

EFFICIENT METHODS FOR FACIAL MICRO-EXPRESSIONS DETECTION & CLASSIFICATION

Rahul Yadav

Research Scholar, ECE Department, DCR University Science & Technology,
Murthal, Sonapat, Haryana 123401, India
rahuldevyadav@outlook.com

Priyanka

Professor, ECE Department, DCR University Science & Technology,
Murthal, Sonapat, Haryana 123401, India
drpriyanka.ece@dcrustm.org

Priyanka Kacker

Assistant Professor, IBS, National Forensic Science University,
Gandhinagar, Gujarat, India
drkacker@gfsu.edu.in

Abstract

Micro-expressions (ME) are a form of facial expression that lasts just a few milliseconds. Due to its use in revealing implicit human intentions, particularly in high-stakes situations, ME detection and classification is a burgeoning field of research. In recent studies, it has been found that the apex frame contains maximum facial movement conveying expression information. But they last for very few milliseconds, hence detecting the apex frame and classifying the ME is a very challenging task. This paper addresses this problem by proposing a training-free method for ME detection and an ME classifier using hybrid features. For ME detection, video frames are preprocessed to remove head movement and align faces. Then, from five local facial regions, Local Binary Patterns (LBP) are extracted. These regions are selected to reduce the effects of head eye blinks and changing eye gaze. Then the apex frame is detected using the Frobenius norm on extracted LBP features. For ME classification, hybrid features: Frame Difference (FD), Optical Flow Magnitude (OF-Mag), and Optical Flow Histogram of Oriented Gradients are extracted from regions around facial landmarks and are used to classify micro-expressions. These regions are selected to reduce the effects of head movement, eye blinks, and changing eye gaze. The proposed methods are tested on publicly available benchmark datasets CASME II. On the CASME II, the ME detection method achieved 0.1137 Normalized Mean Absolute Error (NMAE), 0.0720 Normalized Standard Error (NSE), and the ME classifier achieved 67.07 percent and 0.69, accuracy and a f1-score respectively. The results reveal that the proposed methods excelled the existing methods.

Keywords: Apex Frame Detector; Machine Learning; Micro-Expression Classification; Local Binary Pattern (LBP); Support Vector Machine;

1. Introduction

Involuntary facial movements in response to emotional stimuli are known as Micro-Expressions (MEs) [1]. ME can be used for clinical diagnosis, national security, and interrogation, revealing concealed emotions in high-stakes circumstances. ME are distinct from ordinary facial expressions in that they are transient (0.25 to 0.5 second), low-intensity, and appear with a limited number of facial action units [2]. ME are hard to identify and characterize due to the aforementioned characteristics.

In general, ME analysis consists of two key tasks: detection and classification. The detection task is used to discover the frames having ME, whereas the classification task is used to classify ME into respective emotional expressions[3]–[5]. The majority of current ME recognition research relies on full video frames[6]–[10]. A snapshot taken at the apex of an expression can provide vital information about facial expression recognition [11]. For instance, it can transmit an emotion message. An expression's apex can provide critical information about a facial expression's recognition. This is because capturing the apex can transmit an emotion message.

Redundant information in clips of the ME classification system can reduce the recognition performance of the individual [3]. The frames containing ME, on the other hand, provide useful information for ME classification.

Furthermore, ME classifier using apex frame and bi-weighted orientation optical flow feature presented by [12]. However, there have been few research that look into the role of the apex frame in ME classification. In recent years, CNN (Convolutional Neural Networks) methods are effectively used for facial expressions classification [13], [14] and then some researchers started using CNN for ME classification [15], [16]. However, as compared to hand-crafted methods, their proposed methods significantly reduce performance[7]. It's because ME databases are quite limited, and changes in ME are quite minimal.

In this paper, initially ME apex frame is detected. Most known spontaneous ME apex frame spotting approaches, such as optical flow-based methods[16], estimate the facial muscle change to detect the apex frame in given video frames. This method is prone to errors since it does not detect the temporal dimension of the muscle change.

The ME classification problem is introduced in the second half of the study. Humans can be trained to classify ME, but their accuracy is always weak [17]. Humans' limited ability to understand ME effectively necessitates the development of automated systems that can do so automatically. Automatic ME classification is the use of computer-based approaches for ME classification. Supervised machine learning algorithms such as Random Forest, Neural Networks, Decision Trees, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) [18] have been used for ME classification problems in the past. SVMs are widely used for ME classification tasks due to their high generalization efficiency.

To classify spontaneous micro-expressions, hybrid features, which are a combination of three features, Frame Difference (FD), Optical Flow Magnitude (OF-Mag), and Optical Flow Histogram of Directed Feature (OF-HOG), are utilized. Face landmark points are used to extract these features from Regions of Interest (ROI). To account for global head movement, eye blinks, and changing gaze, these ROIs are chosen.

