Autonomous Apex Detection and Micro-Expression Recognition using Proposed Diagonal Planes

Vida Esmaeili[a], Mahmood Mohassel Feghhi[a,*], Seyed Omid Shahdi[b]

[a]Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran.
[b]Faculty of Electrical, Biomedical and Mechatronics Engineering, Qazvin Branch, Islamic Azad University, nokhbegan Blvd., Qazvin, Iran.

Abstract

Micro-expression as the main way of non-verbal communication occurs quickly and subtle in high-risk situations. Since it cannot be misleading, it discloses the real human aim. Nonetheless, feature extraction is an arduous task due to its two particular features. To resolve this problem in this paper, we propose Local Binary Pattern on Four Diagonal Planes. In fact, we utilize it after motion magnification for feature extraction in apex detection task. Also, we combine it and optical flow for micro-expression recognition. Simulation results show that automatic micro-expression apex frame detection and micro-expression recognition is promising using our method on CASME2 and CASME in comparison with other methods.

Keywords: Apex frame detection, local binary pattern on four diagonal planes, micro-expression recognition, motion magnification, optical flow.

1. Introduction

Nowadays, people hide their true feelings to maintain social status or avoiding dangers. However, the expression of happiness, sadness, fear, surprise, hatred, anxiety, or anger appears on their faces in a fraction of a second, involuntarily. Thus, it can be useful in court, police interrogation, and etc [13].

For the first time, this uncontrollable facial expression discovered by Haggard and Isaacs while scanning psychotherapy films in 1966 [9]. Three years later, Ekman et al. [4] named it Micro-Expression (ME). They saw that a glad psychiatric patient looks sad at only 0.16 seconds in the interview film. Then, they found out that the patient had a suicide plan but she wanted to hide her true emotion. Thus, they understood the ME and deception relationship.

---

*Corresponding author

Email addresses: v.esmaeili@tabrizu.ac.ir (Vida Esmaeili), mohasselfeghhi@tabrizu.ac.ir (Mahmood Mohassel Feghhi), shahdi@qiau.ac.ir (Seyed Omid Shahdi)

In fact, a person cannot fake or imitate ME. Therefore, ME can disclose true human intention. Nevertheless, ME recognition is arduous by the naked eye and computer vision due to its short time duration (occurring in a fraction of a second from 0.04 to 0.3 seconds) and low intensity. Actually, subtle motion intensity of ME is usually as low as noise. Therefore, giving different weights to features is not suitable. Because image noise weights may be similar to ME feature weights.

To address this problem, Wang et al. [19] did compute motion intensities by the temporal accumulation of optical flows. Then, the computed weights were multiplied with Local Binary Pattern from Three Orthogonal Planes (LBP-TOP) and fed into a Linear Support Vector Machine (LSVM). They obtained 68.75% accuracy for ME recognition on CASME2. In addition, Chang et al. [2] suggested Local Optical Flow Binary Patterns on Three Orthogonal Planes (LOFBP-TOP) for ME recognition. They obtain 52.44% accuracy using this method on CASME2.

Although ME recognition is interesting due to its relationship with deception and lie detection, it still a relatively new research field with many challenges. Among the ME researches, Apex Frame Detection (AFD) is important. Because this frame shows the most motion and change of facial muscles in the ME image sequence. Hence, it shows the facial expression and we don't need to analyze all frames to ME recognition [15, 12]. Also, we can detect the ME from ordinary Facial Expression only by the time duration of the first frame (neutral face image) and the apex frame.

Since ME is short in time and subtle in intensity, the AFD is challenging. So far, the ground-truth apex frame determines manually. Manually detection is very time-consuming. Newly, researchers have tried to AFD, automatically.

In 2017, Ma et al. [16] proposed the Region Histogram of Oriented Optical Flow (RHOOF) to detect this frame automatically. They extracted HOOF feature in five key ROIs. To extract the HOOF [3] features, the OF vectors based on the orientation are captured as histogram bins and weighted by the OF magnitude. The histogram is then normalized to sum up to 1. They achieved a 30.77% and 19.04% improvement in comparison with the previous works in the CASME and CASME2 datasets. Also, using this method the Mean Absolute Error (MAE) was reduced. But the detected apex index was long from the ground-truth apex index.

Afterward, Liong et al. [14] determined the apex with the difference of a frame from the labeled apex using LBP and the divide-and-conquer procedure on a sample of CASME2. They reported the sixty-fourth frame as apex, while ground-truth apex was 63.

The RHOOF [16] ignores spatial features and calculates only the difference of motion between frames. The LBP considers only spatial features. Therefore, the above-mentioned works could not detect the exact apex. Therefore, both spatial and temporal features should be considered.

The LBP feature descriptor is simple and fast to compute. It applies to the two-dimensional gray-scale image (XY plane). In fact, LBP compares the Center Pixel (CP) gray value and the neighboring pixels gray value in one image. Thus, it codes appearance changes. However, subtle changes and motions can observe better in the image sequences. Therefore, the temporal planes could be essential to detect subtle muscle movements. XT and YT planes extract temporal features. But they show movements in one row or column of pixels.

