Adaptive Mask for Region-based Facial Micro-Expression Recognition

Walied Merghani[1] and Moi Hoon Yap[2]
[1] Computer Science and Information Technology, Sudan University of Science and Technology, Khartoum, Sudan
[2] Department of Computing and Mathematics, Manchester Metropolitan University, Manchester, United Kingdom

***Abstract*— Facial micro-expression can be characterized by its short duration and subtle movements. In facial micro-expression recognition, these subtle movements require more specific feature descriptors due to only a few parts of the face produce information that helps us to recognize micro-expressions. Over the past decade, researchers designed different Region of Interests (ROIs) to study specific face regions in micro-expressions recognition. To further study this aspect, we proposed a region-based method with an adaptive mask for facial micro-expression recognition. Based on the most frequent Action Units on the two publicly available datasets, i.e. CASME II and SAMM, 14 ROIs are defined where the adaptive mask is created by calculating the optical flow after Gaussian smoothing. Further, LBP-TOP features are extracted from each ROIs and Sequential Minimal Optimization is used to classify the micro-expressions. When evaluating our proposed method on CASME II and SAMM, we achieved the accuracy of 68.2% and 56.1%. In terms of F1-Score, our proposed method achieved 0.57 on CASMEII and the best performance of 0.50 on SAMM.**

I. INTRODUCTION

Facial micro-expression (FME) has become an active research area in recent years. Micro-expressions occur when a person attempts to conceal their true emotion [1], [2]. Upon the realisation that a facial expression is occurring, a person may try to suppress the facial expression because showing the emotion may not be appropriate or seen as a negative due to a cultural display rule [3]. During the suppression of a facial expression, a person may mask over the original expression and cause a micro-facial expression to appear in its place. During a scenario where the risks or ‘stakes’are high, micro-expressions have a higher tendency to appear based on the real emotion the person was feeling. The importance of this area comes from the benefits that can be use for a variety of disciplines including for instance security, interrogation and psychiatry [4], [5], [6]. Previous work on the micro-expressions recognition using face regions in data analysis have provided promising results on successful detection techniques, however there is room for improvement.

This paper introduces a new FME recognition algorithm by creating a new adaptive mask based on optical flow and defined new 14 ROIs based on the most frequently Action Units (AUs) related to FME occurs which has been found after a comprehensive analysis on FME datasets.

This work was supported by Manchester Metropolitan University and Sudan University of Science and Technology PhD studentship.

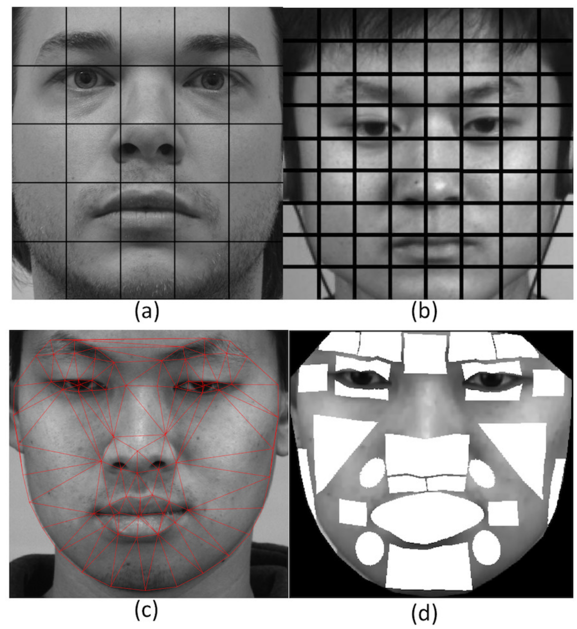


Fig. 1. Illustration of face regions: (a) 5×5 blocks, (b) 8×8 blocks, (c) Delaunay triangulation, and (d) FACS-based regions.

II. RELATED WORK

The state of the art can be categorized into:

- *Four quadrants*. Shreve et al. [7] split the face into 4 quadrants and analyse each quarter as individual temporal sequences. The advantage of this method is that it is simple to analyse larger regions, however the information to retrieve from the areas are restricted to whether there was some form of movement in a more global area.
- $m \times n$ *blocks*. Another method is to split the face into a specific number of blocks [8], [9], [10]. The movement on the face is analysed locally, rather than a global representation of the whole face, and can focus on small changes in very specific temporal blocks. A disadvantage to this method is that it is computationally expensive to process the whole images as $m \times n$ blocks. It can also include features around the edge of the face, including hair, that do not relate to movement but could still effect the final feature vector. Merghani et al. [11] try to enhance $m \times n$ *blocks* by removing

some of unnecessary blocks and evaluated their method using objective classes [12]. Figure 1(a) and Figure 1(b) illustrate the samples of block-based face regions.

