

# A Boost in Revealing Subtle Facial Expressions: A Consolidated Eulerian Framework

Wei Peng<sup>1+</sup>, Xiaopeng Hong<sup>2,1+</sup>, Yingyue Xu<sup>1</sup>, and Guoying Zhao<sup>1§</sup>

<sup>1</sup> Center for Machine Vision and Signal Processing, University of Oulu, Finland

<sup>2</sup> Xian Jiaotong University, Xian, P. R. China

**Abstract**—Facial Micro-expression Recognition (MER) distinguishes the underlying emotional states of spontaneous subtle facial expressions. Automatic MER is challenging because that the intensity of subtle facial muscle movement is extremely low and the duration of ME is transient.

Recent works adopt motion magnification or temporal interpolation to resolve these issues. Nevertheless, existing works divide them into two separate modules due to their non-linearity. Though such operation eases the difficulty in implementation, it ignores their underlying connections and thus results in inevitable losses in both accuracy and speed. Instead, in this paper, we propose a consolidated Eulerian framework to reveal the subtle facial movements. It expands the temporal duration and amplifies the muscle movements in micro-expressions simultaneously. Compared to existing approaches, the proposed method can not only process ME clips more efficiently but also make subtle ME movements more distinguishable. Experiments on two public MER databases indicate that our model outperforms the state-of-the-art in both speed and accuracy.

## I. INTRODUCTION

Micro-expression recognition (MER) is an emerging facial affective analysis task to distinguish the subtle spontaneous facial expressions. Automatic MER has many potential applications like clinical diagnosis, lie detection, and human computer interaction (HCI). MER is challenging since people tend to hide their emotions in case of being found. Studies of ME in psychology indicate that even a trained human can only achieve a MER accuracy lower than 50% [5].

Promising progresses have been made over the past decade, in aspects such as dataset construction [15], [34], [4], spatial-temporal descriptors either by hand-crafting [3], [36], [14], [39], [9], [33], [7], [40], or deep learning [23], [26], [16], [31], [32], [8], [29], etc. However, it is still far away from satisfaction. The difficulties are two-fold: 1). The ME related muscle movements are remarkably subtle. ME, unlike general facial expressions, only covers a small area of face so that the movements are too imperceptible to be detected. 2). ME duration is quite short. Generally, ME only lasts approximately from 1/25 to 1/3 seconds [25].

Recent works tend to employ motion magnification (MAG) or temporal interpolation as pre-processing steps for MER. MAG [19], [30], [35] amplifies the intensity of motion and makes subtle motion much easier to observe. Li *et al.* [14] amplified the subtle facial movements by employing

MAG before extracting features and improved the accuracy by around 10% on the CASMEII [34] database. Obvious improvements are also observed in [21], [20], [22] as a result of introducing MAG. Besides MAG, temporal interpolation also attracts increasing attention. Pfister *et al.* [24] proposed to use Time Interpolation Model (TIM) [38], [37] for MER problem, which allows to obtain sufficient frames for feature extraction even for very short expressions. Methods [15], [34], [13], [18], [14], [12], [10], [11] also follow a similar mechanism and get promising results.

Despite these progresses, MAG and TIM are usually treated as two separate modules [34], [11], [14]. Such assumption eases the difficulty in implementation. However, there are several problems. Firstly, it is prone to having the diluted problem<sup>1</sup> [13] into the succeeding process. Secondly, the calculation of intermediate products increases the computational costs. Finally, what makes it worse, ignoring the underlying connections will inevitably result in the loss of recognition accuracy.

To address these issues, we explore the underlying joint formulations for MAG and TIM and propose a consolidated framework for revealing the subtle spatial-temporal information in ME clips. The proposed model, which is called ME-Booster, can accomplish temporal interpolating and facial muscle motion magnification simultaneously. Compared to the traditional MER systems, our ME-Booster not only avoids unnecessary separation of the two modules but also eliminates the side effect brought by the intermediate process. We investigate our framework on two popular spontaneous MER databases: SMIC [15] and CASMEII [34]. Experimental results indicate that our model is computationally economic and substantially outperforms the state-of-the-art approaches.

