Deep Learning for Automatic Facial Expression Recognition: Leveraging Local Features and Convolutional Neural Networks

Pothula Jayanthi

Department of Computer Science Sri Venkateswara University, Tirupati

Abstract— *Facial expressions are crucial for human interactions and emotional analysis. Recognizing them automatically from images is challenging but has diverse applications like driver safety, human-computer interaction, and healthcare. We propose a deep-learning method comprising local feature extraction using a gravitational force descriptor and a deep convolutional neural network (DCNN) with two branches for geometric and holistic feature extraction. A score-level fusion technique computes the final classification score.*

I. INTRODUCTION

Facial expression recognition (FER) is vital for its resemblance to human coding skills and its role in affective computing. FER is essential for understanding non-verbal communication and sentiment analysis, automatically detecting emotions like neutral, anger, disgust, fear, happiness, sadness, and surprise from images or videos. This recognition is crucial for various applications such as driver safety, healthcare, and cognitive science, emphasizing its importance in community interactions and emotional analysis.

II. LITERATURE REVIEW

2.1 Pattern of Local Gravitational Force (Plgf): A Novel Local Image Descriptor

D. Bhattacharjee and H. Roy (2019)

This article introduces the Pattern of Local Gravitational Force (PLGF), a novel image descriptor inspired by the Law of Universal Gravitation. It combines two components: the Pattern of Local Gravitational Force Magnitude (PLGFM) and the Pattern of Local Gravitational Force Angle (PLGFA). PLGFM encodes force magnitude, while PLGFA encodes force angles within a local neighborhood. A new noise-resistant and edge-preserving binary pattern, neighbors to center difference binary pattern (NCDBP), is proposed for magnitude encoding. Experimental results demonstrate the effectiveness of PLGF across various databases, outperforming other descriptors.

2.2 Reliable Crowdsourcing and Deep Locality Preserving Learning for Unconstrained Facial Expression Recognition.

S. Li and W. Deng (2019).

This article introduces the Real-world Affective Face Database (RAF-DB), comprising around 30,000 facial images reflecting uncontrolled poses and lighting conditions from diverse individuals. An expectation-maximization algorithm is developed to estimate emotion labels accurately, revealing the prevalence of compound emotions. To address recognition of multi-modal expressions, a novel deep locality-preserving convolutional neural network (DLP-CNN) is proposed, enhancing deep feature discriminative power by preserving locality while maximizing inter-class differences. Results demonstrate DLP-CNN outperforms existing methods for wild expression recognition.

2.3 A Thermal Infrared Face Database with Facial Landmarks and Emotion Labels.

M. Kopaczka, R. Kolk, J. Schock (2018).

In this article, the Paper providing a large number of labeled images would allow the application of current image processing methods on the example of solving challenging face analysis tasks. We introduce a high-resolution thermal facial image database with extensive manual annotations and explore how it can be used to adapt methods from the visual domain for infrared images. An evaluation of algorithm performance shows that learning algorithms either outperform available solutions or allow completely new applications that could previously not be addressed. In our conclusion, we prove that investing the effort into acquiring appropriate training data and adapting competitive algorithms is not only a viable approach in analyzing thermal infrared images but can also allow outperforming specific task-designed solutions.

2.4 Identity-Adaptive Facial Expression Recognition Through Expression Regeneration Using Conditional Generative Adversarial Networks.

H. Yang, Z. Zhang, and L. Yin (2018)

This article introduces IA-gen, a method to mitigate subject variation by regenerating facial expressions from input images. Conditional generative models produce six prototypic expressions while preserving identity information. Fine-tuning a CNN (FER-Net) for expression classification further refines the process. Evaluation on multiple databases confirms the method's effectiveness in person-independent facial expression recognition.

2.5 Facial Expression Recognition with Neighborhood-Aware Edge Directional Pattern (NEDP).

M. Abdullah-Al-Wadud, B. Ryu (2018)

In this article, we propose a novel local descriptor named Neighborhood-aware Edge Directional Pattern (NEDP) to overcome such limitations. We introduce template-orientations for the neighboring pixels, which give importance to the gradients in consistent edge directions, prioritizing the specific neighbors falling in the direction of the local edge to represent the shape of the local textures, unambiguously. Moreover, due to the effective management of the featureless regions, no such region is erroneously encoded as a feature by NEDP. Experiments of the performances for person-independent recognition on benchmark expression datasets also show that NEDP performs better than other existing descriptors, and thereby, improves the overall performance of facial expression recognition.

