International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal

# REAL TIME-EMPLOYEE EMOTION DETECTION SYSTEM (RTEED) USING MACHINE LEARNING

[1]KOPPULA RAM KISHORE,[2]S.K.ALISHA

[1]MCA Student,B V Raju College, Bhimavaram,Andhra Pradesh,India

[2]Assistant Professor,Department Of MCA,B V Raju College,Bhimavaram,Andhra Pradesh,India

**ABSTRACT:**

The paper introduced the present status of speech emotion recognition. In order to improve the single-mode emotion recognition rate, the bimodal fusion method based on speech and facial expression was proposed. The emotional databases of Chinese speech and facial expressions were established with the noise stimulus and movies evoking subjects' emtion. On the foundation, we analyzed the acoustic features of Chinese speech signals under different emotional states, and obtained the general laws of prosodic feature parameters. We discussed the single-mode speech emotion recognitions based on the prosodic features and the geometric features of facial expression. Then, the bimodal emotion recognition was obtained by the use of Gaussian Mixture Model. The experimental results showed that, the bimodal emotion recognition rate combined with facial expression was about 6% higher than the single-model recognition rate merely using prosodic features.

**Keywords: CNN, Emtion, Gaussian Mixture model, trained image.**

## 1. INTRODUCTION:

With the wide application of computers in various fields, speech recognition as the key technology of human-computer interaction has attracted more and more attention. However, current research on emotion recognition does not go far enough. The studies on many aspects have not led to a systematic theory such as the establishment of the emotional speech databases, the selection and parameter extraction of emotional features, emotion recognition methods [1]. The related studies on English, Japanese, and etc are comparatively more than those on Chinese feature selection is mainly on prosodic parameters. Emotional analysis methods mainly include Principal Component Analysis, Gaussian Mixture Models, Hidden Markov Models, Support Vector Machines, and etc [2]. There have got been certain results in those areas. But there has been less research on Multi-modal speech emotion recognition which integrates facial expression and human physiological signals.Because of the inherent defects of voice in the emotion detection, using voice signals to identify the emotional state, the recognition rate can only reach about 80%, and the robustness of the recognition results can not be guaranteed. Emotion detection from a single channel has become increasingly unable to meet the actual needs of the project [3-5]. Therefore, complementary features extracting from the dual-mode has become new ways to improve speech emotion recognition rates.

This paper presents a dual-mode recognition method based on prosodic features

and facial expression to increase the rate of speech emotion recognition and robustness.

## 2. LITERATURE SURVEY

Facial emotions are important aspects in human communication that help us to understand the intentions of others. Facial expressions convey Non-verbal Cues which play an important role to maintain interpersonal relations. According to different surveys verbal component (speech) convey one-third of human and Non-Verbal components (Facial emotions, Gestures) convey two-third of human communication. Facial emotion detection became a well attempted research topic now days due to its prospective accomplishments in many domains such as Medical engineering, Vehicles, Robotics and Forensic applications etc. Emotion Recognition will help to understand the inner feelings for people by using their facial expression.

[1] 2017 IEEE 4th InternationalConference on KnowledgeBased Engineering and Innovation (KBEI) https:/ /ieee xplore. ieee.org/document/8324974

[2] Communication 2019 1st on Innovations in Informationand Communication Technology (ICIICT). https://ieeexplore. ieee. org/document/8741491

[3] International Journal of MachineLearning and Computing, Vol.9, No. 1,February2019http://www.ijmlc. org/ vol9/759-L0 179.pdf

[4] D Y Liliana, Published under licence by IOP Publishing Ltd Journal ofPhysics: Conference Series, Volume1193, conference https://iopscience.iop.org/article/10.1088/1742 -6596/1193/ 1/ 0 1 2004.