The contributions of this paper are two-fold: 1). We propose a consolidated video motion revealing model for MER, where TIM and MAG are, for the first time, jointly formulated; 2). We provide fast implementations of MAG and TIM methods by exploring their underlying linear forms. It is thus computationally efficient.

## II. METHODOLOGY

In this section, we describe a consolidated Eulerian subtle motion revealing model, termed ME-Booster, to boost micro-expression recognition. Suppose there is a ME video clip  $V = (I(1), I(2), \dots, I(T))$ , where  $I(t) \in \mathbb{R}^d$  denotes the

<sup>+</sup> Authors contributed equally to this work    <sup>§</sup> Corresponding author

<sup>1</sup>A phenomenon that the facial muscle movements fall into decline.

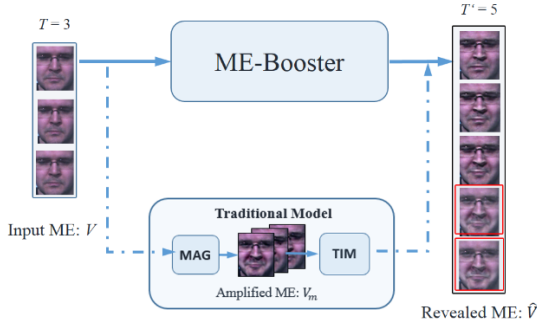


Fig. 1: Overview of the proposed ME-Booster. The original three-frame input can hardly be distinguished, while ME-Booster (top) expands it to five frames and makes it much easier to recognize by eyes. The traditional model (bottom) treats MAG and TIM as two parts. So there is an intermediate product: the amplified ME without temporal extension. On the contrary, our ME-Booster model treats these processes as an entirety.

$t$ -th frame of the video for  $t = 1, \dots, T$ . Our goal is to produce a video  $\hat{V}$  where the subtle facial movements can be easily observed.

The model, in its general form, is formulated by a video-specific function  $f$  with a set of parameters  $\Theta$  as follows

$$\hat{V} = f(V; \Theta), \quad (1)$$

where the function  $f: \mathbb{R}^{d \times T} \mapsto \mathbb{R}^{d \times T'}$  maps the input video to a fixed length ( $T'$ ) one, where the subtle facial muscle movements in  $V$  are revealed.

Considering the non-linear nature of the representative motion magnification and temporal interpolation methods, such as the Eulerian based MAG method [30] and the graph spectral TIM method [2], existing MER approaches [34], [11], [14] usually treat MAG and TIM as two separate modules and adopt them in a successive manner. On the contrary, in this paper we will show that under certain but reasonable circumstances, MAG and TIM can be embedded in Eq. (1) by a scale-wise linear function. Firstly, the Eulerian MAG model [19], [30], [35] is multi-scale and non-linear, and thus a complicated system. However, we realize that it is feasible to implement it as a series of scale-wise linear functions, since in each of the scales it approximates movements using the Taylor expansion. Secondly, we will show that the interpolated video can be obtained without explicitly solving the generalized eigenvalue problem as the TIM method [38] does. Thus, w.l.o.g., in the following context, we formulate our model with a single-scale setting and consider the  $f(\cdot)$  as a linear function<sup>2</sup>

$$\hat{V} = V\mathbf{W}. \quad (2)$$

To obtain the parameter matrix  $\mathbf{W} \in \mathbb{R}^{T \times T'}$ , we decouple it as follow

$$\mathbf{W} = \mathbf{W}^M \mathbf{W}^I, \quad (3)$$

in which  $\mathbf{W}^M \in \mathbb{R}^{T \times T}$  and  $\mathbf{W}^I \in \mathbb{R}^{T \times T'}$  are the magnification matrix and the interpolation matrix, respectively. The

<sup>2</sup>Multi-scale analysis can be easily reached by applying the model to each single scale.

order of the right hand side of Eq. (3) is consistent with the one in [14].

