Real-Time Emotion Classification using Facial Expression Recognition: A Survey

Abhishek Sanjay Gurav[1], Prof. Pramila M. Chawan[2]

[1]M.Tech Student, Dept of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India
[2]Associate Professor, Dept of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India

***

Abstract - *For a long time, expressions have been something that human beings are proud of. That is an essential difference between us and machines. Facial Expressions can be classified using various technical factors and we've studied various such factors in our surveys as well. In this paper we have surveyed various machine learning techniques for facial expression recognition and have proposed a method to tackle both of the mentioned issues. This model can be used for prediction of expressions of both still images and real time video.*

Key Words: **Machine Learning, Emotion Classification, Facial Expression Recognition, Facial Action Units, Real-time.**

1. INTRODUCTION

Human language is divided into natural language and body language. Facial expression is part of body language. Therefore, accurately identifying the expression of others is critical to successful human–computer interaction. When people see different people's faces, they can easily recognize the same expression, which is called facial expression recognition. The main application fields of facial expression recognition technology include human–computer interaction, intelligent control, security, medical, communication, education, fatigue detection, political election and other fields. In these fields recognizing facial expressions is a task so crucial that it needs to be instantaneous and accurate simultaneously. In this paper we are going to propose a solution to issues like latency and accuracy based on our surveys conducted on the existing methods.

1.1 Supervised Machine Learning

Machine learning algorithms especially are of two types supervised and unsupervised machine learning algorithms. A supervised machine learning model gets trained on the labelled dataset. The dataset in our proposed methodology will be a labelled dataset as well.

Images of every particular expression will be placed in a separated folder labelled with the name of that expression.

1.2 Performance Metric

This is a multi-class classification problem with 7 different classes, so we have considered three performance metrics:

a. Multi-Class Log-loss: We have used deep learning model with cross-entropy layer in the end with seven softmax units, so therefore our goal is to reduce the multi-class log loss/cross-entropy loss.

b. Accuracy: Measure of how accurate our model will perform expression prediction.

c. Confusion Metric: Since the problem is multi-class classification, so confusion metric will helps us to know which classes are more dominant over others or towards which class the model is more biased. This gave us the clear picture of the prediction result of the model.

2. LITERATURE REVIEW

In [1], an FER system based on machine learning is proposed. ROI rearrangement process minimized the environmental change factor and the hierarchical structure of the person, and facial expression classification improved the classification rate. Based on this study, it could be applied to the personalized vehicle interfaces to prevent traffic accidents by combining this proposed system with vehicle.

In [2], 1:1 combination is observed for the classification i.e., one test image is tested with the one trained image for each expression, this demerits to obtain the best match for the expressions. Proposed work executed on 50 training images and tested with 42 query image. 1:N comparison of the test and query image will be performed. Hence accurate result can be obtained and according to the results they took 11 facial expression images of happy, 07 of Angry, 04 of Sad, 10 0f Disgust, 02 of Fear and surprise, and 06 of Neutral. For every query image they almost got perfect match with existing training images and hence by using PCA methodology which was found to be the best method to work on Eigen faces and would give more accuracy in classification.

In [3] they have presented a pilot study for real-time testing of conventional and deep learning approaches in facial emotion recognition. Preliminary results show better generalization power and better performance in real-time application of fine-tuned AlexNet CNN and Affdex CNN than

SVM and MLP approaches. Commercial Affdex CNN has overall superior accuracy, but AlexNet and SVM had better “anger” recognition (96.77% vs. 70.97%). FER-CNN had the lowest overall accuracy but high accuracy for “sadness”, comparable with Affdex CNN result (81.82% vs. 84.85%). In further research we will test this fact in a larger group of volunteers and for more than four emotions.

In [4], an average weighting method is proposed to avoid potential errors in real-time facial expression recognition based on the traditional convolutional neural network. Using a camera with high frame rates, the influence of noise from the environments can be reduced. Moreover, because of the average weighting method, facial expression recognition results become more robust from frame to frame. As a result, the accuracy of facial expression recognition is improved. Experimental results have shown that the proposed facial expression recognition system is more reliable than the traditional CNN approach.

In [5], they proposed and implemented a facial expression recognition framework which could be able to improve the accuracy of recognition of *anger and sadness* expressions. The experiments showed that our proposed system outperforms other state-of-the-art methods on the CK+ database. The main idea behind this successfulness is to select salient facial regions relative to each particular expression of six basic emotions.