- *Delaunay triangulation*. Delaunay triangulation, as shown if Figure 1(c), has also been used to form regions on just the face and can exclude hair and neck [13], however this approach can still extract areas of the face that would not be useful as a feature and adds further computational expense.
- *FACS-based region*. A more recent and less researched approach is to use defined ROIs to correspond with one or more FACS AUs [14], [15]. These regions have more focus on local parts of the face that move due to muscle activation. Some examples of ROI selection for micro-expression recognition and detection include discriminative response map fitting [16], Delaunay triangulation [13] and facial landmark based region selection [17]. Unfortunately, currently defined regions do not cover all AUs and miss some potentially important movements such as AU5 (Upper Lid Raiser), AU23 (Lip Tightener) and AU31 (Jaw Clencher). To overcome the problem, Davison et al. [18] proposed FACS-based regions to improve local feature representation by disregarding face region that do not contribute to facial muscle movements. The defined region is presented in Figure 1(d).

Figure 1 compares different face region splitting methods. Due to FACS-based region is more relevant to facial muscle movements and suitable for AUs detection, more research should be focusing on FACS-based region than split the face into $m \times n$ blocks.

III. Methodology

Figure 2 describes the methodology of the proposed method, each part in the figure will be described into more details in the following sections.

A. Facial Landmarks Detection

Faces have quite distinct features, such as eyebrows, mouths, noses and eyes. Most humans have these features in about the same position, and so identifying the points where these features occur on the face is an interest research problem. Zhou et al. [19] proposed a way of detecting facial points using a deep learning approach named convolutional neural networks (CNN). The research toolkit developed requires an Internet connection to allow for the images to be processed on the Face++ servers. We use Face++ to detect the facial landmarks. Figure 3(a) shows the facial landmark points.

B. FACS-Based Regions

Analysis have been done on two of the most popular publicly available micro-expressions datasets, CASMEII [8] and SAMM [20], to find the most frequently occur AUs in micro-expressions. Table I summarizes the frequency of the AUs with the highest occurrence on AU4 and the lowest occurrence on AU31. This step is to ensure only relevant movements are detected. Additionally, the advantage of this step is that the features can be locally analyzed without processing insignificant parts of the face. Table II summarises the name of each region and its correspondence AUs. ROIs have located on face after detect facial landmarks points. Figure 3(b) illustrates our proposed ROIs.

TABLE I
A Summary of AUs frequency on CASME II and SAMM.

AU	CASME II	SAMM	Total
1	26	6	32
2	22	18	40
4	129	23	152
5	2	10	12
6	13	5	18
7	38	46	84
9	13	5	18
10	16	6	22
12	34	30	64
13	0	3	3
14	27	13	40
15	16	4	20
17	25	7	32
18	0	4	4
20	0	7	7
24	2	10	12
25	2	7	9
26	0	6	6
31	0	2	2

TABLE II
The region number, name and its associated AUs.

Region Number	Region Name	Associated AU(s)
1	Right Brow - Right	2,4
2	Right Brow - Left	1,4
3	Left Brow - Right	1,4
4	Left Brow - Left	2,4
5	Right Eye	5,7
6	Glabella	1, 4, 9
7	Left Eye	5,7
8	Right Cheek	6, 12
9	Left Cheek	6, 12
10	Dimple - Right	12, 13, 14, 18, 20
11	Upper Lip	10
12	Dimple - Left	12, 13, 14, 18, 20
13	Mouth	12, 13, 14, 15, 16 17, 18, 19, 20, 22 23, 24, 25, 26, 28
14	Chin	15, 17, 25, 26

C. Smoothing

Due to the subtle nature of micro-expression and the noise could affect the extracted features, noise reduction should be applied. One of the simplest method in noise reduction is using the smoothing algorithm. For our work, Gaussian