In the following paragraphs, we will derive the solutions of  $\mathbf{W}^M$  and  $\mathbf{W}^I$  respectively in detail and show that both  $\mathbf{W}^M$  and  $\mathbf{W}^I$  can be determined once the lengths of the input and output video are given.

### Solution to the magnification matrix $\mathbf{W}^M$

In this subsection, we aim to get the solution for  $\mathbf{W}^M$ . Let  $W_{i,j}^M$  be the entry of  $\mathbf{W}^M$  at  $(i, j)$  for  $i, j = 1, \dots, T$ . The magnified frame  $I_m(t)$  can be described as

$$I_m(t) = \sum_{i=1}^T I(i) W_{i,t}^M. \quad (4)$$

To get the solution of  $\mathbf{W}^M$ , we revisit the Eulerian magnification model [30]. It regards the first frame in a video as a 1D digital signal  $s(x)$ , and every other frame as a displacement of this signal, i.e.,  $I(x, t) = s(x + \delta(t))$ , where  $I(x, t)$  is the pixel intensity at location  $x$  of the  $t$ -th frame and  $\delta(t)$  is a tiny displacement function. It is then approximated by the first-order Taylor expansion

$$s(x + \delta(t)) = s(x) + \delta(t) \frac{\partial s}{\partial x}, \quad (5)$$

where every frame in the ME video is an addition of the base signal and the multiplication of the displacement and the first derivative. Here,  $\frac{\partial s}{\partial x} = I(t) - I(t-1)$  represents the ME motion. It is fulfilled using a recursive algorithm to take all the preceding frames into account and get the ‘historical’ motion  $\mathbf{B}(t)$

$$\mathbf{B}(t) = L_1(t) - L_2(t). \quad (6)$$

$L_1(t)$  and  $L_2(t)$ , for  $t = 2, \dots, T$  are auxiliary variables

$$\begin{cases} L_1(t) = w_1 I(t) + (1 - w_1) L_1(t-1) \\ L_2(t) = w_2 I(t) + (1 - w_2) L_2(t-1), \end{cases} \quad (7)$$

where the  $L_1(1) = I(1)$ ,  $L_2(1) = I(1)$ . The hyper-parameters  $w_1$  and  $w_2$  are in the range  $(0, 1)$  and  $w_1 > w_2$ . Then applying a magnification factor  $\alpha$  to  $\mathbf{B}(t)$ , we get the magnified frame  $I_m(t)$

$$I_m(t) = I(t) + \alpha \mathbf{B}(t). \quad (8)$$

The Eulerian magnification model in Eq. (8) is clearly a linear model. We solve out a parameter matrix operated on the input video by expanding the recursive form in Eq. (7). As a result, by combining Eq. (4) and Eq. (8), we get the solution for  $\mathbf{W}^M$

$$W_{i,j}^M = \begin{cases} \alpha(1 - w_1)^a w_1^b - \alpha(1 - w_2)^a w_2^b & \text{if } j > i \\ \alpha(w_1 - w_2) + 1 & \text{if } j = i \\ 0 & \text{else,} \end{cases} \quad (9)$$

where  $a = j - i$  and  $b = \min(1, i - 1)$ . As it is in Eq. (9),  $\mathbf{W}^M$  is an upper triangular matrix since ME motion is only related to the preceding video frames. As the parameters are clearly independent of the input video, we thus compute  $\mathbf{W}^M$  in advance and store it by a look-up table (LUT) to further improve the computational efficiency. To get a better

magnification result, one can employ  $\mathbf{W}^M$  on the multi-scale input video, where for different scales, the factor  $\alpha$  in Eq. (9) may be truncated.