This paper proposes a training-free ME detection system and a ME classifier based on hybrid features, with the following notable and novel contributions:

- A training free automatic ME Detection system for frame-wise ME apex frame detection is proposed.
- An automatic classification system for ME is proposed. This system uses Hybrid features to classify ME.
- Pre-processing processes are meticulously developed to correct the face's positioning and register it in order to minimize the influence of head movement.
- The most relevant face Region of Interest (RoI) for feature extraction is obtained using action units (AUs). This allows you to collect the most ME motion-related information while removing the unnecessary eye blinking motion.

The work is divided into five sections: Section II presents the previous literature, while sections III discuss the proposed methods for detecting and classifying MEs. The experiments are discussed and the findings are analyzed in Section IV. The suggested methods and results are summarized in Section V.

2. Literature Review

This section provides a summary of the recent research on Micro-Expression detection and classification. ME is composed of Onset, apex, and offset frames. The onset frame depicts the moment when a facial expression first appears on the face, the apex frame depicts a face with a peak expression, and the offset frame depicts a face with fading expression [19]. In the CASMEII [20] dataset, onset, apex and offset frames are manually annotated. Apex frame is very informative because it consists maximum facial movement. As a result, apex frame detection is important for ME classification and researchers intended to automate ME apex frame detection methods.

Yen *et. al.* [21] proposed first apex detection method using LBP features and Constrained Local Model (CLM). The apex frame was detected using highest frame distance between two consecutive frames. The drawback of this approach, the average distance between the ground truth and detected apex frame was presented. This led to incorrect results. Liong *et. al.* [16] used many facial regions and LBP features with optical flow to detect apex frame. The efficiency of ME apex frame detector is further improved using HOOOF (Histogram Oriented Optical Flow) features [22]. [23] used the 3D FFT (Fast Fourier Transform) method for ME detection. The Methods achieved good results but they merely considered comparative small changes with respect to neighboring frames. The ME apex detection performance can further be improved if all the frames in video can be compared instead of neighboring frame only.

For ME classification, many machine learning models are available in the literature. The majority classification method used LBP features and its variants for the task. LBP-TOP (LBP from Three orthogonal planes) [24] used both spatial and temporal information's. LBP-TOP method is further improved by using six intersection points in LBP, it reduces the superfluous information from LBP-TOP. Other features like quantized patterns [25] and image gradients histograms [26] are also explored for ME classification. In the all of the above methods, all the video frames of the videos were used. But still, these method does not perform significantly and still lot of improvement can be done.

Recently many deep-learning algorithms [4], [23], [27]–[29] are also used for ME classification. Most of the method used transfer learning, CNN (Convolutional Neural Networks) architecture trained on special features like

optical flow [28] and FFT features [23]. While some method used pre-trained CNN models for classification[4], [29]. Since the labeled data for spontaneous Micro expression is limited, e.g., CASME II data has total 254 videos having seven expressions. For this reason, deep-learning method is not suitable for the ME classification. This indicate that machine learning method are still is more better choice than deep learning methods. For using machine learning methods, we need use hybrid features that can be used in combination for overcome their limitations.

3. Proposed Method

In this section, proposed methods for ME apex frame detection and classification are explained in details. The general data pipeline is shown in Figure 1. The proposed methods contain to processes, Apex frame detection and ME classifier. In the subsequent subsections, each process is explained in detail.

3.1. ME Apex Frame Detector

The ME apex frame detection has mainly for steps, Preprocessing input video frames, feature extraction, post processing and apex frame detection. In the following subsection these steps are explained in details.

3.1.1. Preprocessing video frames

In this step, first face is detected using [30] Convolutional Neural Network (CNN) face detection algorithm for