The LBP-TOP [24] method contains XY, XT, and YT planes (Figure 1). These orthogonal planes intersected each other in CP. In this method, the sequential dynamic texture images are considered as the XY planes. Their horizontal and vertical pixels with corresponding locations are placed into the XT and YT planes, respectively. Then, the LBP code is computed on each plane. Finally, the obtained histograms from three planes are concatenated into a single histogram. Therefore, it combines extracted features from these planes. This method had well-performed under subtle facial motions [17]. Utilizing the re-parameterization of second-order Gaussian jet on it creates a more robust and reliable histograms against noise. But it has computational complexity [18].
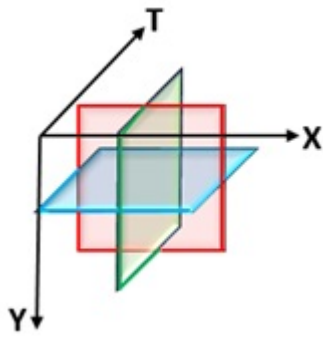
Figure 1: XY, XT, and YT planes in the LBP-TOP method shown in different colors.
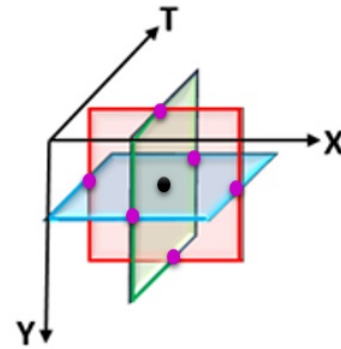

Figure 2: The intersecting points that are the neighbour points of CP shared by all three planes.

For reducing the computational complexity and redundant information, LBP with Six Intersection Points (LBP-SIP) proposed in [20]. According to Figure 2, the specified points are computed once instead of twice in this method. In other words, these pink points are sharing in the planes. Thus, it is enough that they are considered once in LBP computation. Hence, LBP-SIP reduces feature extraction time. Nonetheless, ME recognition accuracy was no too different by using LBP-SIP relative to LBP-TOP [20].

Recently, Esmaeili and Shahdi [5] proposed Cubic-LBP that combines spatial and temporal features. In fact, they computed LBP on fifteen planes in a small cube. Using this method the Mean Absolute Error (MAE) and the Standard Error (SE) were reduced from previous AFD works. However, using fifteen planes increase feature extraction time. To our knowledge, previous relevant spatio-temporal methods contain XT and YT planes [5, 6, 18, 20, 24]. However, we can exploit only key planes that each one combines spatial and temporal features alone. These planes can show motions and changes in row and column, concurrently. Hence, we reduce computation and feature extraction time in comparison with some methods such as Cubic-LBP. Meanwhile, we improve the results. In this paper, initially, the apex frame is detected from the image sequences. Afterward, ME recognition is done using the apex frame and its previous and next frame. The framework of the proposed method has shown in Figure 3. For AFD, we magnify muscle motions using Eulerian emotion magnification in ME video. Then, we convert the video to an image sequence and we apply the proposed four diagonal planes on them. The LBP is computed on these planes for feature extraction. Finally, we use the Sum of Squared Differences (SSD) to feature differences computation. The frame that shows a maximum motion is the apex. This frame and one frame before and one frame after it are used to ME recognition.

For the recognition task, we combine the optical flow and the LBP from four diagonal planes to extract spatiotemporal features. Finally, we use the LSVM for classification.

The contributions of this paper are as follows:

- A new feature descriptor method is proposed to reveal tiny muscle motions in image sequence. This method considers the spatial and temporal information, concurrently.

- Using motion magnification for small motions exaggeration in AFD task. As far as we know, motion magnification didn't use in the AFD task.

- The ground-truth apex frame is automatically detected in several samples. Unlike Liong et al. [14] and Ma et al. [16] work, the detected apex is exactly the ground truth in many samples
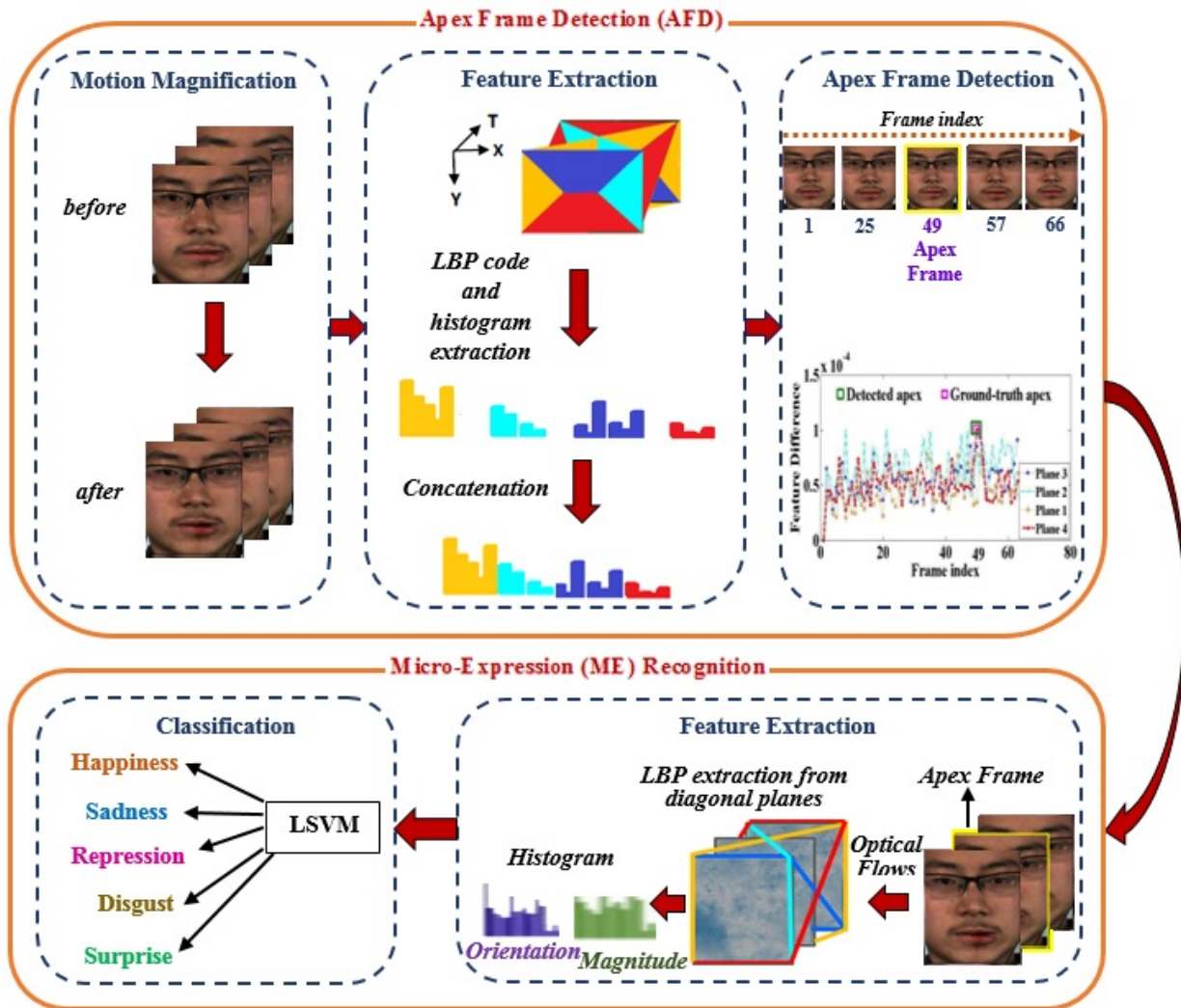
Figure 3: our proposed method.