2.6 Facial Expression Recognition with Faster R-CNN.

J. Li, D. Zhang, J. Zhang, J. Zhang (2017).

In this article, we proposed a method of Faster R-CNN (Faster Regions with Convolutional Neural Network Features) for facial expression recognition in this paper. Firstly, the facial expression image is normalized and the implicit features are extracted by using the trainable convolution kernel. Then, the maximum pooling is used to reduce the dimensions of the extracted implicit features. Finally, the Softmax classifier and regression layer is used to classify the facial expressions and predict boundary box of the test sample, respectively. The dataset is provided by Chinese Linguistic Data Consortium (CLDC), which is composed of multimodal emotional audio and video data. Experimental results show the performance and the generalization ability of the Faster R-CNN for facial expression recognition. The value of the mAP is around 0.82.

2.7 Facial Expression Recognition Via Deep Learning.

A. Fathallah, L. Abdi (2017).

In this article, we demonstrated the new architecture network based on CNN for facial expressions recognition. We fine-tuned our architecture with Visual Geometry Group model (VGG) to improve results. To evaluate our architecture, we tested it with many largely public databases (CK+, MUG, and RAFD). Obtained results show that the CNN approach is very effective in image expression recognition on many public databases which achieve an improvement in facial expression analysis.

2.8 Survey On RGB, 3d, Thermal, And Multimodal Approaches for Facial Expression Recognition.

C. A. Corning, M. O. Simon, J. F. Cohn (2016).

In this article, the paper presents a general overview of automatic RGB, 3D, thermal and multimodal facial expression analysis. We define a new taxonomy for the field, encompassing all steps from face detection to facial expression recognition, and describe and classify the state-of-the-art methods accordingly. We also present the important datasets and the bench-marking of most influential methods. We conclude with a general discussion about trends, important questions and future lines of research.

2.9 Automatic Pain Assessment with Facial Activity Descriptors.

P. Werner, A. Al-Hamadi (2016)

In this article, we propose a novel feature set for describing facial actions and their dynamics, which we call facial activity descriptors. We apply them to detect pain and estimate the pain intensity. Automatic recognition systems may contribute to overcome this problem by facilitating objective and continuous assessment. The proposed method outperforms previous state-of-the-art approaches in sequence-level pain classification on both, the BioVid Heat Pain and the UNBC-McMaster Shoulder Pain Expression database. We further discuss major challenges of pain recognition research, benefits of temporal integration, and shortcomings of widely used frame-based pain intensity ground truth.

2.10 Facial Expression Recognition in Video with Multiple Feature Fusion.

J. Chen, Z. Chen, Z. Chi (2016).

In this article, we analyse an effective framework to address this issue in this paper. In our study, both visual modalities (face images) and audio modalities (speech) are utilized. A new feature descriptor called Histogram of Oriented Gradients from Three Orthogonal Planes (HOG-TOP) is proposed to extract dynamic textures from video sequences to characterize facial appearance changes. And a new effective geometric feature derived from the warp transformation of facial landmarks is proposed to capture facial configuration changes. We applied the multiple feature fusion to tackle the video-based facial expression recognition problems under lab-controlled environment and in the wild, respectively. Experiments conducted on the extended Cohn-Kanade (CK+) database and the Acted Facial Expression in Wild (AFEW) 4.0 database show that our approach is robust in dealing with video-based facial expression recognition problems under lab-controlled environment and in the wild compared with the other state-of-the-art methods.

Problem Definition: Facial expressions are vital for community interactions and emotional analysis, with applications ranging from driver safety to healthcare. Facial expression recognition (FER) is crucial for affective computing, enabling the automatic recognition of seven basic emotions. Despite efforts, recognizing expressions, especially in real-world settings, remains challenging. Proposed methods include various techniques like Gobar wavelets, Haar wavelet, Local Binary Pattern (LBP), Histogram of Oriented Gradients (HOG), Histogram of Bunched Intensity Values (HBIV), and Dynamic Bayesian Network (DBN). Geometric features are also utilized, transforming images into primitives like corners, edges, and curves.