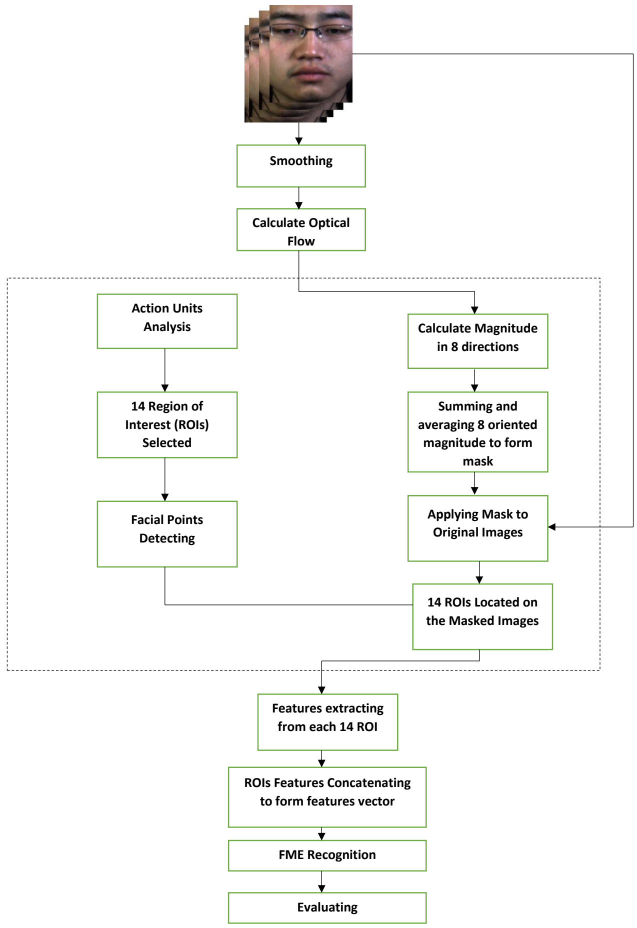
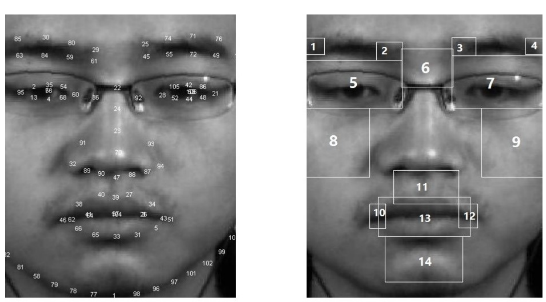


Fig. 2. Methodology of proposed method.



(a) (b)

Fig. 3. (a) Facial landmark points on a sample subject of CASMEII (b) 14 ROIs based on the frequency of AUs occurrences.

(a) (b)

Fig. 4. Preprocessing step: (a) Before smoothing (b) after smoothing.

smoothing operator have been applied to reduce the noise, as shown in Figure 4.

D. Optical-flow Mask

The optical flows are calculated from the frame sequences after smoothing to represent motion information. The optical flows are computed from each pair of neighbouring frames between the first frame and the rest, as illustrated in Figure

5. Then the magnitude is calculated in eight orientations, these eight oriented magnitude is then go through an averaging process as shown in Figure 6 to form the mask (in the centre of Figure 6). The mask was adaptive because every single frame in the sequence has its own mask due to the different oriented magnitude for each frame. This mask is then applied to the original image to produce a masked image before locating the ROIs, as shown in Figure 7.

Fig. 5. Illustration of the optical flow process.

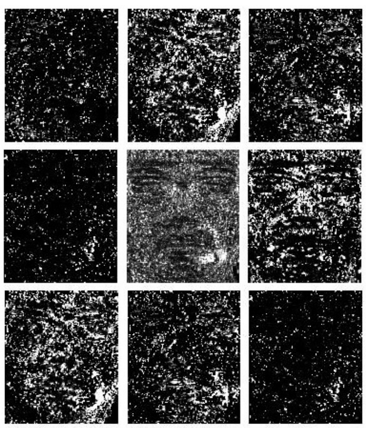


Fig. 6. Illustrations of the magnitude calculation of optical flow in 8 directions (surrounding images), the centre image shows the average of 8 surrounding images.