**Existing System**

In the existing system affective computing is the “computing that relates to, arises from, or influence emotions”, or in the other words, any form of computing that has something to do with emotions. The creation of automatic classifier involves collecting information, extracting the features which are important and finally training the data, so it classify and recognize some patterns. To build a model have to extract emotion of happiness and sadness from facial expression and have to feed the model with pictures of people smiling, tagged with ”happiness”, and with pictures of people frowning, tagged with “sadness”. After that, when it receives a picture of a person smiling or frowning, it identifies the shown emotion as “happiness” or “sadness”. Emotion detection using speech, gathering emotional information from the user of a system is their voice. Any emotion from the speaker’s speech is represented by the large number of parameters which is contained in the speech and changes in these parameters will result in corresponding changes in emotions which is quite difficult.

**Disadvantages of existing system**

• Creation of model in real life is difficult.

• Voice recognition software won’t always put your words on the screen completely accurately.

• Programs cannot understand the context of language the way that humans can, leading to errors that are often due to misinterpretation.

## 3. METHODOLOGY

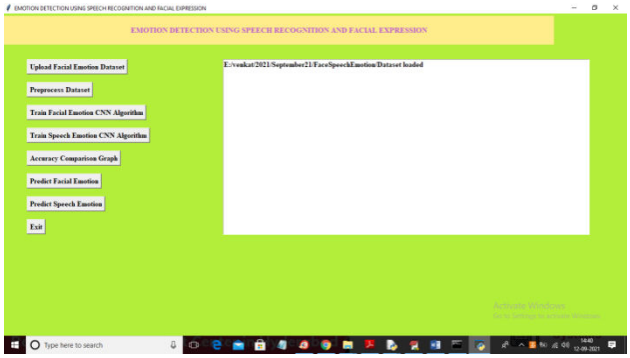
To overcome the existing drawbacks, comparing the traditional machine learning approaches, deep learning based methods have

shown better performance in terms of accuracy and speed of processing in image recognition. We have used a modified Convolutional Neural Network (CNN). CNN is mostly used in image and face recognition. CNN is a kind of artificial neural networks that employs convolutional methodology to extract features from the input data to increase the number of features from live video streaming. That captures each frame and test them and is trained by CNN model and later classified into different emotions. With computational power of Graphical Processing Units (GPU's), CNN has achieved remarkable cutting edge results in image recognition.

**Emotion and Features:**

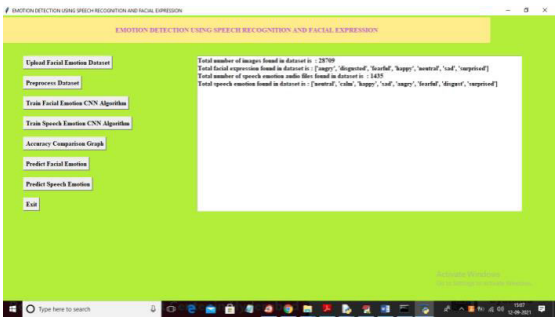
Psychological studies have shown that changes in human emotions reflect through prosodic parameters of speech. Generally, acoustic features associated with the emotions including pitch, duration time, energy, formant, and average, maximum, minimum, intermediate values, ranges, the first derivative, the second derivative and change rates derived from them [7]. After repeated experiments, this paper eventually selected the following prosodic features: phonation time, speech rates, basic frequency averages, basic frequency ranges, basic frequency change rates , Amplitude averages, Amplitude change ranges, formant change averages, formant change ranges, and formant change rates.

Face features generally include three kinds of: Geometric features, physical features, mixing features. The physical features refer to the features using the whole face image pixels, reflect the underlying information of face images, and focus on extracting the subtle changes of local features [8]. However, the number of feature point? extracted is too many that resuling to the higher dimension and the complex calculations. Mixing features combine the geometric features with physical features. The calculation of it is also complex, and the initial point is difficult to obtain [9]. The recognition effect of the geometric features requires a higher accuracy of the Datum point extracted. The recognition effect of requiring a higher accuracy of the Datum point extracted. Meanwhile extracting the geometric features ignores the other information of faces (such as skin texture changes etc.) But it can describe the macro structural changes of the face, and the easy way to extract and the lower dimension making it quite comply with the requirements of our emotional system.