Drawbacks

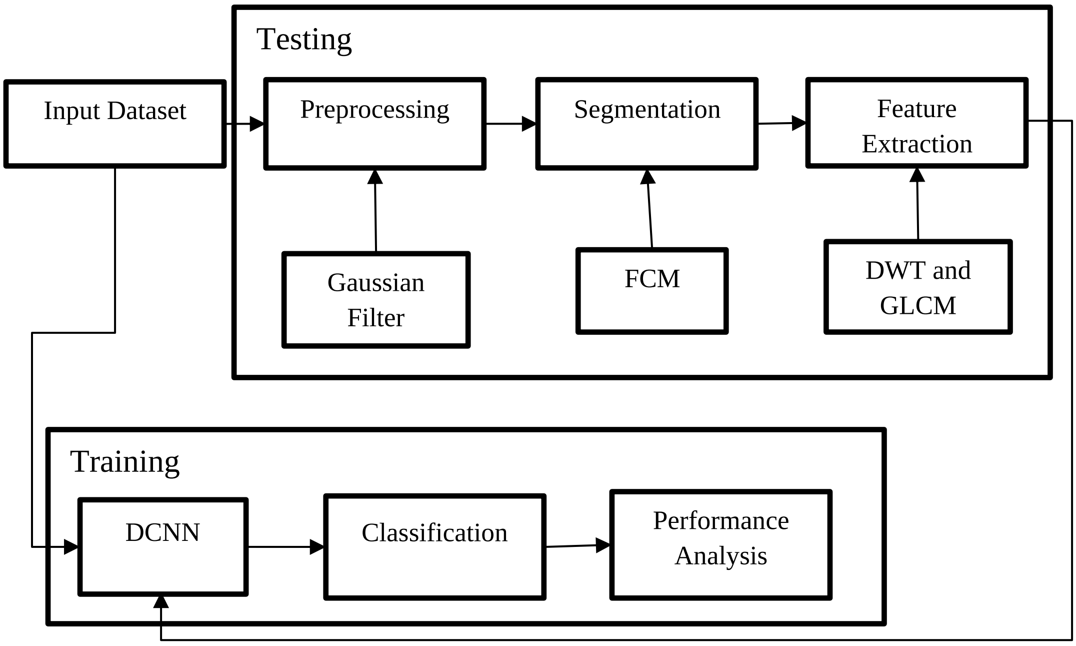
- Difficult to achieve better performance on large data.
- Traditional approaches are not up to the real FER application requirements.
- Require high computational cost.
- Occupied Large Space.

III. PROPOSED WORK

Facial expressions are crucial for social interactions and emotion analysis, with applications spanning driver safety to healthcare. Our method employs deep learning to identify facial expressions, using a combination of local gravitational force descriptor and deep convolutional neural networks (DCNN). DCNN, known for its effectiveness in image classification, extracts both geometric and holistic features. Score-level fusion is then used to compute the final classification score.

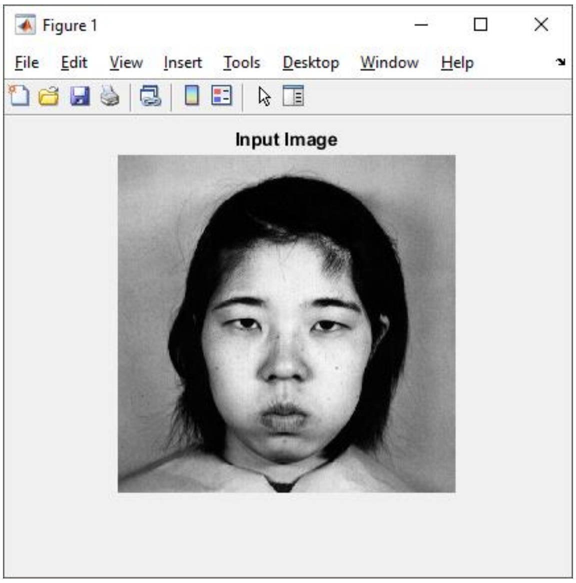
3.1 Advantages

- DCNN for FER was designed to provide better discrimination ability by combining the central loss function and the verification recognition model.
- Better Performance.
- Accurate results can be obtained, when compare to the other system.