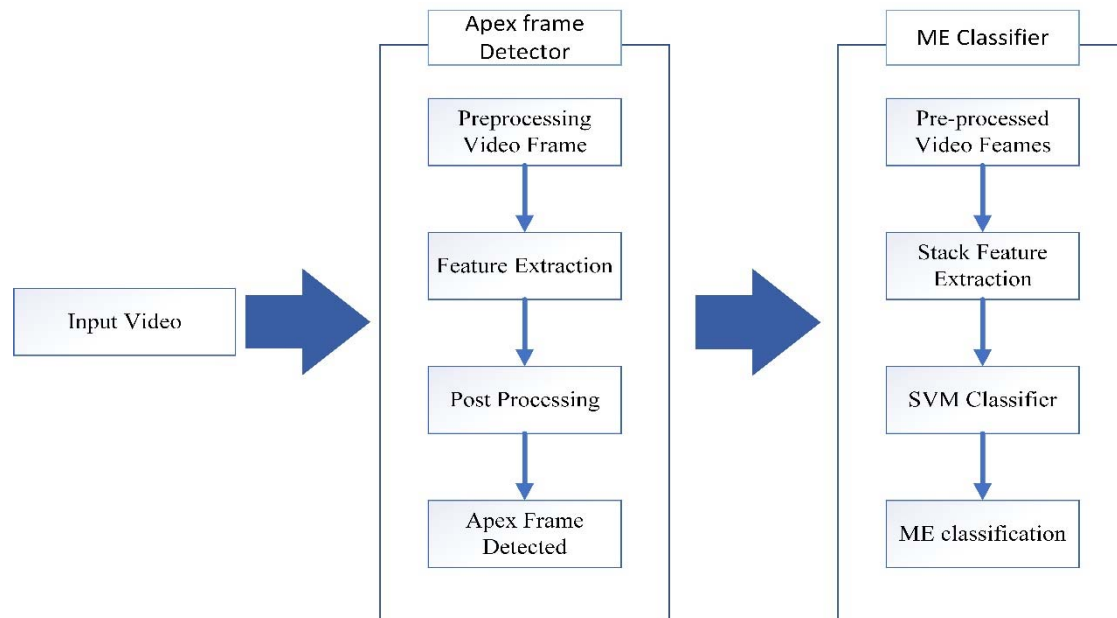


Fig. 1. General frame work for Micro-expression Analysis, i.e., Apex frame detection and ME classification.

face detection. The face detector is light and has higher processing speed. Once face is detected then the facial landmarks are obtained using [31]. The facial landmarks are used for correcting the orientation and align face. In rectifying face orientation, we correct the roll of the face, i.e., roll is 0°. Orientation of face is done to reduce the effect of head movements. This is done using eye key-points (landmark points obtained). If (x_l, y_l) and (x_r, y_r) are coordinates of left eye and right eye points respectively. Then orientation of face is calculated as follows:

$$(X_{lc}, Y_{lc}) = \left(\frac{1}{l} \sum_{l=1}^6 x_l, \frac{1}{l} \sum_{l=1}^6 y_l \right) \quad (1a)$$

$$(X_{rc}, Y_{rc}) = \left(\frac{1}{r} \sum_{r=1}^6 x_r, \frac{1}{r} \sum_{r=1}^6 y_r \right) \quad (1b)$$

$$dX = X_{rc} - X_{lc} \quad (1c)$$

$$dY = Y_{rc} - Y_{lc} \quad (1d)$$

$$\Phi = \tan^{-1} \frac{dY}{dX} \quad (1e)$$

Given the coordinates of eyes points, the Center coordinates of each eye are estimated using Eq. 1a and 1b. Using center coordinates slope line Φ can be easily be calculated using equations Eq. 1c-e. Image is rotated by Φ degree to rectify the orientation.

Once the orientation of faces is rectified, they are aligned using warp affine transformation. Nose tip point is used for align faces. For transformation, the offset between reference frame and current frame nose point is calculated. After calculating offset 2×3 translation matrix is formed for transformation. Using this translation matrix, the frames are aligned.

3.1.2. Feature Extraction

The Local Binary Pattern (LBP) features [32] are extracted from five local facial regions are cropped according to Facial Action Coding System (FACS) [33], as shown in Figure 2.

LBP are 2-Dimensional textural features which measures, spatial patterns by comparing the pixel values in local neighborhood. LBP texture feature extractor is a convenient and effective texture operator that labels the pixels in an image by thresholding the pixels in its direct proximity and treating the result as a binary number. For any given pixel location (x_c, y_c) in an image for P sampling points and R , the LBP feature is given as:

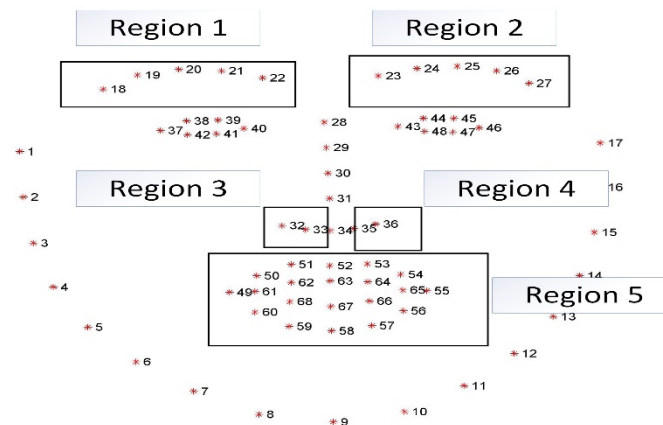


Fig. 2. Local facial regions used for feature extraction.

$$LBP_{P,R} = \sum_{p=0}^P s(x_c - x_p) 2^p \quad (2)$$

Where,

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

After extracting LBP features from five regions, they are horizontally stacked.