### Solution to the interpolation matrix $\mathbf{W}^I$

In this subsection, we will derive  $\mathbf{W}^I$  to form our efficient temporal interpolation module for extending  $V_m$  to  $T'$  frames. Just for clarification, we introduce the symbols  $V_m = (I_m(t))$  for  $t = 1, 2, \dots, T$  to describe the intermediate magnification output, though they are finally eliminated by Eq. (3) in practice. The recovered video  $\hat{V}$  is

$$\hat{V} = V_m \mathbf{W}^I. \quad (10)$$

Applying a graph spectral method [2], the traditional TIM approach [38] finds projection functions by which the frame adjacency priority is kept in the low latent space. That is

$$\mathcal{F}_{map}(I_m(t)) = (\mathbf{M}^{-1} \mathcal{V}^\top \mathbf{U}^\top)(I_m(t) - \bar{I}_m), \quad (11)$$

where the function  $\mathcal{F}_{map} : \mathbb{R}^d \mapsto \mathbb{R}^{T-1}$  describes a projection from an image to  $T - 1$  points. Here,  $\mathbf{U}$  is an unitary matrix from SVD [6] on input,  $\mathbf{M}$  and  $\mathcal{V}$  are graph Laplacian related diagonal matrix and eigenvectors, respectively.  $\bar{I}_m$  is the average frame of all the frames in video  $V_m$ . Readers may refer to [38] for more details.

In this way, every frame becomes a group of discrete points in a latent space. According to [2], the optimal projection is a group of sine functions. Let  $\mathcal{F}(t) = \mathcal{F}_{map}(I_m(t))$ . By sampling from the latent space and projecting the points back to image space, the input is reconstructed. According to Eq. (11), the intermediate magnification video can be reconstructed by

$$V_m = \mathbf{A} \mathbf{Y} + \bar{V}_m, \quad (12)$$

where  $\bar{V}_m = V_m(\frac{1}{T} \mathbf{1}_{T \times T})$ ,  $\mathbf{Y} = [\mathcal{F}(\frac{1}{T}), \mathcal{F}(\frac{2}{T}), \dots, \mathcal{F}(1)]$  are  $T$  groups of points in latent space. In [38],  $\mathbf{A} = \mathbf{U}(\mathcal{V}^{-1})^\top \mathbf{M}$ . However, one can avoid solving generalized problem since  $\mathbf{Y}$  is known. We adopt the right inverse matrix to compute  $\mathbf{A}$

$$\mathbf{A} = V_m(\mathbf{I} - \frac{1}{T} \mathbf{1}_{T \times T}) \mathbf{Y}^\top (\mathbf{Y} \mathbf{Y}^\top)^{-1}. \quad (13)$$

So sampling  $T'$  groups of points, say  $\mathbf{Y}' = [\mathcal{F}(\frac{1}{T'}), \mathcal{F}(\frac{2}{T'}), \dots, \mathcal{F}(1)]$ , the video is extended to  $T'$  frames. That is

$$\hat{V} = \mathbf{A} \mathbf{Y}' + \bar{V}'_m, \quad (14)$$

where  $\bar{V}'_m(t) = V_m(\frac{1}{T'} \mathbf{1}_{T' \times T'})$ . Since  $\mathbf{Y}'$  contains groups of discrete points, it satisfies the requirement for a linear TIM model. Substituting Eq. (13) into Eq. (14), we obtain the solution for  $\mathbf{W}^I$ , that is

$$\mathbf{W}^I = (\mathbf{I} - \frac{1}{T'} \mathbf{1}_{T' \times T'}) \mathbf{Y}'^\top (\mathbf{Y} \mathbf{Y}^\top)^{-1} \mathbf{Y}' + \frac{1}{T'} \mathbf{1}_{T' \times T'}. \quad (15)$$

Here Eq. (15) provides a linear solution to  $\mathbf{W}^I$  and it is independent to the input video, which means  $\mathbf{W}^I$  also can be computed in advance.