IV. IMPLEMENTATION

- Datasets
- Preprocessing
- Segmentation
- Feature Extraction
- Classification



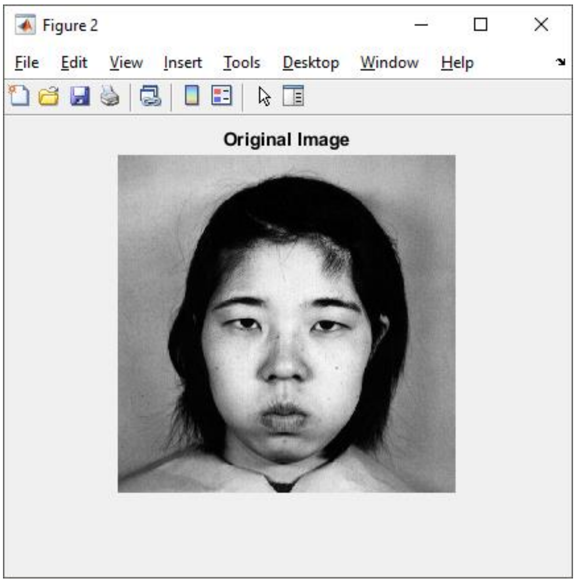



Figure 3
File Edit View Insert Tools Desktop Window Help
Gaussian Filtered Image

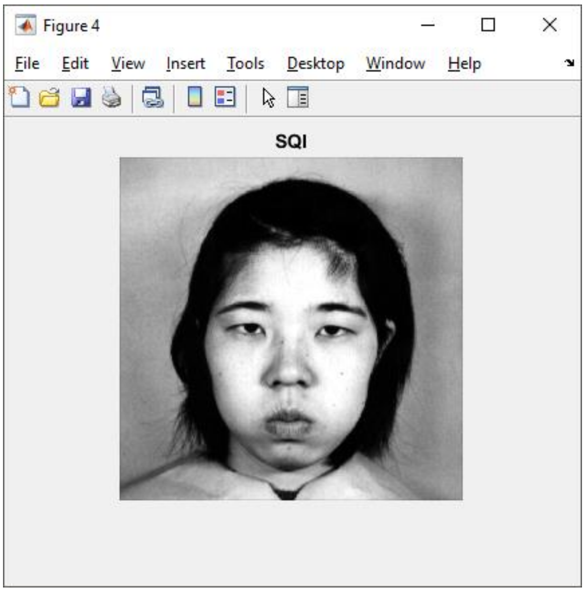


Figure 4
File Edit View Insert Tools Desktop Window Help
SQI

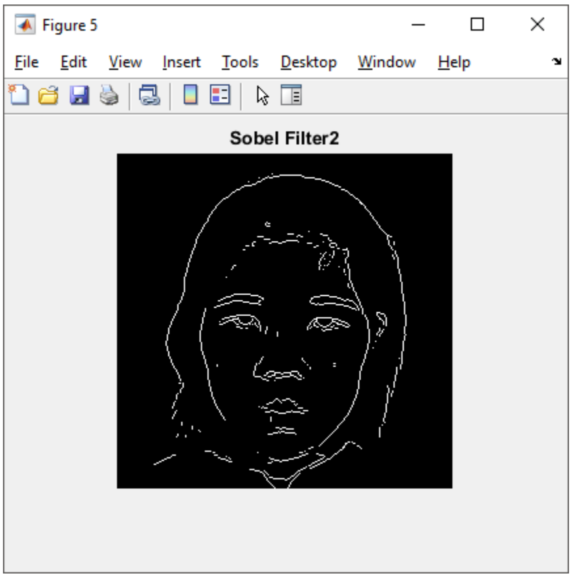


Figure 5
File Edit View Insert Tools Desktop Window Help