3.1.3. Post-processing & Apex frame detection

To detect to apex frame, Frobenius norm is used. This used measure the similarity frame stacked facial region LBP features. Since the apex frame local regional textural features differs from rest frame, it will be highly dissimilar from other frames and would have minimum norm value. For any given vector \mathbf{A} , it is given as:

$$\|A\| = \sqrt{\sum_{i=0}^m \sum_{j=0}^n |a_{i,j}|^2} \quad (3)$$

Norm results are further processed using Savitsky Golay filter [34] which reduces the false positives. The apex frame is then detected using selecting frame with minimum norm values.

3.2. ME Classification

Method proposed for classifying ME combines, OF (Optical Flow) magnitude and frame difference information, and OF-HOG (Optical Flow Histogram Oriented Gradients) features extracted from facial regions, these features are then feed in SVM classifier for classification. The framework of the ME classification method is shown in Figure 3. In this subsection, feature extraction method followed by description of ROI are discussed in subsequent subsections. The preprocessing step is same as used for apex detection.

3.2.1. Feature Extraction

Three features are used in the proposed methodology, FD (Frame Difference), OF (Optical-Flow) magnitude and OF-HOG (Optical Flow - Histogram of Oriented Gradients). FD features are extracted by calculating the pixel-

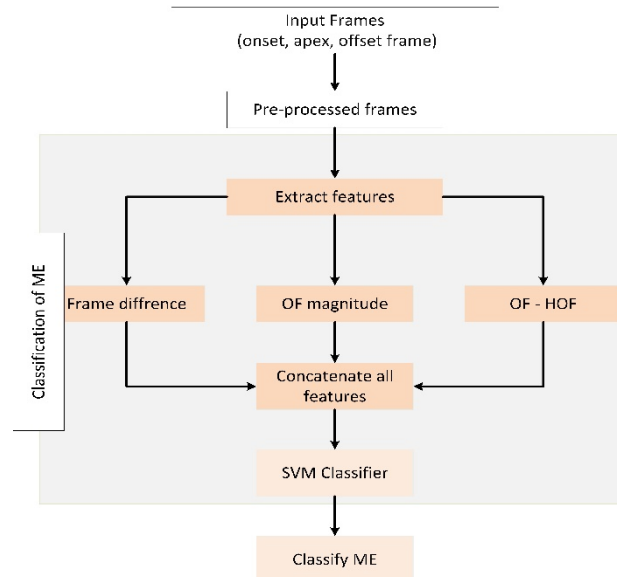


Fig. 3. General frame work of the proposed ME (Micro-Expression) classification.

wise difference between frames. The frame difference between apex and onset frame are calculated. If I_{apex} is apex frame and I_{onset} is onset frame then FD, δ is given as:

$$\delta = I_{apex} - I_{onset} \quad (4)$$

For remaining two features we use dense OF algorithm proposed by [35]. Optical flow shows the instantaneous velocity of an object's movement on the observation plane. It uses the relationship between the frames and the adjacent frames to determine the object's motion. It assumes that the object does not change in color in the two consecutive frames. If the, object pixel intensity in (x,y) location at time instance t is defined as $I(x,y,t)$ is moved to new location $(\Delta x, \Delta y)$ at time instance (Δt) , then assuming that the pixel intensities between the two frames are unchanged, i.e.:

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \quad (5)$$

Using Taylor expansion of R.H.S of eq. (5):

$$\frac{dI}{dx} \delta x + \frac{dI}{dy} \delta y + \frac{dI}{dt} \delta t = 0 \quad (6)$$

Optical flow equation is obtained by dividing equation (6) by δt :

$$\frac{dI}{dx} u + \frac{dI}{dy} v + \frac{dI}{dt} = 0 \quad (7)$$

Where u and v are optical flow vector in x , y directions. $\frac{dI}{dx}$, $\frac{dI}{dy}$ and $\frac{dI}{dt}$ are gradients of pixel intensities in x , y and t axis respectively. The magnitude of optical flow, OF-mag, is obtained as follows:

$$OF - MAG = \sqrt{v^2 + u^2} \quad (8)$$

OF-MAG is calculated for apex and onset frame. For OF-HOG features OF for is measured between onset frame, apex frame and between apex frame and offset frame. Then HOG features [36] are extracted.

3.2.2. SVM Classifier

Support Vector Machine (SVM) is a robust and scalable Machine Learning model that may be used to handle linear or nonlinear classification tasks. SVM's uses hyper-planes in high-dimensional features space for classification. The hyper-planes constructed by SVM classifier has a margin's called support vectors. We used SVM classifier for classification of emotion classes. The optimal hyper-parameters for SVM classifier were obtained by exhaustive grid search. Following parameter were used for grid search:

- Regularization parameter: [1, 10, 1e2, 1e3, 5e3]
- Kernel: ['linear', 'rbf', 'poly']
- Kernel coefficient: [1e-4, 5e-4, 1e-3, 5e-3, 0.01, 0.1, 1]

4. Experiments

In this section, experiments with apex frame detection and classification on the CASME II [20] are carried out, and the findings are analyzed.