by using our proposed method.

- The SE and MAE of our method are less than the error of the previous methods. It means that the wrong detected apex frame number has a short distance from the ground-truth frame number.

- The proposed feature descriptor method, LBP on four diagonal planes, is combined with optical flow for motion feature extraction. This work generates a robust histogram. Unlike LOFBP-TOP, our proposed method receives more meaningful information.

- Increasing the ME recognition accuracy.

The remaining parts of the paper are organized as follows: We express our proposed method in section II. In Section III, we present experimental results. Finally, the paper's conclusion and suggest future directions are given in Section IV.
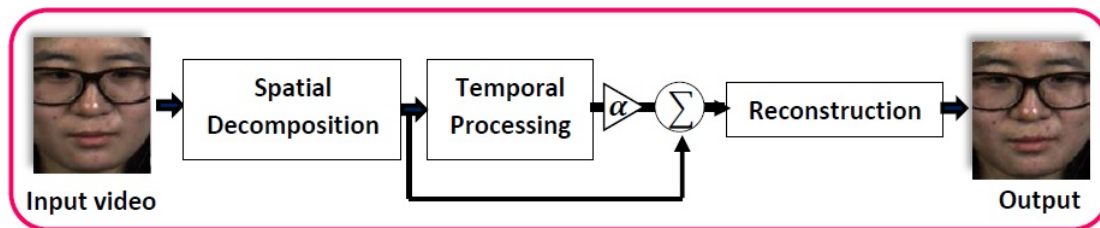
Figure 4: The EMM framework.

2. THE PROPOSED METHOD

In this section, initially, the proposed AFD is presented. Afterward, the ME recognition is done using proposed method.

2.1. AFD USING THE PROPOSED METHOD

For the proposed AFD, initially, we glimpse into the Eulerian magnification method. Then, we introduce the proposed feature extraction method. Finally, the feature differences computation is expressed to find the apex

2.1.1. Motion Magnification

The intensity of facial muscle contractions and movements in ME is so subtle as to be arduous to analyze or describe [8]. For resolving ME's subtleness that makes it impossible to identify, the motions should be magnified. In other words, magnification reveals undetectable variations and ambiguities. Thus, we use Eulerian Motion Magnification (EMM) [21]. In this method, the ME video is taken as input. It decomposes using the Laplacian pyramid structure into different spatial frequency bands. Then, low order IIR band-pass filter is applied to each band. The extracted signal is multiplied with the amplification factor. The original is added to it, and then all of them reconstructed to form output (Figure 4). Mathematically, EMM is [11]:

$$
\begin{aligned}
&\hat{I}\left(3D'\right) = I\left(3D\right) + \sum_{n} \alpha_n \frac{\partial I\left(3D\right)}{\partial T}, \\
&3D' \quad is \quad X, Y, T+1. \\
&3D \quad is \quad X, Y, T.
\end{aligned}
\tag{2.1}
$$

where I is motion intensity and α is magnification factor.

2.1.2. Feature extraction using the proposed LBP on diagonal planes method

For feature extraction, the output video is converted to a frame sequence. These frames are considered as 2D+T dimensional volume (X, Y, T) The four proposed diagonal planes are applied to it.

These planes intersect in CP. They show the motions of the facial muscle simultaneously in both horizontal and vertical directions. Four proposed planes are illustrated in Figure 3. In this figure, Plane1, Plane2, Plane3, and Plane4 are shown by mustard yellow, cyan, blue, and red colors, respectively.

Suppose $g_{T_C,C}$ is CP gray value. Coordinates of this pixel are (X_C, Y_C, T_C). The neighboring point coordinates around CP in Plane1, Plane2, Plane3, and Plane4 is given by:
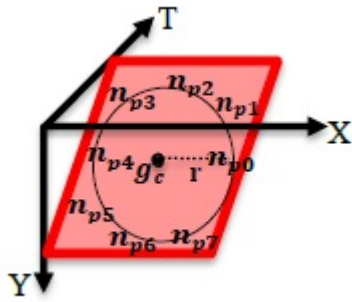
Figure 5: The neighboring points of CP on the circle with r radius.