Sobel Filter2

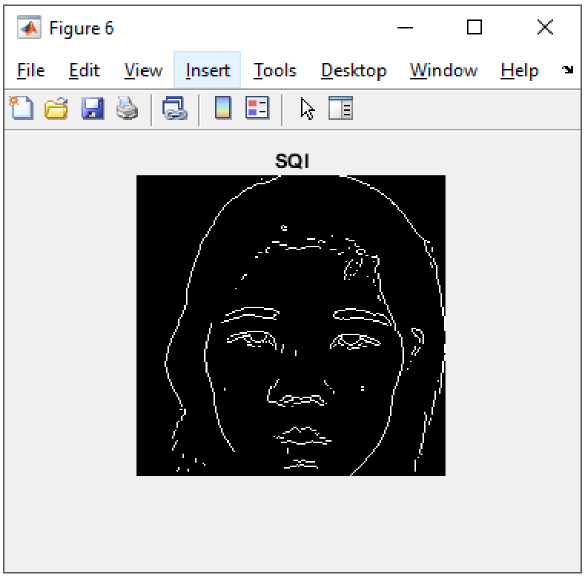


Figure 6
File Edit View Insert Tools Desktop Window Help
SQI

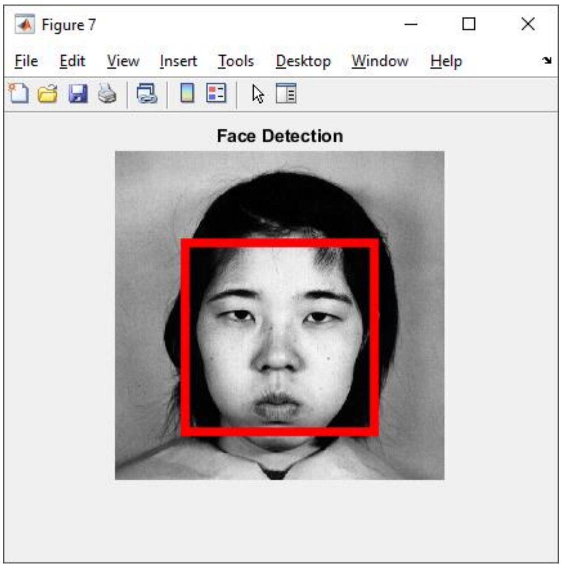


Figure 7
File Edit View Insert Tools Desktop Window Help
Face Detection

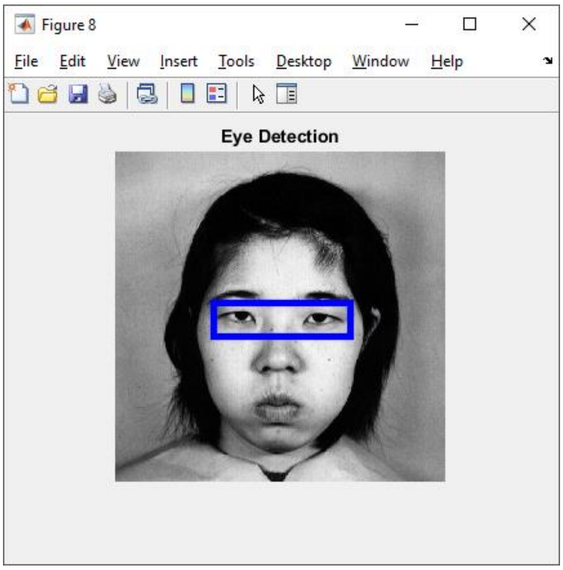
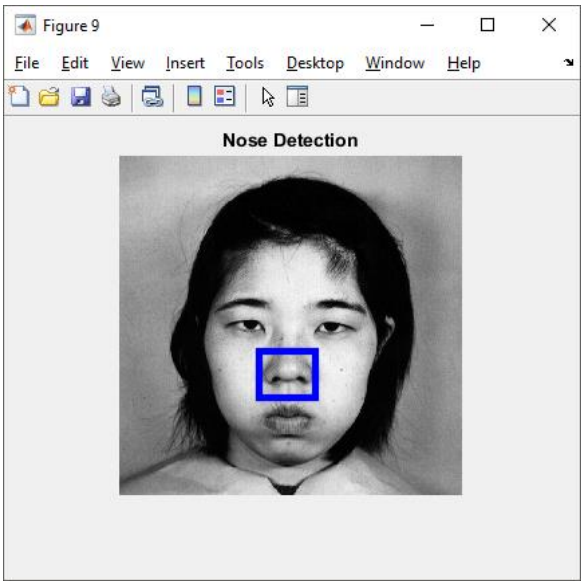
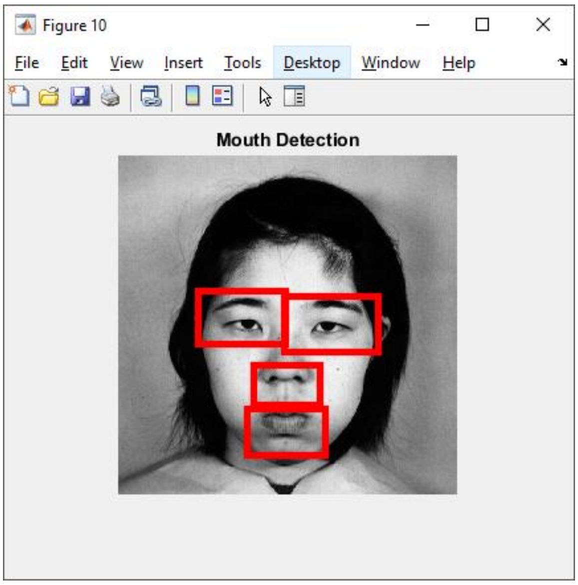
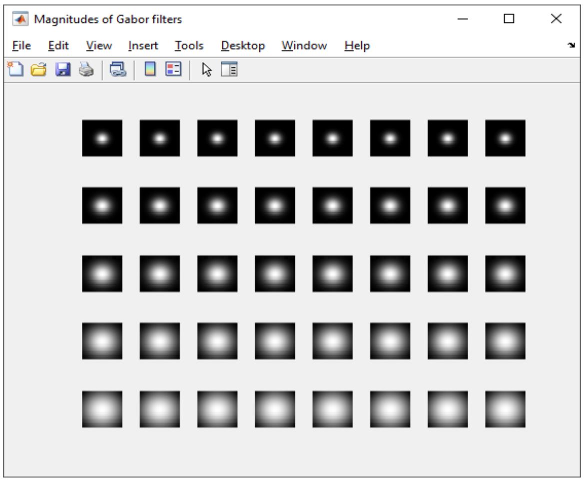
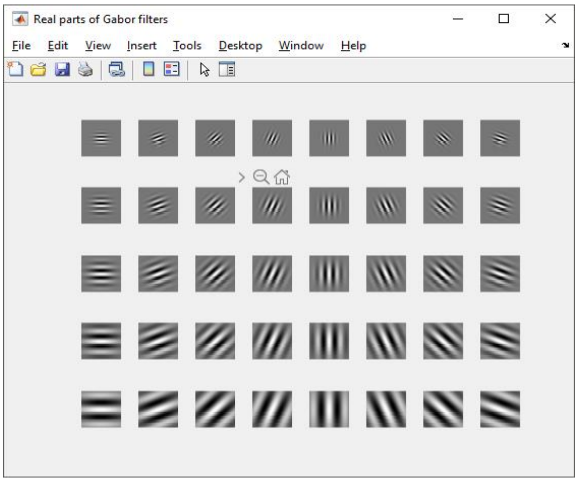