4.1. Dataset

Fu Xiaolan's CASME II [20] micro-expression database, which included 255 micro-expression video clips, was employed in the experiment. The frame rate is 200 frames per second, and the resolution is 280×340 pixels. All the subjects in dataset are of Asian ethnicity. Dataset contain total 63409 video frame but only 254 are apex frame as shown in Figure 4a.

The frame annotation like onset, apex and offset are provided in the dataset. Every subject data is used for apex frame detection except subject 4 file EP12_01f (apex frame annotation is missing). The distribution of emotion classes in dataset is shown in Figure 4b. For micro expression classification, emotions with more than 10 samples are considered as suggested in [23].

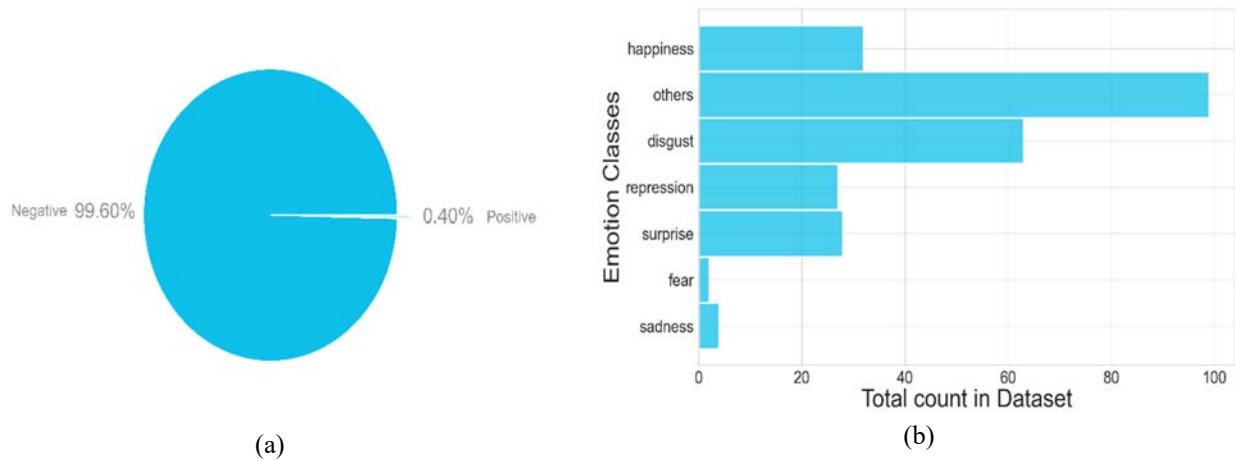


Fig. 4. (a) Distribution of apex (Positive) and non-apex frame (Negative) in given dataset. (b) Total count/numbers of samples of each emotion class in CASMEII dataset.

4.2. Experimental Settings & Implementation details

All the experiments were performed on Intel i5-8300H CPU hardware with 16GB RAM, 64-bit operating system, and NVIDIA GeForce GTX 1060 GPU (Graphics Processing Unit). The experiments are programmed using PYTHON programming language. For experiments opencv-python, dlib, skimage, scipy and scikit-learn libraries are used. For calculating features for ME classification, we used Dense Optical Flow algorithm with 5 pyramid levels of 0.5 scale and window size of 11 pixels. The 5-pixel neighborhoods are used to find polynomial expansion with 1.1 standard deviation of gaussian polynomial expansion. For calculating OF-HOG features we used optical flow vectors calculated using above parameters and HOG feature with 9 orientation bins on 2-pixel blocks of 8×8 pixel cells.

4.3. Evaluation Protocol

To comparing the results for ME apex frame detection we used *Normalized Mean Absolute Error (NMAE)* and *Normalized Standard Error (NSE)* [23]. These evaluation metrics are defined as:

$$NMAE = \frac{1}{K} \sum_{n=1}^K e'_i \quad (9)$$

Where K is total number video samples and e'_i is absolute normalized distance between actual and estimated apex frame of i^{th} video sample. e'_i is calculated by absolute distance between ground truth and apex frame (e_i) and dividing by total number of frames in i^{th} (len), i.e. $e'_i = \frac{e_i}{len}$.

