dataset also known as labelled faces in the wild, consists of over 12000 images of faces which are collected from the web. These images are categorised into four different set of images namely :- i) original/normal images. ii) Aligned images, where the images are aligned in a proper manner such that the face detection can be done more accurately. iii) Funnelled images, which are aligned images but each part or pose are collected in different

angles in different images. iv) Deep funnelled images, are the aligned images which are developed using machine learning techniques where each feature can be tuned to represent the images which are at different resolutions. Results obtained are as shown in the Table II.

TABLE II. ACCURACY ANALYSIS FOR THE FDDB /LFW DATASET

Total Images	12620
Frontal Faces Detected	11232
Partial Detection	232
False Positive	0

Since GPU is tailor made for image and video processing and image processing [17], GPU can significantly improve the speed of processing. The improvement in frame rate achieved with and without GPU is shown in the Table III.

TABLE III. PERFORMANCE OF PROPOSED SYSTEM

FPS with both CPU and GPU	17
FPS with only CPU	9
Speed Up Factor	≈ 2

*FPS – Frames Per Second

Once the face detection is done for multiple faces using both CPU and GPU, the corresponding emotion is detected. We are detecting the emotion not only for single face but also for multiple faces in a single frame. Thus the average emotional value per frame for a sample video is shown in the Figure 4.

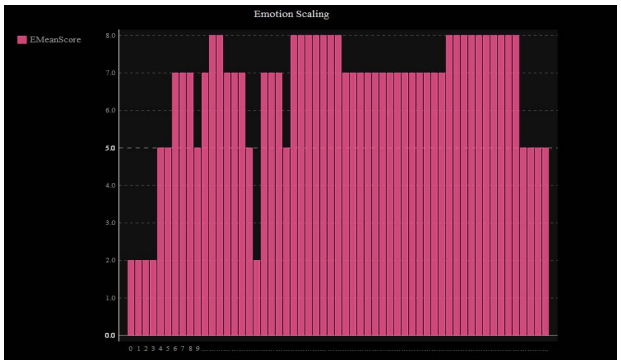

Fig. 4. Plot of Average Emotional Value per Frame

The x-axis represents the times/frames and y-axis denotes average emotional value (EMean Score). This value obtained from scaling of emotion is utilized to suggest changes in teaching strategy. E.g. if the scale value is 9 then it represents the surprise. If the scale value is 1 then it represents sleepy, in turn it suggests to modify the teaching strategy.

CONCLUSION AND FUTURE WORK

The proposed method detects facial emotion recognition for multiple faces in a single frame. GPU along with the CPU helps in speeding up this process. Hence this makes eLearning a better educational system by recommending an improvised teaching strategy for a group of eLearning students.

A lot of improvements can be done in near future w.r.t. proposed eLearning system for improving the performance in the teaching strategy which is not focused much in the present methodology. The present work will be adapted in traditional classroom systems and thereby considering suitable teaching strategy for improving the efficiency of teaching learning process. Further improvements will be done in face detection using background subtraction and individual statistics based on emotions for face recognition.

REFERENCES

[1] Whitehill, Jacob, et al. "The Faces of Engagement: Automatic Recognition of Student Engagementfrom Facial Expressions." Affective Computing, IEEE Transactions on 5.1 (2014): 86-98.

[2] Scheenstra, Alize, Arnout Ruifrok, and Remco C. Veltkamp. "A survey of 3D face recognition methods." *Audio-and Video-Based Biometric* Person Authentication. Springer Berlin Heidelberg, 2005.

[3] Daradoumis, Thanasis, Marta Arguedas, and Fatos Xhafa. "Current Trends in Emotional e-Learning: New Perspectives for Enhancing Emotional Intelligence." *2013 7th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS)*. IEEE, 2013.

[4] Jingjing Chen; Qi Luo "Research on e-learning system model based on affective computing"Computer-Aided Industrial Design and Conceptual Design, 2006. CAIDCD '06. 7th International Conference

[5] Aiqin, Zhu, and Qi Luo. "Study on E-learning System Model based on Affective Computing." Information and Automation, 2006. ICIA 2006. International Conference on. IEEE, 2006.

[6] SL Happy et al. "Automatic facial expression recognition using features of silent facial patches" Affective Computing, IEEE Transactions on vol.6 No.1 (2015).

[7] Jingjing Chen; Qi Luo "Research on e-learning system model based on affective computing"Computer-Aided Industrial Design and Conceptual Design, 2006. CAIDCD '06. 7th International Conference

[8] Duo, Sun, and Lu Xue Song. "Research on E-learning system based on affective computing." Information Management and Engineering (ICIME), 2010 The 2nd IEEE International Conference on. IEEE, 2010.

[9] Calvo, Rafael, Giuseppe Riva, and Christine Lisetti. "Affect and Wellbeing: Introduction to Special Section." Affective Computing, IEEE Transactions on 5.3 (2014): 215-216.

[10] Li, Lan, Li Cheng, and Kun-xi Qian. "An e-learning system model based on affective computing." Cyberworlds, 2008 International Conference on. IEEE, 2008.

[11] Saleh, Ahmed Farouk Mohamed, and Al Farouk Ahmed Farouk. "Students with Disabilities' Attitudes towards E-Learning Courses in Developing Countries." e-Learning" Best Practices in Management, Design and Development of e-Courses: Standards of Excellence and Creativity", 2013 Fourth International Conference on. IEEE, 2013.

[12] Feidakis, Michalis, et al. "Measuring the Impact of Emotion Awareness on e-learning Situations." Complex, Intelligent, and Software Intensive Systems (CISIS), 2013 Seventh International Conference on. IEEE, 2013.

[13] Vermun, Kartik, et al. "Gesture-Based Affective and Cognitive States Recognition Using Kinect for Effective Feedback during e-Learning." Technology for Education (T4E), 2013 IEEE Fifth International Conference on. IEEE, 2013.

[14] http://vision.ucsd.edu/

[15] http://vis-www.cs.umass.edu/fddb/

[16] http://tamaraberg.com/faceDataset/

[17] http://docs.opencv.org