

EULERIAN EMOTION MAGNIFICATION FOR SUBTLE EXPRESSION RECOGNITION

Anh Cat Le Ngo, Yee-Hui Oh, Raphael C.-W. Phan, John See

Multimedia University,
Cyberjaya 63100, Selangor, Malaysia

ABSTRACT

Subtle emotions are expressed through tiny and brief movements of facial muscles, called micro-expressions; thus, recognition of these hidden expressions is as challenging as inspection of microscopic worlds without microscopes. In this paper, we show that through motion magnification, subtle expressions can be realistically exaggerated and become more easily recognisable. We magnify motions of facial expressions in the Eulerian perspective by manipulating their amplitudes or phases. To evaluate effects of exaggerating facial expressions, we use a common framework (LBP-TOP features and SVM classifiers) to perform 5-class subtle emotion recognition on the CASME II corpus, a spontaneous subtle emotion database. According to experimental results, significant improvements in recognition rates of magnified micro-expressions over normal ones are confirmed and measured. Furthermore, we estimate upper bounds of effective magnification factors and empirically corroborate these theoretical calculations with experimental data.

Index Terms— Subtle emotion, Motion magnification, Micro-expression recognition, Classification

1. INTRODUCTION

Human beings learn to recognize others' emotions through facial expressions i.e. motions of facial muscles. Recognition of an emotion state would very much depend on magnitudes of such motions. While a normal emotion with significant expressions is easily recognized, subtle emotions i.e. involuntary, sudden and brief (1/25s to 1/3s duration) and small expressions are much more difficult to recognize for both human beings and machines. This difficulty was quantified by Frank et al.'s psychological experiments [1] in which untrained and trained human subjects recognise five subtle expressions with accuracy rates of 32% and 47% respectively. Meanwhile, studies of Russell et al. [2] show that *normal* expression recognition can be recognized by human subjects from different races with 75% accuracy in the same five-class task.

In the first FERA Challenge [3], Yang et al. reported an accuracy rate 84% of their artificial system for recognising five normal spontaneous expressions. In contrast, for the case of subtle emotions, Le Ngo et al. [4] and [5] reported a recog-

niton rate of 44% with leave-one-subject-out (LOSO) validation protocol and 65% with leave-one-video-out (LOVO) validation on the CASME II corpus [5] of subtle spontaneous expressions. As the LOSO protocol is broadly employed in other facial expression recognition studies [6], Le Ngo et al.'s recognition rate 44% is used for consistent comparison with previous studies. The difference of recognition rates indicates that large facial motions of normal expressions are likely to play a significant role in their superiority of recognition rate over subtle emotions with small magnitudes of expressions. Intuitively, the more facial motions an expression has, the more recognizable it is.

Therefore, it seems plausible that motion magnification could boost the recognition rate of subtle emotions. This research question is the focus of the current paper, wherein we show that motion magnification of micro-expressions improves recognizability of subtle emotions.

To the best of our knowledge, the only other previous work applying magnification to emotion recognition is by Park et al. [7], where they reported increment of recognition rate when motion vectors at facial key-points are magnified according to predefined motion magnification factors; then, corresponding points at later frames are shifted accordingly. Nevertheless, since such a magnified expression considered in the Park et al. approach, as shown in Fig. 6 of [7], is synthesized from an original expression by piece-wise transformation, the result would visually appear to be unnatural. The distorted expression is due to local shifts at particular facial points with ad-hoc magnification factors as well as inaccurate estimation of complicated motions. Moreover, experiment results reported in [7] are unreliable since Park et al. [7] classify these unnaturally magnified emotions from an undisclosed facial expression databases [8] without psychological ground truth and evaluate with k -fold cross-validation, which is a subject-dependent evaluation technique. Also, in very recent independent works, both Li et al. [9] and Park et al. [10] proposed to apply the amplitude-based Eulerian motion magnification technique of [11] for micro-expression recognition. In contrast, our work here is based on more advanced Eulerian motion magnification [12].