NSE is calculated as follows:

$$NSE = \frac{\sqrt{e_i - \bar{e}_i}}{\sqrt{K}} \quad (10)$$

Where is \bar{e}_i mean of e_i . In the experiments for ME classification, LOOCV (Leave One Out Cross validation) protocol is used because of limited data. In this protocol one sample out for testing while remaining samples are used as training and validation data. In contemporary research, this validation method is most commonly utilized, as it maximizes data utilization by using the largest number of samples for training in each iteration; sampling is deterministic. For ME classification we used accuracy and f1-scores as comparison metrics.

$$Accuracy = \frac{1}{n} \sum_{i=0}^n 1 (predictedvalue = groundtruth) \quad (11)$$

where ‘n’ is total number of samples. F1-score is defined as:

$$F1 - score = 2 \times \frac{precision \times recall}{precision + recall} \quad (12)$$

4.4. Result & Discussion

In this section, results obtained for both ME detection and classification method respectively are presented. The obtained results are then compared with state of art methods using evaluation metrics as discussed in section 4.3 above.

Results obtained for ME apex frame detection method are compiled in TABLE I. The proposed method is compared against state of art methods such as [16], [22], [23], [32]. The proposed method performed much better than other proposed methods. The Local Binary Pattern (LBP) features with, an 8-pixel neighborhood and radius of radius 1. The proposed method performed better than other available state of art methods. The proposed method is compared against recent methods using machine learning algorithm [16], [22], [32] and deep-learning algorithm [23]. The results of ME apex frame detector have shown promising results. The lower value of NMAE confirms that the preponderance of apex frames is accurately detected. However, NSE is high due to outliers; for a few cases, the difference between the identified apex frame and ground truth is significant. The overall performance of the ME apex frame detector is excellent.

Methods	NMAE	NSE
LBP [32]	0.2037	0.0158
RHOOF [22]	0.1656	0.0159
OS-ROI [16]	0.1964	0.0118
LGCon [23]	0.1471	0.008
Proposed Method	0.1137	0.0072

Table 1. Performance comparison of proposed ME apex frame detection method with existing state of art methods using NMAE & NSE.

The ME classification method results are compiled in TABLE II. The proposed method is compared with machine learning based method [3], [7] and deep-learning based methods [4], [9], [23], [27]–[29]. The proposed method is able to achieves accuracy of 67.07% and f1-score of 0.68. The proposed method outperforms deep-learning method. This is because of two reasons, first deep-learning methods requires large dataset and second data-set used is highly imbalance.

Methods	Accuracy	F1-Score
Bi-WOOF [3]	59.67	-
STLCQP [7]	58.39	0.52
3D-FCNN [27]	59.11	-
TIM-DCNN [9]	64.90	-
VGGMAG [28]	63.21	0.59
CNN-LSTM [4]	60.96	-
LGCon [23]	65.02	0.64
LGConD [23]	62.14	0.60
Improved Mobilenet[29]	53.01	0.52
Proposed Method	67.07	0.68

Table 1. Using accuracy and the f1-score, the proposed ME-classification approach is compared to existing state-of-the-art methods.

5. Conclusion

The majority of the representative information regarding the sort of facial expressions is carried by the apex frame. The contribution of applying selective RoI's for apex frame identification and stacked features retrieved from apex frame for ME classification was investigated in this paper. A comprehensive detection and classification pipeline is proposed, first detecting the apex frame by evaluating textural properties (LBP features) taken from ROIs, and then classifying ME using stacked features and an SVM classifier. The proposed approach is tested on the CASME

II, and the results show that it is suitable for ME identification and classification. The conclusion textural features contain enough information to detect apex frames, and stacking the features that overcome each other's limitations can be employed for ME classification. In the future, we will investigate the effects of individual local regions on action unit and ME Classification.