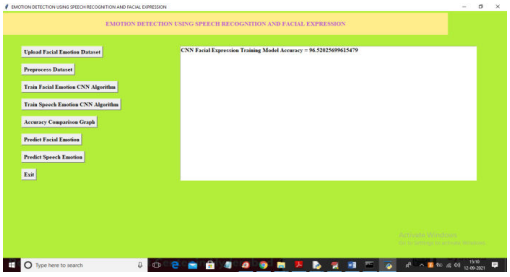



In above screen dataset loaded and now click on 'Preprocess Dataset' button to read all images and then resize them to equal size and then extract MFCC features from dataset and then build trained model.

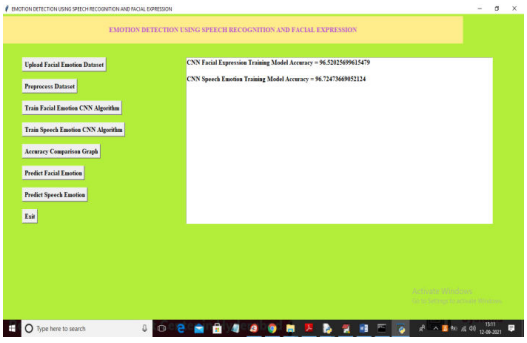
International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal



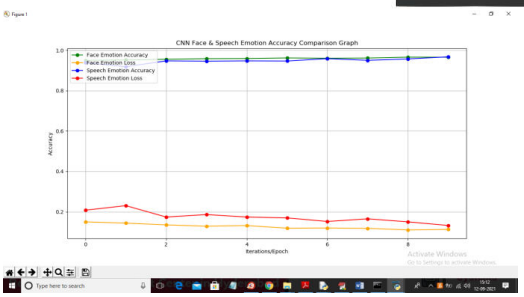

In above screen both datasets are processed and we can see total number of images and audio files available in both datasets and now dataset is ready and now click on 'Train Facial Emotion CNN Algorithm' button to train Facial dataset with CNN and to get below screen



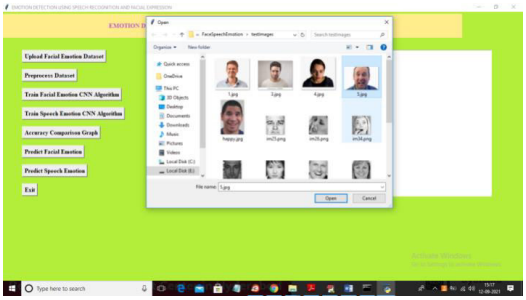

In above screen training CNN with Facial images got 96.52% accuracy and now click on 'Train Speech Emotion CNN Algorithm' button to train CNN with audio features and to get below output



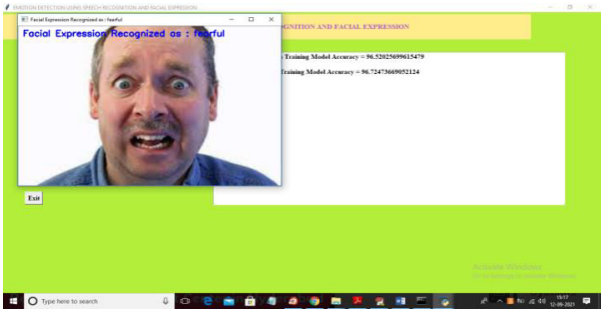

In above screen with CNN speech Emotion we got 96.72% accuracy. Now click on 'Accuracy Comparison Graph' button to get below graph




In above graph x-axis represents EPOCH and y-axis represents accuracy and loss values and we can see both algorithms accuracy reached to 1 and both algorithms loss values reached to 0. In above graph green line represents face emotion accuracy and blue line represents speech accuracy. Now click on "Predict Facial Emotion" button to upload face image and will get below result