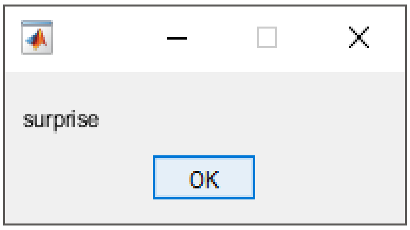
Figure 8
File Edit View Insert Tools Desktop Window Help
Eye Detection











V. CONCLUSION

Facial expressions in natural settings differ from lab-controlled environments. This study focuses on JAFFE, CK+, and KDEF databases. A novel DCNN framework extracts holistic features, preceded by a GF-based edge descriptor for low-level features. The softmax classifier computes probabilities for seven expressions, and score-level fusion combines outputs. The method achieves improved recognition accuracy, showing the effectiveness of combining local and holistic features for facial expression recognition.

REFERENCES

[1] M. Owayjan, R. Achkar and M. Iskandar, "Face Detection with Expression Recognition using Artificial Neural Networks," *2016 3rd Middle East Conference on Biomedical Engineering (MECBME)*, 2016, pp. 115-119, doi: 10.1109/MECBME.2016.7745421.

[2] Pavithra, P & Ganesh, A.. (2011). DETECTION OF HUMAN FACIAL BEHAVIORAL EXPRESSION USING IMAGE PROCESSING. ICTACT JOURNAL ON IMAGE AND VIDEO PROCESSING. 1. 976-9102. 10.21917/ijivp.2011.0023.

[3] Kumbhar, Mahesh & Jadhav, Ashish & Patil, Dr. Manasi. (2012). Facial Expression Recognition Based on Image Feature. International Journal of Computer and Communication Engineering,. 1. 117-119. 10.7763/IJCCE.2012.V1.33.

[4] C. Juanjuan, Z. Zheng, S. Han, and Z. Gang, "Facial expression recognition based on PCA reconstruction," in International Conference on Computer Science and Education, 2010, pp. 195-198.

[5] Cohen, N. Sebe, A. Garg, L.S. Chen, and T.S. Huang, "Facial expression recognition from video sequences: temporal and static modeling," Computer Vision and Image Understanding, vol. 91, no. 1, pp. 160-187, 2003.

[6] Y. Wang, H. Ai, B. Wu, and C. Huang, "Real time facial expression recognition with adaboost," in Proceedings of the 17th International Conference on Pattern Recognition, 2004, pp. 926-929.

[7] D.T. Lin, "Facial expression classification using PCA and hierarchical radial basis function network," Journal of information science and engineering, vol. 22, no. 5, pp. 1033-1046, 2006.