Figure 6: Plane3 of the proposed method.

$$\left(X_C + r_X \cos\left(\frac{2\pi n_p}{N_P}\right), Y_C - r_Y \sin\left(\frac{2\pi n_p}{N_P}\right), T_C + r_T \cos\left(\frac{2\pi n_p}{N_P}\right)\right) \tag{2.2}$$

$$\left(X_C + r_X \cos\left(\frac{2\pi n_p}{N_P}\right), Y_C - r_Y \sin\left(\frac{2\pi n_p}{N_P}\right), T_C - r_T \cos\left(\frac{2\pi n_p}{N_P}\right)\right) \tag{2.3}$$

$$\left(X_C - r_X \cos\left(\frac{2\pi n_p}{N_P}\right), Y_C - r_Y \sin\left(\frac{2\pi n_p}{N_P}\right), T_C - r_T \sin\left(\frac{2\pi n_p}{N_P}\right)\right) \tag{2.4}$$

$$\left(X_C + r_X \cos\left(\frac{2\pi n_p}{N_P}\right), Y_C - r_Y \sin\left(\frac{2\pi n_p}{N_P}\right), T_C + r_T \sin\left(\frac{2\pi n_p}{N_P}\right)\right) \tag{2.5}$$

where n_p is CP's neighboring point. N_P is the number of these points. r is distance from CP to the neighbor points on the circle (Figure 5).

In fact, Figure 5 show Plane4. This plane takes bottom pixels of the frontal image, CPs of the middle image, and upper pixels of the rear image. Plane3 captures upper pixels of the frontal image, CPs of the middle image, and bottom pixels of the rear image (Figure 6). Plane1 and Plane2 contain right, middle, and left pixels in images. Thus, each of the four planes can display both of the changes in the row and the column, concurrently.

LBP code is computed on each of the introduced planes. We can get LBP using (2.6):

$$LBP = \sum_{n_p=0}^{Np-1} s(g_p - g_c)2^{n_p}$$

$$s(g_p - g_c) = \begin{cases} 1 & g_p - g_c \geq 0 \\ 0 & g_p - g_c \prec 0 \end{cases} \tag{2.6}$$

Then, final histogram of LBP obtains by concatenating four planes histogram:

$$Hist_{j,k} = \sum_{3D} g\{f_j(3D) = j\},$$

$$3D \quad is \quad X, Y, T. \tag{2.7}$$

$$j = 0, ..., n_k - 1; k = 0, ..., 3; g(C) = \begin{cases} 1, \text{if } C \text{ is true} \\ 0, \text{if } C \text{ is false} \end{cases}$$

where $f_j(3D)$ and n_k are the LBP code of CP and the number of different labels processed by the LBP operator in the k-th plane, respectively. Finally, the final histogram is normalized:

$$NHist_{j,k} = \frac{Hist_{j,k}}{\sum\limits_{m=0}^{n_k-1} Hist_{m,k}} \tag{2.8}$$

2.1.3. Feature differences computation and AFD

For AFD, the final histograms ($NHist_i$) should be compared with the neutral frame histogram ($NHist_1$). The frame that shows the majority of wrinkles is the reported apex. To find the difference between two frames, we compute SSD of two histograms as follows:

$$\sum_{i=1}^{n} (NHist_i - NHist_1)^2 \tag{2.9}$$

2.2. ME RECOGNITON USING THE PROPOSED METHOD

In this section, we present our proposed method for ME recognition. First, the apex frame and one frame before and one frame after it are taken. Then, features are extracted using the proposed method and optical flow combination. Finally, the feature vectors are fed to LSVM to facial expression recognition.

2.2.1. Feature extraction by the combination of the optical flow and the proposed LBP on diagonal planes method

The apex frame, the 1st frame prior, and the 1st frame after it are used for feature extraction. In fact, these frames are considered as 2D+T dimensional volume. Initially, optical flow computes on these frames. Since the muscle movements in ME are small, the image constraint at pixel intensity $I(X, Y, T)$ with Taylor series can be developed to get:

$$I(X + \Delta X, Y + \Delta Y, T + \Delta T)$$
$$= I(X, Y, T) + \frac{\partial I}{\partial X}\Delta X + \frac{\partial I}{\partial Y}\Delta Y + \frac{\partial I}{\partial T}\Delta T \tag{2.10}$$
$$+ high \ order \ term$$

Rearrange (2.10) to get the optical flow constraint equation:

$$\frac{\partial I}{\partial X}V_X + \frac{\partial I}{\partial Y}V_Y + \frac{\partial I}{\partial T}V_T = 0 \tag{2.11}$$

where V_Y, V_X are the vertical and horizontal components of the optical flow field. To solve (2.11), the MFSF method [7] is used for optical flow estimation. Then, the vertical and horizontal components of the optical flow field are obtained. Also, the orientation and magnitude components are calculated into 8×8 cells.

$$m = \sqrt{V_X^2 + V_Y^2}$$
$$\theta = \tan^{-1}(\frac{V_X}{V_Y})$$
$$...$$

(2.12)

Next, LBP on diagonal planes is applied to m and θ of all cells. After the LBP code computation, corresponding histograms are concatenated to form the feature vector. Therefore, optical flow and LBP from diagonal planes are combined.

2.2.2. Classification

LSVM [1] as the classifier is used for ME recognition. For the parameter selection, cross-validation is done on the training data to select a good parameter setting. For CASME and CASME2, samples are classified into 7 and 5 categories, respectively.

3. EXPERIMENTAL RESULTS

Our method implementation is performed in MATLAB 2017 and tested on a Core i7 system with 8GB RAM onboard. The details of implementation and its performance are discussed in the following.

3.1. The used datasets

Experiments are performed on two datasets. CASME [22] dataset is comprised of one hundred ninety-five samples of MEs at a 60 fps frame rate from 22 Chinese males and 13 Chinese females. It is a spontaneous dataset. For collecting spontaneous ME datasets are asked the persons to conceal their real feeling while watching the film. With this work, MEs appears involuntary.