Table -1: Salient Regions of various expressions

Expression	Saliant Facial Regions
Anger	Mouth, eye and nose regions
Fear	Mouth and eye regions.
Disgust	Nose, mouth and eye regions. Wrinkles on nose region gets little more attention than the other two regions.
Sadness	Mouth and eye region. Biased towards mouth region
Surprise	Mouth region
Happiness	Mouth region

In [6], a potential ability for recognizing facial expression based on landmarks was explored, which may could prove that the human brain can recognize facial expressions by using only 68 points instead of all pixels of a face image. We use the landmarks provided by CK+ dataset for experiment. Input vectors were normalized by multiple origins or single origin, then trained into MLP for training to classifying expressions. The result in test samples shows that the method based on landmarks also have comparable performance with CNN based methods. But, the performance of the proposed method is also related to the precision of landmark extracting algorithm.

In[7], instead of the texture information of the whole face, the proposed method just considers the geometric information in 2D space structured by the feature points. Thus, there is a chance of improvement in interception speed. Extensive experiments have been carried out on BHU facial expression database and MMI facial expression database. The results are mainly analysed by interception accuracy and speed, which validate that the proposed FESI is improved over existing methods.

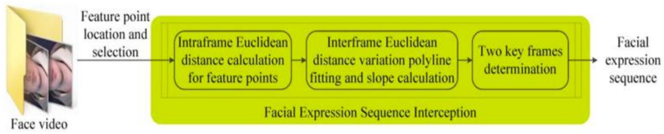

Fig -1: FER using Feature point movement

Although the proposed method performs well on FESI, it fails to handle the non-frontal face images, which is the most challenging task for all of the researches. Besides, the feature point location must be exact in this method, and this problem is solved by the manually correcting feature point coordinates.

In [8], a hierarchical framework based on Dynamic Bayesian Network for simultaneous facial feature tracking and facial expression recognition was proposed. By systematically representing and modelling inter relationships among different levels of facial activities, as well as the temporal evolution information, the proposed model achieved significant improvement for both facial feature tracking and AU recognition, compared to state of the art methods. For six basic expressions recognition, the result wasn’t as good as that of state of the art methods, since they did not use any measurement specifically for expression, and the global expression is directly inferred from AU and facial feature point measurements and from their relationships.

In [9], Expression recognition has a good performance, but there is still a gap compared to the actual application. experimental results showed that the correct rate of the neural network after 150 training sets is 85% after 1000 trainings, after 2000 training the correct rate is 99%, the accuracy of the test set after 3000 iterations of training had again dropped to 89%.

In [10], the generalization performance to new subjects for recognition of full facial expressions of emotion in a 7-way forced choice was 93.3%, which is the best performance reported so far on this publicly available dataset. The ML-based system presented here can be applied to recognition of any facial expression dimension given a training dataset. They also applied the system to fully automated facial action coding, and obtained a mean agreement rate of 94.5% for 18

AU's from the Facial Action Coding System. This is the first system that we know of for fully automated FACS coding of images without an infrared eye position signal. The outputs of the expression classifiers change smoothly as a function of time, providing information about expression dynamics that was previously intractable by hand coding.

In [11], the proposed system employs both detection and tracking algorithms together to make the system fast enough to perform accurately in real time. The framework uses landmarks to localize a few facial patches from which the discriminating feature vectors are computed. Based on the extracted features, the SVM classifier classifies the input face image into either a neutral face or the expression label. The performance of the proposed framework gives promising results in both CK+ and RafD databases.

In [12], based on facial landmarks detected by IntraFace algorithm, seven regions of interest (ROI), corresponding to the main components of face, are first extracted to represent face image. A pre-processing stage is then applied on these ROIs for resizing and partitioning them into blocks, before performing feature extraction to build face feature descriptor. Finally, a multiclass SVM classifier is utilized to infer emotion state.

3. PROPOSED SYSTEM

3.1 Problem Statement

To improve facial expression detection accuracy without compromising on latency and accuracy of the results.

3.2 Problem Elaboration

Computer animated agents and robots bring new dimension in human computer interaction which makes it vital as how computers can affect our social life in day-to-day activities. Face to face communication is a real-time process operating at a time scale in the order of milliseconds. The level of uncertainty at this time scale is considerable, making it necessary for humans and machines to rely on sensory rich perceptual primitives rather than slow symbolic inference processes.

In this project we are presenting the real time facial expression recognition of seven most basic human expressions: ANGER, FEAR, HAPPY, NEUTRAL, SAD, DISGUST, SURPRISE.