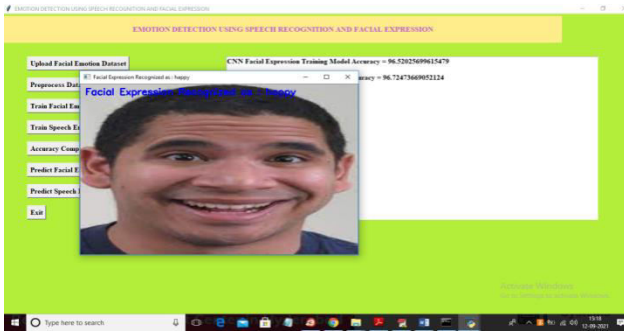



In above screen selecting and uploading '5.jpg' image and then click on 'Open' button to get below result

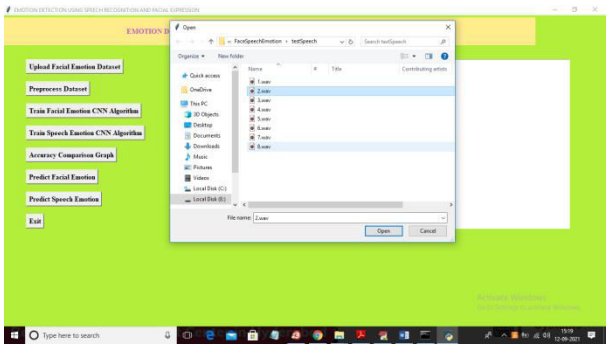



In above screen facial emotion or expression predicted as 'Fearful' and now test other image

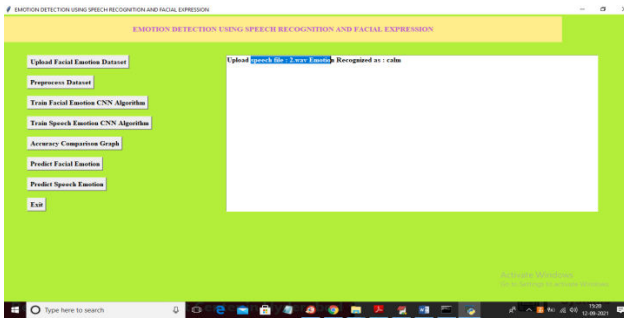
International Journal For Recent Developments in Science & Technology
2581-4575
Crossref
A Peer Reviewed Research Journal



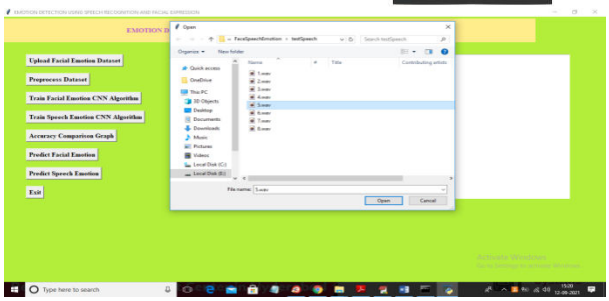

In above screen facial expression predicted as 'happy' and similarly you can upload other images and test. Now click on 'Predict Speech Emotion' button to upload audio file and get below result



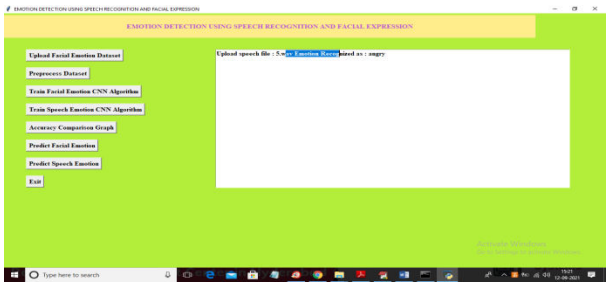

In above screen selecting and uploading '2.wav' file and below is the result




In above screen uploaded audio file emotion predicted as 'calm' and now test other file




In above screen uploading '5.wav' file and below is the prediction result




In above screen uploaded file emotion predicted as 'angry' and similarly you can upload other files and test.