In the CASME dataset, videos recorded using two cameras with 1280×720 and 640×480 pixels spatial resolutions. Then, videos have been converted to sequential images. Some of its profits are having neutral faces before and after each ME sample, labelling clips on onset (the occurrence of ME), apex and offset (the disappearance of ME) frames, and containing images without the head movement and irrelevant mouth movement (i.e. talking). An example of a part of the frame sequence from this dataset is shown in Figure 7. In this figure, (a), (b), and (c) labelled on the onset, apex, and offset frame, respectively.

Another spontaneous dataset that has the apex label is CASME2 [23]. It contains two hundred forty-seven ME samples. Its images are recorded by Grey GRAS -03K2C camera with a 200 fps frame rate and 280 × 240 resolution.

3.2. Motion Magnification

We convert the sequential images of the dataset to video. Then, we magnify them at $\alpha = 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, and, 21$. An example of magnification is shown in Figure 8.

According to the results, $\alpha = 7$ is suitable. Because magnification at higher levels magnifies unwanted motions such as head movements. Also, it makes artifacts. Against, magnification at 1, 3, and 5 levels don't reveal subtle muscle movements sufficiently.
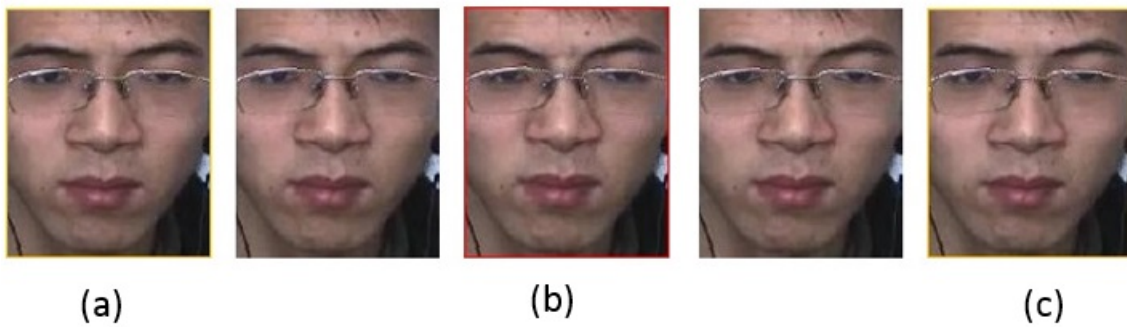
Figure 7: An example for a part of the frame sequence from the CASME [22] dataset including onset frame (a), apex frame (b), and offset frame (c).

Figure 8: An example of motion magnification.

3.3. Feature Extraction

For feature extraction, we convert the magnified ME video to sequential images. These same sized images are then converted to gray level. They are taken as a 3-dimensional array. We assume circles with unit radius along Y, X, and T axes. In addition, we consider 8 neighboring points around each CP due to its best performance. LBP code is computed on 4 proposed planes and the corresponding histograms are concatenated as a feature vector. The final histogram is obtained as a matrix with the size $4 \times 2^{Np=8}$. The elapsed time is 0.8 seconds based on the simulation run time.

3.4. AFD

The histograms should be compared with first frame histogram. Because the first frame doesn't show any motion. Feature differences are computed using SSD. The frame that has a maximum difference is the apex.

3.5. ME Recognition

In the feature extraction stage of recognition task, we compute optical flow on three frames (the apex frame and its previous and next frame). Then, we apply diagonal planes to them and compute LBP. The radii values in axes Y, and X range from 1 to 4, and $r_Y = r_X$. However, r_T is 1 due to three consecutive frames. According to the results, our method achieves the best recognition rate when $r_T = 1, r_X = r_Y = 4$, and N_P is 8. Finally, we use LSVM for classification. The datasets (CASME and CASME2) are divided randomly into two training and test sets in a ratio of 3/4 and 1/4, respectively.

3.6. Results and Discussion

A sample of the AFD by using each plane is illustrated in Figure 9. The ground-truth apex is 12 in it (i.e., sub08-EP12-2-2 in CASME [22] dataset). Also, we put the obtained results of all planes in a figure to determine the exact apex by combining four planes feature (Figure 10). As can be seen, reported apex is exactly 12. In Figure 9 and Figure 10, the vertical axis shows the Feature Difference (FD) of each frame from the first frame, and the horizontal axis is the number of frames. The percentage of the AFD using every four planes of the proposed method and total of them on CASME are shown in Figure 11. According to the results, the percentage of AFD by our proposed method is 45%. It means that in 45% of all CASME samples, the detected apex is the ground-truth apex.

Moreover, we apply the proposed method on CASME2. Our method has a good performance on it. A sample of AFD on the CASME2 illustrated in Figure 12. The forty-ninth frame is the detected and ground-truth apex in it. Finally, we compare our method with other methods by using MAE and SE computation. The MAE estimates the number of frames that the detected apex frame is off the ground-truth. It is defined for S sample size as [5, 16]:

$$MAE = \frac{1}{S} \sum_{i=1}^{S} |e_i| \qquad (3.1)$$

where e_i is the deviation from the ground-truth. If σ be sample deviation, SE is defined as [5, 16]:

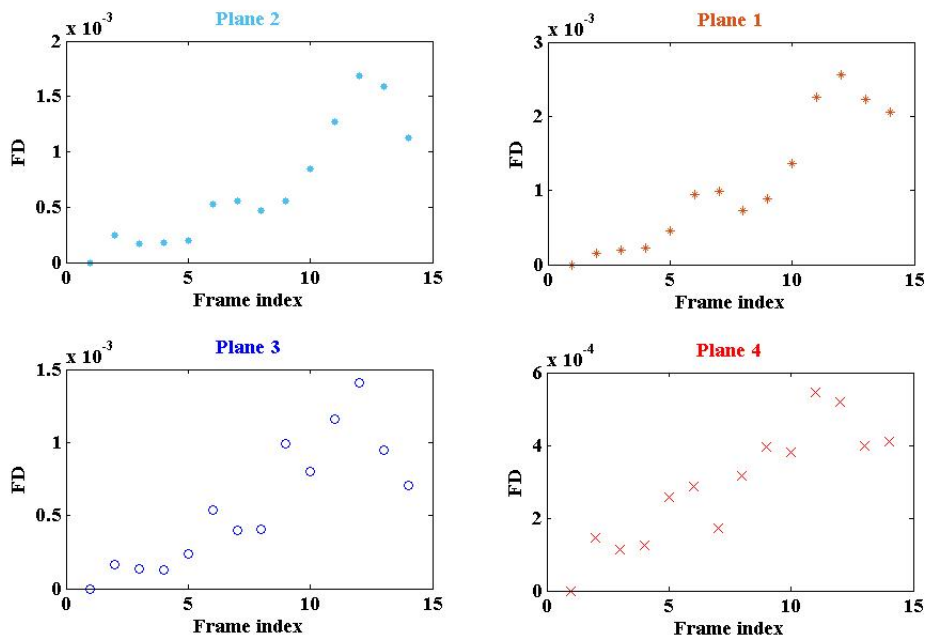$$SE = \frac{\sigma}{\sqrt{S}} \qquad (3.2)$$

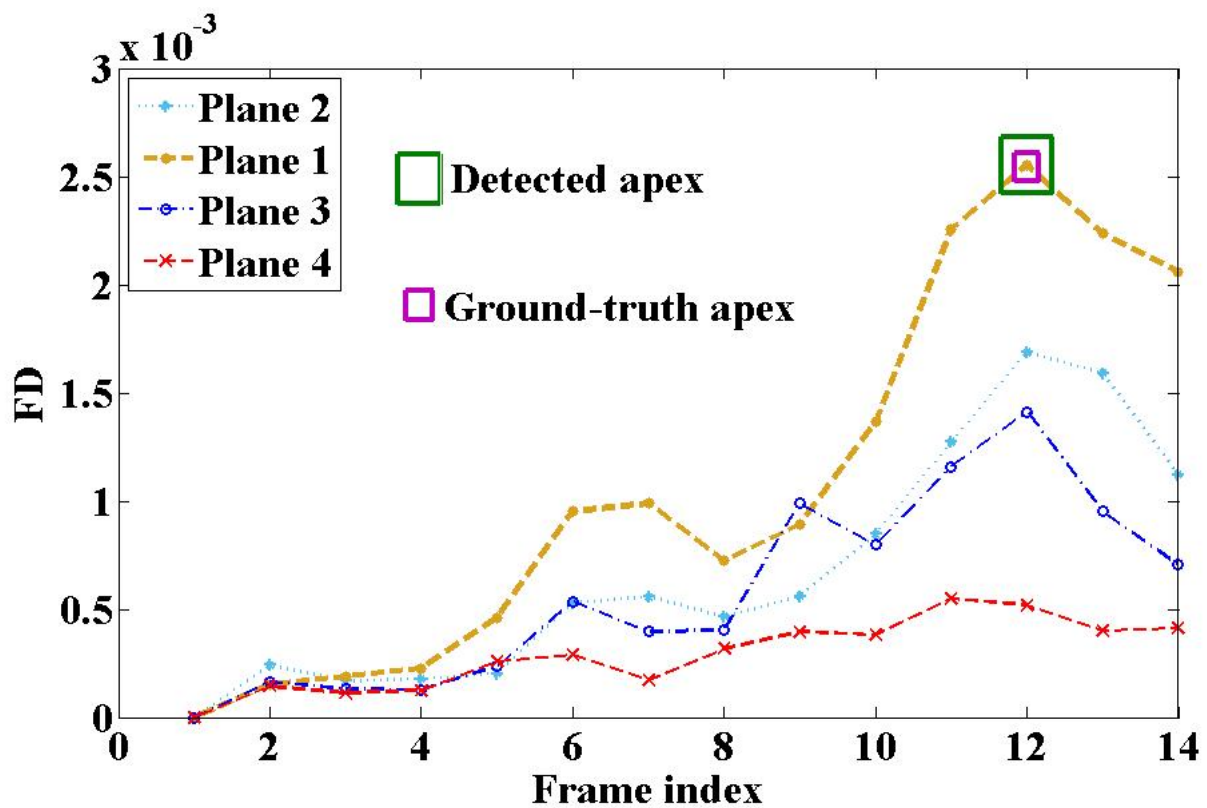Figure 9: The results of AFD in 4 planes of our proposed method on CASME.

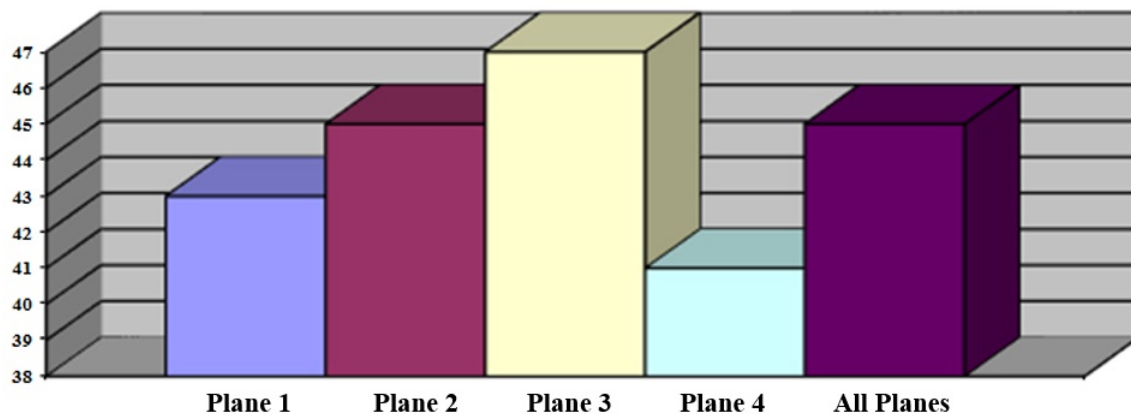Figure 10: AFD using our proposed method on CASME.