References

- [1] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion.," *J. Pers. Soc. Psychol.*, vol. 17, no. 2, pp. 124–129, 1971, doi: <https://doi.org/10.1037/h0030377>.
- [2] P. Ekman, "Lie catching and microexpressions," *Philos. Decept.*, vol. 1, no. 2, p. 5, 2009.
- [3] S.-T. Liong, J. See, K. Wong, and R. C.-W. Phan, "Less is more: Micro-expression recognition from video using apex frame," *Signal Process. Image Commun.*, vol. 62, pp. 82–92, 2018, doi: <https://doi.org/10.1016/j.image.2017.11.006>.
- [4] D. H. Kim, W. J. Baddar, and Y. M. Ro, "Micro-Expression Recognition with Expression-State Constrained Spatio-Temporal Feature Representations," in *Proceedings of the 24th ACM International Conference on Multimedia*, 2016, pp. 382–386, doi: [10.1145/2964284.2967247](https://doi.org/10.1145/2964284.2967247).
- [5] F. Xu, J. Zhang, and J. Z. Wang, "Microexpression Identification and Categorization Using a Facial Dynamics Map," *IEEE Trans. Affect. Comput.*, vol. 8, no. 2, pp. 254–267, 2017, doi: [10.1109/TAFFC.2016.2518162](https://doi.org/10.1109/TAFFC.2016.2518162).
- [6] S.-J. Wang et al., "Micro-Expression Recognition Using Color Spaces," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 6034–6047, 2015, doi: [10.1109/TIP.2015.2496314](https://doi.org/10.1109/TIP.2015.2496314).
- [7] X. Huang, G. Zhao, X. Hong, W. Zheng, and M. Pietikäinen, "Spontaneous facial micro-expression analysis using Spatiotemporal Completed Local Quantized Patterns," *Neurocomputing*, vol. 175, pp. 564–578, 2016, doi: <https://doi.org/10.1016/j.neucom.2015.10.096>.
- [8] G. Zhao and M. Pietikäinen, "Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 915–928, 2007, doi: [10.1109/TPAMI.2007.1110](https://doi.org/10.1109/TPAMI.2007.1110).
- [9] V. Mayya, R. M. Pai, and M. M. Manohara Pai, "Combining temporal interpolation and DCNN for faster recognition of micro-expressions in video sequences," in *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2016, pp. 699–703, doi: [10.1109/ICACCI.2016.7732128](https://doi.org/10.1109/ICACCI.2016.7732128).
- [10] Y. Zong, X. Huang, W. Zheng, Z. Cui, and G. Zhao, "Learning From Hierarchical Spatiotemporal Descriptors for Micro-Expression Recognition," *IEEE Trans. Multimed.*, vol. 20, no. 11, pp. 3160–3172, 2018, doi: [10.1109/TMM.2018.2820321](https://doi.org/10.1109/TMM.2018.2820321).
- [11] P. Ekman, "Facial expression and emotion.," *Am. Psychol.*, vol. 48, no. 4, p. 384, 1993.
- [12] S.-T. Liong, J. See, K. Wong, and R. C.-W. Phan, "Automatic Micro-expression Recognition from Long Video Using a Single Spotted Apex," in *Computer Vision -- ACCV 2016 Workshops*, 2017, pp. 345–360.
- [13] F. Zhang, T. Zhang, Q. Mao, and C. Xu, "Joint Pose and Expression Modeling for Facial Expression Recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2018.
- [14] Y. Li, J. Zeng, S. Shan, and X. Chen, "Occlusion Aware Facial Expression Recognition Using CNN With Attention Mechanism," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2439–2450, 2019, doi: [10.1109/TIP.2018.2886767](https://doi.org/10.1109/TIP.2018.2886767).
- [15] D. Patel, X. Hong, and G. Zhao, "Selective deep features for micro-expression recognition," in *2016 23rd International Conference on Pattern Recognition (ICPR)*, 2016, pp. 2258–2263, doi: [10.1109/ICPR.2016.7899972](https://doi.org/10.1109/ICPR.2016.7899972).
- [16] S.-T. Liong, J. See, K. Wong, A. C. Le Ngo, Y.-H. Oh, and R. Phan, "Automatic apex frame spotting in micro-expression database," in *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, 2015, pp. 665–669, doi: [10.1109/ACPR.2015.7486586](https://doi.org/10.1109/ACPR.2015.7486586).
- [17] M. Frank, M. Herbasz, K. Sinuk, A. Keller, and C. Nolan, "I see how you feel: Training laypeople and professionals to recognize fleeting emotions," in *The Annual Meeting of the International Communication Association. Sheraton New York, New York City*, 2009, pp. 1–35.
- [18] I. P. Adegun and H. B. Vadapalli, "Facial micro-expression recognition: A machine learning approach," *Sci. African*, vol. 8, p. e00465, 2020, doi: <https://doi.org/10.1016/j.sciaf.2020.e00465>.
- [19] M. F. Valstar and M. Pantic, "Fully Automatic Recognition of the Temporal Phases of Facial Actions," *IEEE Trans. Syst. Man, Cybern. Part B*, vol. 42, no. 1, pp. 28–43, 2012, doi: [10.1109/TSMCB.2011.2163710](https://doi.org/10.1109/TSMCB.2011.2163710).
- [20] W.-J. Yan et al., "CASME II: An improved spontaneous micro-expression database and the baseline evaluation," *PLoS One*, vol. 9, no. 1, p. e86041, 2014, doi: [10.1371/journal.pone.0086041](https://doi.org/10.1371/journal.pone.0086041).
- [21] W.-J. Yan, S.-J. Wang, Y.-H. Chen, G. Zhao, and X. Fu, "Quantifying Micro-expressions with Constraint Local Model and Local Binary Pattern," in *Computer Vision - ECCV 2014 Workshops*, 2015, pp. 296–305.
- [22] H. Ma, G. An, S. Wu, and F. Yang, "A Region Histogram of Oriented Optical Flow (RHOOF) feature for apex frame spotting in micro-expression," in *2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, 2017, pp. 281–286, doi: [10.1109/ISPACS.2017.8266489](https://doi.org/10.1109/ISPACS.2017.8266489).
- [23] Y. Li, X. Huang, and G. Zhao, "Joint Local and Global Information Learning With Single Apex Frame Detection for Micro-Expression Recognition," *IEEE Trans. Image Process.*, vol. 30, pp. 249–263, 2021, doi: [10.1109/TIP.2020.3035042](https://doi.org/10.1109/TIP.2020.3035042).
- [24] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikäinen, "A Spontaneous Micro-expression Database: Inducement, collection and baseline," in *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 2013, pp. 1–6, doi: [10.1109/FG.2013.6553717](https://doi.org/10.1109/FG.2013.6553717).
- [25] Y. Wang, J. See, R. C.-W. Phan, and Y.-H. Oh, "LBP with Six Intersection Points: Reducing Redundant Information in LBP-TOP for Micro-expression Recognition," in *Computer Vision -- ACCV 2014*, 2015, pp. 525–537.
- [26] X. Li et al., "Towards Reading Hidden Emotions: A Comparative Study of Spontaneous Micro-Expression Spotting and Recognition Methods," *IEEE Trans. Affect. Comput.*, vol. 9, no. 4, pp. 563–577, 2018, doi: [10.1109/TAFFC.2017.2667642](https://doi.org/10.1109/TAFFC.2017.2667642).
- [27] J. Li, Y. Wang, J. See, and W. Liu, "Micro-expression recognition based on 3D flow convolutional neural network," *Pattern Anal. Appl.*, vol. 22, no. 4, pp. 1331–1339, 2019, doi: [10.1007/s10044-018-0757-5](https://doi.org/10.1007/s10044-018-0757-5).
- [28] Y. Li, X. Huang, and G. Zhao, "Can Micro-Expression be Recognized Based on Single Apex Frame?," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 3094–3098, doi: [10.1109/ICIP.2018.8451376](https://doi.org/10.1109/ICIP.2018.8451376).
- [29] W. Xu, H. Zheng, Z. Yang, and Y. Yang, "Micro-Expression Recognition Base on Optical Flow Features and Improved MobileNetV2.," *KSII Trans. Internet & Inf. Syst.*, vol. 15, no. 6, 2021.
- [30] R. Yadav, Priyanka, and P. Kacker, "High Speed Single-Stage Face Detector using Depthwise Convolution and Receptive Fields," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 2, 2021, doi: [10.14569/IJACSA.2021.0120292](https://doi.org/10.14569/IJACSA.2021.0120292).
- [31] "dlib.net."
- [32] W.-J. Yan, S. Wang, Y.-H. Chen, G. Zhao, and X. Fu, "Quantifying Micro-expressions with Constraint Local Model and Local Binary Pattern," in *Computer Vision - {ECCV} 2014 Workshops - Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings*,