After obtaining  $\mathbf{W}^M$  and  $\mathbf{W}^I$ , we employ Eq. (3) to solve the matrix  $\mathbf{W}$ , which is the only parameter matrix used in

practice, and a consolidated model is constructed. Obviously, ME-booster is a unique model for both video magnification and time interpolation. The relation to the magnification and interpolation is clear. When  $\mathbf{W}^I$  is an identity matrix, the model formulated in Eq. (2) is a MAG model; When  $\mathbf{W}^M$  is an identity matrix, it degenerates to a TIM model.

Our ME-booster model is computationally economic and can make the subtle ME more distinguished. These advantages are mainly contributed by the discovery of the underlying common linear property between TIM and MAG modules. By this way, there is no need of intermediate process or complicated computation of graph Laplacian. Besides, expanding recursive forms and decoupling model parameters from the input also dedicate to these advantages.

## III. EXPERIMENTS

We evaluate the proposed model on two widely used spontaneous ME databases: SMIC [15] and CASMEII [34]. The SMIC database contains 306 ME video clips belonging to four sub-sets, which are SMIC-HS, SMIC-VIS, SMIC-NIR, and SMIC-subHS. There are three classes for all the ME clips : *Positive*, *Negative*, and *Surprise*. The CASMEII database contains 247 ME video clips. The baseline method of CASMEII chooses classes *Happiness*, *Surprise*, *Disgust*, *Repression*, and *Others* for training and testing.

To extensively evaluate the proposed ME-Booster, we perform the following three experiments: A). Consolidation versus separate; B). Comparisons to the state of the arts; and C). Computational cost evaluation. We explore the different values for hyper-parameters  $T'$ ,  $\alpha$ , and  $w_1$  and  $w_2$ . Based on the results, which are found consistent to those in [14], [30], the parameters are empirically set to 10, 16, 0.4, and 0.05, respectively. For the experiments A) and B), a linear SVM [1] is used as classifier for our model, unless further specified. As suggested by [14], in all MER experiments, the Leave-One-Subject-Out (LOSO) protocol is adopted. The micro-average method [27] is employed as evaluation metrics.

### A. Consolidation versus separate

We compare ME-Booster with the current best method [14], which contains both the MAG and TIM modules as two separate ones, as illustrated as the 'traditional model' in Fig. 1. For fair comparison, we use the same classifier and three mainstream temporal-spatial feature descriptors: LBP-TOP[36], HOG-TOP[3] and HIGO-TOP [14], which are extensively evaluated in the state of the art [14].

The comparative results are reported in Fig. 2. There are several important observations. 1) For HIGO-TOP, the accuracy is improved by almost 10.0% on the SMIC-NIR. On the SMIC-subHS, our accuracy is 87.32% , while the current best is only 80.28%. 2) For the HOG-TOP, our model improves the accuracy by 3.0% on average. 3) For the LBP-TOP, ME-Booster is slightly better than the baseline on the SMIC, while it improves by almost 4.5% on the CASMEII. 4) The largest improvements in accuracy are almost 13%, 6%, and 4.5% for the HIGO-TOP, HOG-TOP, and LBP-TOP respectively. From the results, it is safe to conclude that our