Figure 11: The percentage of AFD using our proposed method (total planes) and each plane of that on the CASME dataset.
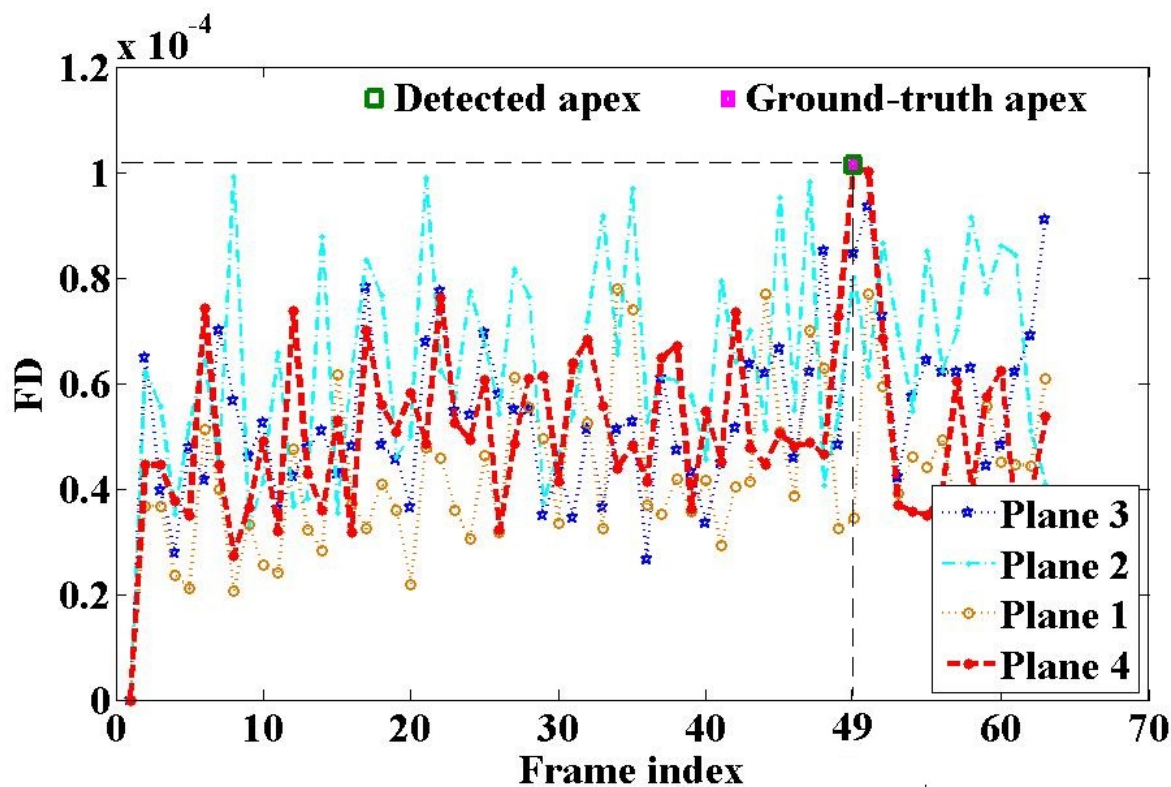

Figure 12: AFD using our proposed method on CASME2.

Autonomous Apex Detection and Micro-Expression Recognition using...
Volume 11, Special Issue, Winter and spring 2020, 483-497

495

Table 1: The results of the AFD using our proposed method on the CASME and CASME2 datasets.

Method	MAE		SE	
	CASME	CASME2	CASME	CASME2
LBP [14, 16]	5.2	13.55	0.58	0.79
RHOOF [16]	3.6	10.97	0.35	0.83
LBP-TOP [5, 24]	2.54	8.38	0.23	0.67
Cubic-LBP [5]	1.93	6.41	0.15	0.62
LBP-SIPl [6]	1.76	-	0.12	-
Proposed Method	1.68	6.13	0.1	0.60

Table 2: The confusion matrix of recognition accuracies using the proposed method for each micro-expression on the CASME2 dataset.

Proposed Method	Ground truth				
	Happiness	Sadness	Surprise	Disgust	Repression
Happiness	0.80	0.02	0.06	0	0.19
Sadness	0.03	0.81	0.05	0.06	0.07
Surprise	0.06	0.03	0.73	0.11	0
Disgust	0.02	0.07	0.08	0.79	0.13
Repression	0.09	0.06	0.07	0.04	0.61

The experimental results are illustrated in Table 1. As can be seen, SE and MAE in our work are smaller than other methods errors on both of the CASME and CASME2 datasets.

The confusion matrix of recognition accuracies using the proposed method for each micro-expression on the CASME2 database is shown in Table 2. The experimental results on two used datasets are illustrated in Table 3 and Table 4. As can be seen, the ME recognition rate is increased in our work compared to the previous methods. In fact, our method can extract more meaningful information by using the apex frame and spatiotemporal diagonal planes.

4. CONCLUSION

The ME could display the true affective states and human goal since it cannot be controlled through one's willpower. However, it could only be effective, when the feature descriptor is able to elicit all tiny motions. For this reason, in this paper, we proposed a new method, which could extract tiny facial movements. We magnified the ME undetectable and subtle motions to visible. Also, we incorporated optical flow and LBP from diagonal planes on three frames especially the apex frame for micro-expression recognition. According to our numerical experiments, the proposed method has superior performance in the AFD and the ME recognition in comparison to the related

Table 3: Comparison of different methods on CASME2 dataset.