E. Feature Decriptors

LBP-TOP which was first introduced by Zhao et al. [21] has been used to extract the features from each of 14 ROI. LBP-TOP Based on the LBP [22] operator, LBP-TOP was first described as a texture descriptor that used XT and YT temporal planes rather than just the 2D XY spatial plane. To begin, an image, with c being the centre pixel and P being neighbouring pixels with a radius of R, has the LBP operator applied

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \tag{1}$$

where g_c is the grey value of the centre pixel and g_p is the grey value of the p-th neighbouring pixel around R. 2^p defines weights to neighbouring pixel locations and is used

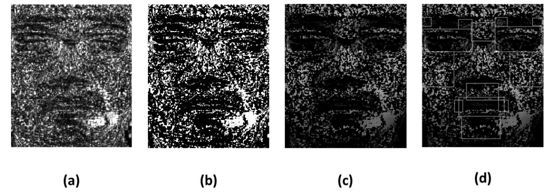


Fig. 7. Optical flow mask: (a) Oriented magnitude (b) thresholding (c) applying mask to original image (d) regions after mask.

to obtain the decimal value, where the binary pattern name derives. The sign function to determine what binary value is assigned to the pattern and is calculated as

$$s(\mathbf{A}) = \begin{cases} 1, & \text{if } \mathbf{A} \geq 0 \\ 0, & \text{if } \mathbf{A} < 0 \end{cases} \tag{2}$$

If the grey value of P is larger than or equal to c, then the binary value is 1, otherwise it will be 0. The full LBP-TOP histogram can be defined as

$$H_{i,j} = \sum_{x,y,t} I\{LBP_j(x,y,t) = i\}, i = 0, \ldots, n_j - 1 \tag{3}$$

where n_j is the number of labels produced by the LBP operator in the jth plane. $j = 0,1,2$ which represents the XY, XT and YT planes respectively. $LBP_i(x,y,t)$ expresses the LBP code of the central pixel (x,y,t) in the jth plane.

F. Classification and Validation

Sequential Minimal Optimization (SMO) [37] has been used for classification. SMO is able to break down large quadratic programming problems into a series of the smallest possible problem, which are solved analytically and avoids using a time-consuming numerical quadratic programming optimisation as an inner loop. SMO is also able to handle large training sets and is one of the computationally fastest methods of evaluating linear SVMs. Leave-one-subject-out (LOSO) has been used to validate the proposed method, where LOSO is well-established and widely used in FME evaluation.

IV. Results and Discussion

Table III shows the result achieved by the proposed method against the state of the art. Our proposed method performs better than the majority of the handcrafted methods and it is comparable to some of the deep learning method. Amongst handcrafted methods, we observe FMBH [31] has better result in terms of accuracy when they evaluated their method on CASME II. However, a manually created mask was used to separate the background from the face, which makes it difficult to use the method in automatic systems, unlike the our proposed method, which is all automatic. In

TABLE III

COMPARISON BETWEEN PROPOSED METHOD WITH THE STATE-OF-THE-ART METHODS ON CASME II AND SAMM.

	Method	CASME II		SAMM	
		Accuracy	F1-Score	Accuracy	F1-Score
Hand-Crafted	LBP-TOP (baseline) [23]	63.4	0.33	41.38	-
	LBP-MOP [24]	66.8	-	42.72	-
	HOOF [25]	44	-	46.13	-
	FDM [26]	45.3	0.47	-	-
	STCLQP [27]	58.39	0.57	-	-
	STLBP-IP [28]	62.75	-	-	-
	Bi-WOOF [29]	62.2	50	-	-
	HIGO[30]	67.20	-	-	-
	MDMO [16]	67.37	-	-	-
	FMBH [31]	69.11	-	-	-
	FHOFO [32]	56.64	-	-	
	Proposed Method	68.2	0.57	56.1	**0.5**
Deep Learning	AlexNet (baseline)	62.96	0.66	52.94	0.42
	SSSN [33]	**71.19**	**0.71**	**56.62**	0.45
	CNN+LSTM [34]	60.98	-	-	-
	OF+CNN [35]	56.94	-	-	-
	ELRCN [36]	52.44	0.55	-	-

addition to that they did not evaluate their method with F1-Score. When compared with other methods in hand-crafted, we achieved the best F1-Score.

In deep learning methods, obviously SSSN [33] achieved the best results on CASME II. However, our proposed method achieved higher F1-score on SAMM. Comparing to region-based method such as MDMO [16] and FHOFO [32], we found that our proposed method performed better.