3.3 Proposed Methodology

We plan to use Facial Action Units for expression recognition, steps:

1. Acquire frame sequence
2. Preprocessing and Segmentation
3. Face points detection
4. AU defining
5. Dominant emotion Classification
6. Expression Labelling

Action Units are highlighted using red dots in the example below:

Fig -2: Facial Action Units

Preprocessing of images is important to efficiently train the model. We will follow 5 simple steps for processing and changing our images such that they become suitable for feeding to the model for training.

Step-1: Converting images to grayscale.

Step-2: Detect face in the image using OpenCV HAAR Cascade.

Step-3: Crop the image to the face.

Step-4: Resize the image to 350*350.

Step-5: Finally save the image.

4. CONCLUSIONS

In this paper, we have performed a survey on various methodologies in FER including CNN, HOG features, AUs, Gabor features, salient regions and also feature-point movement proposed a supervised machine learning based approach which will use facial action units to improves latency and accuracy issues in emotion recognition using facial expression recognition.

REFERENCES

[1] S. Kim, G. H. An and S. Kang, "Facial expression recognition system using machine learning," 2017 International SoC Design Conference (ISOCC), Seoul, 2017, pp. 266-267, doi: 10.1109/ISOCC.2017.8368887.

[2] M. B. S. Divya and N. B. Prajwala, "Facial Expression Recognition by Calculating Euclidian Distance for Eigen Faces Using PCA," 2018 International Conference on

Communication and Signal Processing (ICCSP), Chennai, 2018,pp.0244-0248, doi:10.1109/ICCSP.2018.8524332.

[3] A. Kartali, M. Roglić, M. Barjaktarović, M. Đurić-Jovičić and M. M. Janković, "Real-time Algorithms for Facial Emotion Recognition: A Comparison of Different Approaches," 2018 14th Symposium on Neural Networks and Applications (NEUREL), Belgrade, 2018, pp. 1-4, doi: 10.1109/NEUREL.2018.8587011.

[4] K. Liu, C. Hsu, W. Wang and H. Chiang, "Real-Time Facial Expression Recognition Based on CNN," 2019 International Conference on System Science and Engineering (ICSSE), Dong Hoi, Vietnam, 2019, pp. 120-123, doi: 10.1109/ICSSE.2019.8823409.

[5] A. Vo and B. T. Nguyen, "Facial Expression Recognition Based on Salient Regions," 2018 4th International Conference on Green Technology and Sustainable Development (GTSD), Ho Chi Minh City, 2018, pp. 739-743, doi: 10.1109/GTSD.2018.8595551.

[6] Y. Qiu and Y. Wan, "Facial Expression Recognition based on Landmarks," 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chengdu, China, 2019, pp. 1356-1360, doi: 10.1109/IAEAC47372.2019.8997580.

[7] J. Yi, Y. Sima, M. Zhou and J. Yang, "Facial Expression Sequence Interception Based on Feature Point Movement," 2019 IEEE 11th International Conference on Advanced Infocomm Technology (ICAIT), Jinan, China, 2019, pp. 58-62, doi: 10.1109/ICAIT.2019.8935902.

[8] Y. Li, S. Wang, Y. Zhao and Q. Ji, "Simultaneous Facial Feature Tracking and Facial Expression Recognition," in IEEE Transactions on Image Processing, vol. 22, no. 7, pp. 2559-2573, July 2013, doi: 10.1109/TIP.2013.2253477.

[9] L. Pang, N. Li, L. Zhao, W. Shi and Y. Du, "Facial expression recognition based on Gabor feature and neural network," 2018 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), Jinan, China, 2018, pp. 489-493, doi: 10.1109/SPAC46244.2018.8965443.

[10] M. S. Bartlett, G. Littlewort, C. Lainscsek, I. Fasel and J. Movellan, "Machine learning methods for fully automatic recognition of facial expressions and facial actions," in IEEE International Conference on Systems, Man and Cybernetics, 2004, pp. 592-597.

[11] P. Kumar, S. L. Happy and A. Routray, "A real-time robust facial expression recognition system using HOG features," 2016 International Conference on Computing, Analytics and Security Trends (CAST), Pune, 2016, pp. 289-293, doi: 10.1109/CAST.2016.7914982.

[12] K. Lekdioui, Y. Ruichek, R. Messoussi, Y. Chaabi and R. Touahni, "Facial expression recognition using face-regions," 2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Fez, 2017, pp. 1-6, doi: 10.1109/ATSIP.2017.8075517.

BIOGRAPHIES

Abhishek Gurav,
MTech Computer Engineering VJTI, Mumbai.

Prof. Pramila M. Chawan,
Associate Professor, Department of CE and IT, VJTI, Mumbai.