In more detail, we propose the first known realistic magnification of subtle facial expressions and provide corresponding analysis of the effects on recognition rates of micro-

expressions. In the Eulerian perspective, complex motions of subtle expressions can be effectively magnified as properties of motion e.g. velocity and acceleration are assumed to evolve over time at any image pixel not just facial keypoints. In addition, this realistic magnification also improves the recognition rate of subtle emotions in the CASME II [5] database as compared to state-of-the-art methods equipped with complex features [13][14]. Finally, we propose a simple method of estimating the maximum effective magnification factors and show that these match empirically evaluated data.

2. EULERIAN MOTION MAGNIFICATION

Being able to observe small details is a challenge, not just for the naked eye but also in terms of extracting appropriate features for pattern recognition tasks. Small motions can in fact undergo computational magnification [15] [11] to be better recognisable. In *Lagrangian* approaches to motion magnification, motion vectors need to be estimated at a given time and location explicitly, then frames of videos are warped according to magnified vectors. However, accurate estimation of motion is still computationally intensive and error-prone especially in complex and subtle motions i.e. unnatural magnified subtle emotions in [7]. Meanwhile, Eulerian-inspired approaches [11] [16] [12] do not require explicit motion vectors but simulate motion magnification by magnifying changes of properties i.e. amplitude or phase on the whole image grid. As motion in Eulerian perspective depends on types of signal properties like amplitude or phase, there are amplitude-based (A-EMM) and phase-based (P-EMM) Eulerian Motion Magnification (EMM).

2.1. Amplitude-based Eulerian Motion Magnification

Let $I(x, t)$ denote an image profile at location x and time t . Having undergone a translational motion with a displacement function $\delta(t)$, the image profile is rewritten as $I(x, t) = f(x + \delta(t))$ and $I(x, 0) = f(x)$. The motion in Eulerian perspective is characterised by differences of intensity $B(x, t)$ given that $I(x, t) = I(x) + B(x, t)$; then, a pixel intensity \hat{I} of a magnified motion is computed as

$$\hat{I}(x, t) = I(x) + \alpha * B(x, t) \quad (1)$$

where α is a magnification factor. Assume that only small translational $\delta(t)$ motion occurs, $\hat{I}(x, t)$ can be approximated [11] by the first-order Taylor series as follows.

$$\hat{I}(x, t) \approx f(x) + \sum_k \alpha B(x, t) \quad (2)$$

where k denotes a passband of a temporal filter with a corresponding attenuation factor γ_k and $B(x, t)$ is output of the temporal bandpass filter:

$$B(x, t) = \sum \gamma_k \delta(t) \frac{\delta f(x)}{\delta x} \quad (3)$$



(a) $\alpha = 0$ (b) $\alpha = 3$ (c) $\alpha = 7$
Fig. 1: Amplitude-based Eulerian Motion Magnification



(a) $\alpha = 0$ (b) $\alpha = 3$ (c) $\alpha = 7$
Fig. 2: Phase-based Eulerian Motion Magnification

2.2. Phase-based Eulerian Motion Magnification

In the spectral domain, the shifted image profile $I(x, t) = f(x + \delta(t))$ can be re-written through Fourier series decomposition as follows:

$$f(x + \delta(t)) = \sum_{\omega=-\infty}^{+\infty} A_{\omega} e^{i\omega(x+\delta(t))} = \sum_{\omega=-\infty}^{+\infty} I_{\omega} A_{\omega} e^{i\omega\delta(t)}$$

where $I(x, 0) = \sum_{\omega=-\infty}^{+\infty} I_{\omega}$ represents an image profile at $t = 0$ with I_{ω} as its Fourier coefficients. In this spectral domain, the EMM is realised by magnifying band-passed phase shift $B = \omega\delta(t)$ with magnification factor α . Therefore, the approximated magnified image $\hat{I}(x, t)$ is written [16] as:

$$\hat{I}(x, t) = \sum_{\omega=-\infty}^{+\infty} \hat{I}_{\omega} \approx \sum_{\omega=-\infty}^{+\infty} I_{\omega} e^{i\alpha B} \quad (4)$$

where \hat{I}_{ω} are Fourier coefficients of motion magnified images.