## 4.CONCLUSION

In the experiment, we analyze and compare time, amplitude energy, basic frequency and formant feature parameters under different emotional states, and find out the distribution laws of different emotional signal features. On this basis, we classify five emotional states of calm, sadness, happiness, surprise, and anger. The recognition results show that on these basic Prosodic information we can initially recognize basic emotional catagories, and apply it into the emotion recognition system, which limits the amount of storage and computation and doesn't have strict recognition accuracy. Meanwhile, the prosodic features, integrating with facial expression information, recognizes emotional

categories with Bimodal, reaching a higher recognition rate.

Although the emotional recognition performance combinating with facial expressions has improved, The recognition rate doesn't improve significantly. This is mainly because in the terms of obtaining the emotional information, there is a similar correlation between the adjacent video frames, due to the continuity of the facial expression changes. But we didn't take this correlation into consideration when catching the instant face image to analyze separately. On the other hand, when the facial expression changes, the shape and the location of the organs on the face, will change accordingly. In this paper,although the image analysis method based on Gaussian mixture algorithm has a higher recognition rate for the face contour, it lacks of detailed characterization of changes in the eyes, nose, mouth and other facial organs. Based on the above two reasons, in order to truly improve the system performance, we need to build a correction model associated with the expressions containing a variety of rules, and modified the image recognition results using the model. In addition, in the term of real-time applications, besides enhancing the robustness of the system and improving the accuracy, the efficiency of the recognition algorithm is also a key factor. The strategies such as codebook pruning, data compression can also improve the recognition rate effectively

Multi-modal recognition systems intergrating with images, voice and other emotional information is the inevitable trend of future human-computer interaction development. Although there are still many insurmountable technical problems, with the continuous progress of science and unremitting efforts of the researchers, the real-time systems of multi-modal speech recognition will have more potential development.

## 5.REFERANCES

[1] LS Zhao, Q Zhang, XP Wei. A Research Progress in Speech Emotion Recognition. Computer Application and Research. 2009; 26(2): 428-432.

[2] CW Huang, Y Zhao, Y Jin. A Study of Practical Speech Emotion Features Analysis and Recognition. Electronics and Information Journal. 2011; 33(1): 112-116.

[3] LL Xu, ZX Cai, MY Chen. A Study Review of Emotion Feature Analysis and Recognition of Speech Signals. Circuits and Systems Journal. 2007; 12(4): 77-84.

[4] YM Huang, GB Zhang, HB Liu. Emotion Detection Based on New Bimodal Fusion Algorithm. Journal of Tianjin University. 2010; 43(12):1067-1072.

[5] CW Huang, Y Jin, QY Huang. Mutil-Modal Emotion Recognition Based on Speech Signals and ECG. 2009; 40(5): 895-900.

[6] B Xie. A Study of Key Technologies on Mandarin Speech Emotion Recognition. Zhejiang: Zhejiang University Computer Science and Technology Major. 2006.

[7] YM Huang, GB Zhang, X Li. Speech Emotion Recognition Base on Small Samples of Global Features and Weak-Scale Integration Strategy. Acoustics Journal. 2012; 37(3): 330-338.

[8] LQ Fu, YB Wang, CJ Wang. Speech Emotion Recognition Based on Mutil-Feature

Vectors. Computer Science. 2009; 36(6): 231-234.

[9] XQ Jiang, SY Cui, YH Yin. Speech Emotion Processing in the Man-Machine Speech Interaction. Jounal University of Jinan : Version of nature science. 2008; 22(4): 354-357.

[10] YL Xue, X Mao, Y Guo. A Research Pogress in Facial Expression Recognition. Chinese Image and Graphic Journal. 2009; 14(5): 764-772.

[11] M Fan. Emotional Speech Recognition Based on Facial Expression Analysis. Shandong: Shandong University Circuits and Systems Major. 2009.

[12] TF Zhang, R Min, BY Wang. Facial Expression Recognition Based on Automatic Segmentation of the Characteristic Regions. Computer Engineering. 2011; 37(10): 146-151.