To justify the importance of the operations used in the proposed method, we conduct ablation studies. Table IV compares the effect of smoothing, adaptive mask and number of ROIs on FME recognition. We observed that extracting features locally is better than globally. In addition, we found that the use of adaptive mask has improved the results, as well as the use of smoothing before creating the mask, which has a positive effect on the result as it removes some of the noise (that can be confused with some of the micro-movements). It also shows a sample of experiment using different ROIs (8 ROIs) by combining some ROIs like AU1, AU2, AU3, and AU4, and removing some like AU11 and AU13. The experiment proved that the selected 14 ROIs is a better choice for FME recognition.

TABLE IV

RESULT OF DIFFERENT EXPERIMENTS (S:SMOOTHING,$\tilde{S}$: NOT SMOOTHING, M:MASK, $\tilde{M}$: NO MASK) WITH DIFFERENT NUMBER OF ROIS.

Method	Accuracy	F1-Score
Global HOOF	44	46.31
14 ROIs HOOF($\tilde{S}$,$\tilde{M}$)	56.8	0.23
14 ROIs HOOF(S,$\tilde{M}$)	58.9	0.4
Global LBP-TOP	63.4	0.33
Global LBP-TOP(S,M)	61.3	0.38
8 ROIs LBP-TOP(S,M)	62.2	0.56
14 ROIs LBP-TOP($\tilde{S}$,$\tilde{M}$)	63.4	0.45
14 ROIs LBP-TOP(S,$\tilde{M}$)	63.5	0.51
14 ROIs LBP-TOP($\tilde{S}$,M)	65.09	0.55
14 ROIs LBP-TOP(S,M) Proposed	**68.2**	**0.57**

V. CONCLUSION

This paper proposed a new method for FME recognition. The method is region-based, where 14 ROIs have selected based on AUs analysis on CASMEII and SAMM. ROIs have been proposed for locally analyzed the features to avoid unimportant information of the face. Further, to be more specific to the movement related to FME, adaptive mask based on the micro-motion using optical flow have applied to each frame of ME. LBP-TOP features extracted from each region and SMO is implemented as the classifier. The proposed method evaluated on two of benchmark datasets: CASME II and SAMM, and achieved a promising result. When compared to deep learning approach, it is comparable when evaluated with SAMM. Future work will consider the potential to combine the region-based approach with the deep learning approach.

REFERENCES

[1] P. Ekman, *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*. Norton, 2001.

[2] P. Ekman, "Lie catching and microexpressions," in *The Philosophy of Deception*, C. W. Martin, Ed. Oxford University Press, 2009, pp. 118–133.

[3] D. Matsumoto, S. H. Yoo, and S. Nakagawa, "Culture, emotion regulation, and adjustment." *Journal of personality and social psychology*, vol. 94, no. 6, p. 925, 2008.

[4] M. O'Sullivan, M. G. Frank, C. M. Hurley, and J. Tiwana, "Police lie detection accuracy: The effect of lie scenario." *Law and Human Behavior*, vol. 33, no. 6, p. 530, 2009.

[5] M. G. Frank, C. J. Maccario, and V. l. Govindaraju, "Behavior and security," in *Protecting airline passengers in the age of terrorism*. Greenwood Pub. Group, 2009.

[6] W. Merghani, A. K. Davison, and M. H. Yap, "A review on facial micro-expressions analysis: datasets, features and metrics," *arXiv preprint arXiv:1805.02397*, 2018.

[7] M. Shreve, J. Brizzi, S. Fefilatyev, T. Luguev, D. Goldgof, and S. Sarkar, "Automatic expression spotting in videos," *Image and Vision Computing*, vol. 32, no. 8, pp. 476 – 486, 2014.

[8] W.-J. Yan, X. Li, S.-J. Wang, G. Zhao, Y.-J. Liu, Y.-H. Chen, and X. Fu, "Casme ii: An improved spontaneous micro-expression database and the baseline evaluation," *PloS one*, vol. 9, no. 1, 2014.

[9] A. K. Davison, M. H. Yap, N. Costen, K. Tan, C. Lansley, and D. Leightley, "Micro-facial movements: An investigation on spatio-temporal descriptors," in *Computer Vision-ECCV 2014 Workshops*. Springer, 2014, pp. 111–123.

[10] A. K. Davison, M. H. Yap, and C. Lansley, "Micro-facial movement detection using individualised baselines and histogram-based descriptors," in *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1864–1869.

[11] W. Merghani, A. Davison, and M. Yap, "Facial micro-expressions grand challenge 2018: evaluating spatio-temporal features for classification of objective classes," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, pp. 662–666.