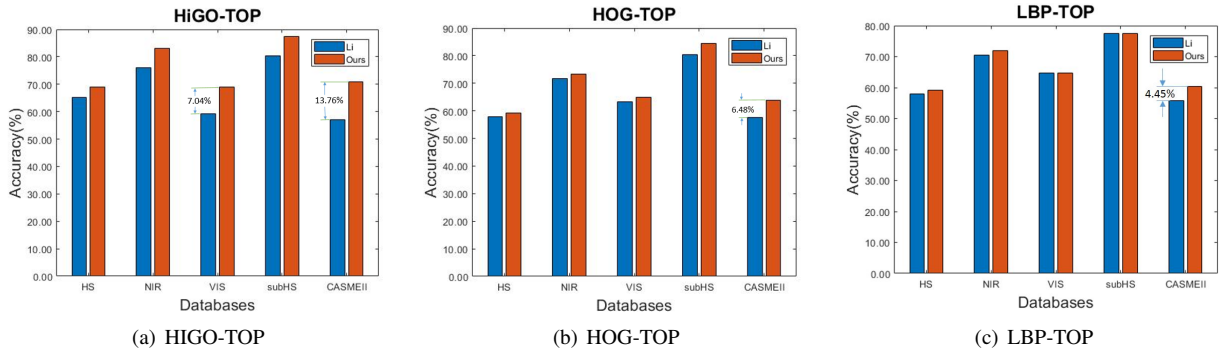


Fig. 2: MER performance comparison using three mainstream feature descriptors: (a) HIGO-TOP, (b) HOG-TOP, and (c) LBP-TOP.

TABLE I: Experimental results of MER comparing to recent state-of-the-art approaches on the SMIC and the CASMEII databases.

	SMIC-HS	CASMEII
Baseline-SMIC [15]	48.8%	-
Baseline-CASMEII [34]	-	63.41%
Adaptive Mag [22]	-	51.9%
Wang <i>et al.</i> [28]	-	62.3%
Liong <i>et al.</i> [17]	53.56%	-
Huang <i>et al.</i> [11]	60.98%	64.37%
DiSTLBP-RIP [10]	63.41%	64.78%
Deep Selective [23]	53.60%	47.30%
Deep Apex [16]	-	60.64%
Li <i>et al.</i> [14]	68.29%	67.21%
<b>Ours</b>	<b>68.90%</b>	<b>70.85%</b>

TABLE II: Experimental results of MER on sub-datasets of SMIC.

	SMIC-subHS	SMIC-VIS	SMIC-NIR
Baseline-SMIC [15]	-	52.1%	38.0%
Li <i>et al.</i> [14]	80.28%	81.69%	67.61%
<b>Ours</b>	<b>87.32%</b>	<b>83.10%</b>	<b>69.01%</b>

model achieves superior accuracy to the current best [14], regardless of the video descriptors chosen.

### B. Comparison to the state of the arts

We compare our method to ten state-of-the-art approaches, including both the deep learning methods [23], [16] and the conventional hand-crafting feature approaches [15], [34], [22], [28], [17], [10], [11], [14]. The current best method [14] provides a standard pipeline for MER tasks. As the HIGO-TOP descriptor is recommended by [14], we also use it as the video descriptors. The results are summarized in Table. I and Table. II. For all the databases, our method achieves the best accuracy against all the other methods. More specifically, the accuracy on SMIC-HS, SMIC-subHS, SMIC-VIS, SMIC-NIR, and CASMEII are 68.90%, 87.32%, 83.10%, 69.01%, and 70.85% respectively, while the second best results achieved by [14] are only 68.29%, 80.28%, 81.69%, 67.61%, and 67.21% respectively. It indicates that our ME-Booster is robust to the substantial variations in appearance brought by different sensors.

### C. Computational cost evaluation

Finally, we compare the computational cost of ME-Booster

to the current best work [14]. The experiment is conducted on MATLAB R2017b platform, and the desktop is with an Intel CPU i5-6500 (3.20GHz) and RAM of 8 GB. Results indicate that our model is more efficient on both databases. For SMIC, there are 164 video clips, the running time are 125.49 seconds (s) for Li *et al.* [14] and 27.42 s for our model. For CASMEII, there are 247 video clips, our model can be even ten times faster than the model in [14], the running time are 420.89 s and 41.61 s, respectively.