Method	ME recognition rate (%)
LOFBP-TOP [2]	52.44
LBP-TOP [13]	55.87
Wang's work [19]	68.75
Our proposed method	79.60

Table 4: Comparison of different methods on CASME dataset.

Method	ME recognition rate (%)
HOOF [3, 10]	57.87
LBP-TOP	64.67
Fuzzy Histogram of Optical Flow Orientations (FHOFO) [10]	65.99
Our proposed method	72.38

existing methods. In the future, this method could be exploited to detect tiny variations in other applications.

References

[1] C.-C. Chang, C.-J. J. A. t. o. i. s. Lin, and technology, LIBSVM: A library for support vector machines, ACM transactions on intelligent systems and technology (TIST), vol. 2, no. 3, pp. 1-27, 2011.

[2] T. Chang, F. Long, and J. Huang, Micro-expression recognition using optical flow and local binary patterns on three orthogonal planes, in Proceedings of the Seventh International Symposium of Chinese CHI, 2019, pp. 44-48.

[3] R. Chaudhry, A. Ravichandran, G. Hager, R. Vidal, editors, Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions, Computer Vision and Pattern Recognition, 2009 CVPR 2009 IEEE Conference on; 2009: IEEE.

[4] P. Ekman and W. V. J. P. Friesen, Nonverbal leakage and clues to deception, Journal for the Study of Interpersonal Processes, DTIC Document, Tech. Rep., Psychiatry, vol. 32, no. 1, pp. 88-106, 1969.

[5] V. Esmaeili, S.O. Shahdi, Automatic micro-expression apex spotting using Cubic-LBP, Multimedia Tools and Applications, 2020.

[6] V. Esmaeili, M. Mohassel Feghhi, S.O. Shahdi, Automatic Micro-Expression Apex Frame Spotting using Local Binary Pattern from Six Intersection Planes, accepted at 2020 International Conference on Machine Vision and Image Processing (MVIP); Faculty of Engineering, College of Farabi, University of Tahran, Iran, 19 & 20 Feb. 2020; 2020.

[7] R. Garg, A. Roussos, and L. J. I. j. o. c. v. Agapito, A variational approach to video registration with subspace constraints, International journal of computer vision, vol. 104, no. 3, pp. 286-314, 2013.

[8] KM. Goh, Ng CH, Lim LL, Sheikh UJTVC, Micro-expression recognition: an updated review of current trends, challenges and solutions, The Visual Computer, 2020,36(3):445-68.

[9] E. A. Haggard and K. S. Isaacs, Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy, in Methods of research in psychotherapy: Springer, pp. 154-165, 1966.

[10] S. Happy and A. J. I. T. o. A. C. Routray, Fuzzy histogram of optical flow orientations for micro-expression recognition, IEEE Transactions on Affective Computing, vol. 10, no. 3, pp. 394-406, 2017.

[11] A. C. Le Ngo, A. Johnston, R. C.-W. Phan, and J. See, Micro-expression motion magnification: Global Lagrangian vs. local Eulerian approaches, in 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG), pp. 650-656, 2018.

[12] 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: IEEE.

[13] X. Li et al., Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods, IEEE Transactions on Affective Computing, vol. 9, no. 4, pp. 563-577, 2017.

[14] S.-T. Liong, J. See, K. Wong, and R. C.-W. Phan, Less is more: Micro-expression recognition from video using apex frame, Signal Processing: Image Communication, vol. 62, pp. 82-92, 2018.

[15] Y. Liu, H. Du, L. Zheng, and T. Gedeon, A neural micro-expression recognizer, in 2019 14th IEEE international conference on automatic face & gesture recognition (FG 2019), 2019, pp. 1-4: IEEE.

[16] 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), pp. 281-286, 2017.

[17] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen, Recognising spontaneous facial micro-expressions, in IEEE International Conference on Computer Vision (ICCV), pp. 1449-1456, 2011.

[18] J. A. Ruiz-Hernandez and M. Pietikäinen, Encoding local binary patterns using the re-parametrization of the

Autonomous Apex Detection and Micro-Expression Recognition using...
Volume 11, Special Issue, Winter and spring 2020, 483-497

497

second order gaussian jet, in 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), pp. 1-6, 2013.

[19] L. Wang, H. Xiao, S. Luo, J. Zhang, and X. J. S. P. I. C. Liu, A weighted feature extraction method based on temporal accumulation of optical flow for micro-expression recognition, Signal Processing: Image Communication, vol. 78, pp. 246-253, 2019.

[20] 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 Asian conference on computer vision, pp. 525-537, 2014.

[21] H.-Y. Wu et al., Eulerian video magnification for revealing subtle changes in the world, ACM Transactions on Graphics, 31(4): 65, 2012.

[22] W.-J. Yan, Q. Wu, Y.-J. Liu, S.-J. Wang, and X. Fu, CASME database: a dataset of spontaneous micro-expressions collected from neutralized faces, in 10th IEEE international conference and workshops on automatic face and gesture recognition (FG), pp. 1-7, 2013.

[23] W-J. Yan, X. Li, S-J. Wang, G. Zhao, Y-J. Liu, Y-H. Chen, et al., CASME II: An improved spontaneous micro-expression database and the baseline evaluation, PloS one, 2014;9(1):e86041.

[24] G. Zhao, M. J. I. T. o. P. A. Pietikainen, and M. Intelligence, Dynamic texture recognition using local binary patterns with an application to facial expressions, IEEE Transactions on Pattern Analysis and Machine Intelligence, no. 6, pp. 915-928, 2007.