[12] A. K. Davison, W. Merghani, and M. H. Yap, "Objective classes for micro-facial expression recognition," *Journal of Imaging*, vol. 4, no. 10, p. 119, 2018.

[13] Z. Lu, Z. Luo, H. Zheng, J. Chen, and W. Li, "A delaunay-based temporal coding model for micro-expression recognition," in *Asian Conference on Computer Vision*. Springer, 2014, pp. 698–711.

[14] S.-J. Wang, W.-J. Yan, X. Li, G. Zhao, and X. Fu, "Micro-expression recognition using dynamic textures on tensor independent color space," in *Pattern Recognition (ICPR), 2014 22nd International Conference on*. IEEE, 2014, pp. 4678–4683.

[15] S.-J. Wang, W.-J. Yan, X. Li, G. Zhao, C.-G. Zhou, X. Fu, M. Yang, and J. Tao, "Micro-expression recognition using color spaces," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 6034–6047, 2015.

[16] Y.-J. Liu, J.-K. Zhang, W.-J. Yan, S.-J. Wang, G. Zhao, and X. Fu, "A main directional mean optical flow feature for spontaneous micro-expression recognition," *IEEE Transaction of Affective Computing*, 2015.

[17] D. Patel, G. Zhao, and M. Pietikäinen, "Spatiotemporal integration of optical flow vectors for micro-expression detection," in *Advanced Concepts for Intelligent Vision Systems*. Springer, 2015, pp. 369–380.

[18] A. Davison, W. Merghani, C. Lansley, C.-C. Ng, and M. H. Yap, "Objective micro-facial movement detection using facs-based regions and baseline evaluation," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, pp. 642–649.

[19] E. Zhou, H. Fan, Z. Cao, Y. Jiang, and Q. Yin, "Extensive facial landmark localization with coarse-to-fine convolutional network cascade," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2013, pp. 386–391.

[20] A. K. Davison, C. Lansley, N. Costen, K. Tan, and M. H. Yap, "Samm: A spontaneous micro-facial movement dataset," *IEEE Transactions on Affective Computing*, vol. 9, no. 1, pp. 116–129, Jan 2018.

[21] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 6, pp. 915–928, 2007.

[22] T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51 – 59, 1996.

[23] 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 *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*. IEEE, 2013, pp. 1–7.

[24] Y. Wang, J. See, R. C.-W. Phan, and Y.-H. Oh, "Efficient spatio-temporal local binary patterns for spontaneous facial micro-expression recognition," *PloS one*, vol. 10, no. 5, p. e0124674, 2015.

[25] R. Chaudhry, A. Ravichandran, G. Hager, and R. Vidal, "Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 1932–1939.

[26] F. Xu, J. Zhang, and J. Z. Wang, "Microexpression identification and categorization using a facial dynamics map," *IEEE Transactions on Affective Computing*, vol. 8, no. 2, pp. 254–267, 2017.

[27] 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.

[28] X. Huang, S.-J. Wang, G. Zhao, and M. Piteikainen, "Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2015, pp. 1–9.

[29] 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.

[30] X. Li, X. Hong, A. Moilanen, X. Huang, T. Pfister, G. Zhao, and M. Pietikäinen, "Reading hidden emotions: spontaneous micro-expression spotting and recognition," *arXiv preprint arXiv:1511.00423*, 2015.

[31] H. Lu, K. Kpalma, and J. Ronsin, "Motion descriptors for micro-expression recognition," *Signal Processing: Image Communication*, vol. 67, pp. 108–117, 2018.

[32] S. Happy and A. Routray, "Fuzzy histogram of optical flow orientations for micro-expression recognition," *IEEE Transactions on Affective Computing*, 2017.

[33] H.-Q. Khor, J. See, S.-T. Liong, R. C. Phan, and W. Lin, "Dual-stream shallow networks for facial micro-expression recognition," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 36–40.

[34] 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 2016 ACM on Multimedia Conference*. ACM, 2016, pp. 382–386.

[35] J. Li, Y. Wang, J. See, and W. Liu, "Micro-expression recognition based on 3d flow convolutional neural network," *Pattern Analysis and Applications*, vol. 22, no. 4, pp. 1331–1339, 2019.

[36] H.-Q. Khor, J. See, R. C. W. Phan, and W. Lin, "Enriched long-term recurrent convolutional network for facial micro-expression recognition," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018, pp. 667–674.

[37] J. Platt *et al.*, "Fast training of support vector machines using sequential minimal optimization," 1999.