## IV. CONCLUSIONS

We explore the underlying relationship between MAG and TIM and realize that both of them can be characterized by a linear system. In line with this important finding, we propose a consolidated Eulerian framework, ME-Booster, for revealing the subtle and fleeting facial expressions. ME-Booster boosts the performance of MER in both speed and accuracy. On one hand, it avoids unnecessary separation of the two modules and thus the computational cost is reduced. On the other hand, the accuracy is further improved since it eliminates the side effects brought by the intermediate process. ME-Booster is online, training-free, and compatible with any succeeding processing methods in the MER pipeline. Experiments on two databases indicate that our model outperforms the state-of-the-art approaches in an efficient manner.

## V. ACKNOWLEDGMENTS

This work was supported by the Academy of Finland, Tekes Fidipro program (Grant No. 1849/31/2015), Business Finland project (Grant No. 3116/31/2017), and Infotech Oulu. It was also supported, in part, by the National Natural Science Foundation of China (No. 61772419 & 61572205). Moreover, the authors wish to acknowledge CSC-IT Center for Science, Finland, for computational resources. Furthermore, we express deep gratitude to the support of NVIDIA Corporation with the donation of GPUs for this research.

## REFERENCES

- [1] C. J. Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998.
- [2] P. J. Burt and E. H. Adelson. The laplacian pyramid as a compact image code. In *Readings in Computer Vision*, pages 671–679. Elsevier, 1987.
- [3] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition(CVPR).2005. IEEE Conference on*, volume 1, pages 886–893. IEEE, 2005.
- [4] 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*, 9(1):116–129, 2018.
- [5] 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.
- [6] G. H. Golub and C. F. Van Loan. *Matrix computations*, volume 3. JHU Press, 2012.
- [7] X. Hong, T.-K. Tran, and G. Zhao. Micro-expression spotting: A benchmark. *arXiv preprint arXiv:1710.02820*, 2017.
- [8] X. Hong, Q.-N. Vo, T.-K. Tran, and G. Zhao. Dense prediction for micro-expression spotting based on deep sequence model. 2019.
- [9] X. Hong, Y. Xu, and G. Zhao. Lbp-top: a tensor unfolding revisit. In *Asian Conference on Computer Vision*, pages 513–527. Springer, 2016.
- [10] X. Huang, S.-J. Wang, X. Liu, G. Zhao, X. Feng, and M. Pietikainen. Discriminative spatiotemporal local binary pattern with revisited integral projection for spontaneous facial micro-expression recognition. *IEEE Transactions on Affective Computing*, 2017.
- [11] X. Huang and G. Zhao. Spontaneous facial micro-expression analysis using spatiotemporal local radon-based binary pattern. In *the Frontiers and Advances in Data Science (FADS), 2017 International Conference on*, pages 159–164. IEEE, 2017.
- [12] X. Huang, G. Zhao, X. Hong, W. Zheng, and M. Pietikainen. Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns. *Neurocomputing*, 175:564–578, 2016.
- [13] 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.
- [14] X. Li, X. Hong, A. Moilanen, X. Huang, T. Pfister, G. Zhao, and M. Pietikainen. Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods. *IEEE Transactions on Affective Computing*, 2017.
- [15] X. Li, T. Pfister, X. Huang, G. Zhao, and M. Pietikäinen. A spontaneous micro-expression database: Inducement, collection and baseline. In *Automatic face and gesture recognition (FG), 2013 10th IEEE International Conference on*, pages 1–6. IEEE, 2013.
- [16] 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)*, pages 3094–3098. IEEE, 2018.
- [17] S.-T. Liong, R. C.-W. Phan, J. See, Y.-H. Oh, and K. Wong. Optical strain based recognition of subtle emotions. In *Intelligent Signal Processing and Communication Systems (ISPACS), 2014 International Symposium on*, pages 180–184. IEEE, 2014.