- Part {I}*, 2014, vol. 8925, pp. 296–305, doi: 10.1007/978-3-319-16178-5_20.
- [33] P. Ekman, W. V Freisen, and S. Ancoli, “Facial signs of emotional experience.,” *J. Pers. Soc. Psychol.*, vol. 39, no. 6, pp. 1125–1134, 1980.
- [34] A. Savitzky and M. J. E. Golay, “Smoothing and Differentiation of Data by Simplified Least Squares Procedures.,” *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, 1964, doi: 10.1021/ac60214a047.
- [35] G. Farneback, “Two-Frame Motion Estimation Based on Polynomial Expansion,” in *Image Analysis*, 2003, pp. 363–370.
- [36] “HOG.” <https://scikit-image.org/docs/dev/api/skimimage.feature.html> (accessed Jun. 20, 2020).

Authors Profile



Rahul Yadav, is currently research scholar in the Department of Electronics & Communication Engineering, D.C.R University of Science & Technology, Murthal, Sonapat, Haryana, India. He has received his M.Tech in 2013. His research interests' areas are pattern recognition, signal & image processing, computer vision and time series data analysis.



Priyanka, is working as Professor at D.C.R. University of Science and Technology, Murthal, Sonipat, India. Her current areas of interest are signal processing, image processing and SAW filter design. Her highest qualification is Ph.D. in Electronics Engineering from Indian Institute of Technology, Delhi, India. She is IEEE member & has published several papers in refereed journals including IEEE Transactions & Conferences.



Priyanka Kacker, is a Senior Assistant Professor at the Institute of Behavioral Science of National Forensic Sciences University, Gandhinagar, since 2013. Before joining GFSU she was Lecturer (4 years) in The Department of Psychology, The M.S. University of Baroda, Gujarat. Micro-expressions link to deception and early detection, intervention, and rehabilitation of Juvenile and Dementia is also part of ongoing studies and research interests. She is a member of the Core Committee of Juvenile Justice Rules of Gujarat Government, which is currently engaged in a developing preliminary assessment module.