- [18] S.-T. Liong, J. See, R. C.-W. Phan, A. C. Le Ngo, Y.-H. Oh, and K. Wong. Subtle expression recognition using optical strain weighted features. In *Asian Conference on Computer Vision*, pages 644–657. Springer, 2014.
- [19] C. Liu, A. Torralba, W. T. Freeman, F. Durand, and E. H. Adelson. Motion magnification. *ACM transactions on graphics (TOG)*, 24(3):519–526, 2005.
- [20] 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 Transactions on Affective Computing*, 7(4):299–310, 2016.
- [21] S. Park and D. Kim. Subtle facial expression recognition using motion magnification. *Pattern Recognition Letters*, 30(7):708–716, 2009.
- [22] S. Y. Park, S. H. Lee, and Y. M. Ro. Subtle facial expression recognition using adaptive magnification of discriminative facial motion. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 911–914. ACM, 2015.
- [23] D. Patel, X. Hong, and G. Zhao. Selective deep features for micro-expression recognition. In *Pattern Recognition (ICPR), 2016 23rd International Conference on*, pages 2258–2263. IEEE, 2016.
- [24] T. Pfister, X. Li, G. Zhao, and M. Pietikäinen. Recognising spontaneous facial micro-expressions. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1449–1456. IEEE, 2011.
- [25] S. Polikovsky, Y. Kameda, and Y. Ohta. Facial micro-expressions recognition using high speed camera and 3d-gradient descriptor. 2009.
- [26] L. Tian, X. Hong, C. Fan, Y. Ming, M. Pietikäinen, and G. Zhao. Sparse tikhonov-regularized hashing for multi-modal learning. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 3793–3797. IEEE, 2018.
- [27] V. Van Asch. Macro-and micro-averaged evaluation measures [[basic draft]]. 2013.
- [28] 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*, 24(12):6034–6047, 2015.
- [29] X. Wei, H. Li, J. Sun, and L. Chen. Unsupervised domain adaptation with regularized optimal transport for multimodal 2d+ 3d facial expression recognition. In *Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on*, pages 31–37. IEEE, 2018.
- [30] H. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. T. Freeman. Eulerian video magnification for revealing subtle changes in the world. *ACM Trans. Graph. (Proceedings SIGGRAPH 2012)*, 31(4), 2012.
- [31] Z. Xia, X. Feng, X. Hong, and G. Zhao. Spontaneous facial micro-expression recognition via deep convolutional network. In *2018 Eighth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6. IEEE, 2018.
- [32] Z. Xia, X. Hong, X. Gao, X. Feng, and G. Zhao. Spatiotemporal recurrent convolutional networks for recognizing spontaneous micro-expressions. *arXiv preprint arXiv:1901.04656*, 2019.
- [33] H. Xiaohua, S.-J. Wang, X. Liu, G. Zhao, X. Feng, and M. Pietikainen. Discriminative spatiotemporal local binary pattern with revisited integral projection for spontaneous facial micro-expression recognition. *IEEE Transactions on Affective Computing*, 2017.
- [34] 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*, 9(1), 2014.
- [35] Y. Zhang, S. L. Pintea, and J. C. van Gemert. Video acceleration magnification. *arXiv preprint arXiv:1704.04186*, 2017.
- [36] G. Zhao and M. Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE transactions on pattern analysis and machine intelligence (TPAMI)*, 29(6):915–928, 2007.
- [37] Z. Zhou, X. Hong, G. Zhao, and M. Pietikäinen. A compact representation of visual speech data using latent variables. *IEEE transactions on pattern analysis and machine intelligence*, 36(1):1–1, 2014.
- [38] Z. Zhou, G. Zhao, and M. Pietikäinen. Towards a practical lipreading system. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 137–144. IEEE, 2011.
- [39] Y. Zong, X. Huang, W. Zheng, Z. Cui, and G. Zhao. Learning from hierarchical spatiotemporal descriptors for micro-expression recognition. *IEEE Transactions on Multimedia*, 2018.
- [40] Y. Zong, W. Zheng, X. Hong, C. Tang, Z. Cui, and G. Zhao. Cross-database micro-expression recognition: A benchmark. *arXiv preprint arXiv:1812.07742*, 2018.