3. MAGNIFIED SUBTLE EMOTION RECOGNITION

The main thesis of this paper is that the challenge of recognizing subtle emotions can be partly eased if this subtle and short-lived motion is properly magnified. To substantiate this, we show the first known analysis of how Eulerian Motion Magnification (EMM) improves recognition of subtle expressions on a publicly available corpus. For such subtle motions, Eulerian magnification induces less noise than the Lagrangian approach as suggested by Rubinstein et al. [17]. Fig.1b,1c and 2b,2c illustrate exaggerated expressions at magnification factors (α) 3,7 when A-EMM and P-EMM magnify original subtle expressions in Fig. 1a and 2a. As normal facial expressions are analysed in a general concep-

tual framework [6]: facial registration, feature extraction, and classification, similar steps are utilised here for classifying EMM-magnified subtle expressions. All facial samples of the CASME II database are registered by Active Appearance Model (AAM); then, identity of human subjects is partly suppressed by warping their faces according to a common template by Local Weighted Mean (LWM) transformation [18]. After that, A-EMM or P-EMM magnifies nearly invisible motion of subtle facial expressions with magnification factors α before Local Binary Pattern with Three Orthogonal Plane (LBP-TOP) [19] extracts feature vectors from these artificially exaggerated emotion. After the motion magnification and feature extraction processes are done for all video samples, the generated LBP-TOP feature vectors are fed into the Support Vector Machine (SVM) with Linear (LIN) kernel for training and evaluating the multi-class classifier of subtle emotions. As the number of video samples for training is much smaller than the dimensionality of LBP-TOP features [19] i.e. histograms of spatio-temporal local blocks, LIN kernel of data is sufficient for forming optimal hyperplanes.

Subject identities will significantly impact the evaluation results if the same subjects appear in both training and testing corpus. Therefore, the classifier is trained with the Leave-One-Subject-Out (LOSO) validation approach. In this approach, training samples are taken from all 25 subjects in the CASME II corpus except one test subject, and the test subject is sequentially selected from all 26 subjects. As a result, a classifier is trained and tested 26 times and its performance is averaged from these 26 evaluation results. To evaluate the trained classifiers, we employ F1 scores, precision and recall rates as numerical results instead of accuracy or recognition rate as suggested by Le Ngo et al. [4]. These evaluation measurements are less prone to bias, caused by skewed sample distributions over emotional classes. Note that all above abbreviations appear in figures, presented in the Section 4.

In order to well-approximate magnified signals, we propose to bound the magnification factors of A-EMM (α_{A-EMM}) and P-EMM (α_{P-EMM}) by $(1 + \alpha_{A-EMM}) * \delta(t) < \frac{\lambda_c}{8}$ [11] and $\alpha_{P-EMM} \delta(t) < \frac{\lambda_c}{4}$ [16] with respect to spatial cut-off wavelength λ_c and motion duration δ_t . The reasoning is as follows: Wu et al. [11] mentions dependence of spatial cut-off wavelength on particular applications. As Mermillod et al. [20] show spatial frequencies lower than 32 cycles per image (CPI) contain the best features for recognition of facial expressions, the cut-off spatial wavelength for micro-expression is $\lambda_c = \frac{D}{32}$ where D is a diagonal length of a sample image. In our experiment, an input frame is resized to 320×240 pixels hence $D = \sqrt{320^2 + 240^2} = 400$ and the corresponding cut-off spatial wavelength is 12.5 pixels per cycle (PPC). Furthermore, durations of micro-expressions last from $\frac{1}{25}$ of a second to $\frac{1}{15}$ of a second, which means $\frac{1}{25} < \delta(t) < \frac{1}{15}$. Given $\lambda_c = 12.5$ and $\delta(t) \in [\frac{1}{25}, \frac{1}{15}]$, we can compute boundaries of A-EMM and P-EMM magnification factors, $\alpha_{A-EMM} < 23$ and $\alpha_{P-EMM} < 47$. In the

following experiments, magnification factors are varied for completeness, between 0 and 90 with a step of 3. The range of factors with maximal recognition rates will be shown to corroborate the above estimations.

4. EXPERIMENTS & RESULTS

In LBP-TOP feature extraction, images are spatially resized to 320×240 pixels resolution and partitioned into non-overlapping 5×5 blocks, then corresponding blocks of all frames are stacked up in 3-D volumes. Spatial-temporal features, histograms of binary patterns appearing in these volumes, are extracted from the volumes by LBP-TOP_{4,4,4,1,1,4}. The first three numbers (4,4,4) represent 4-neighbor connections in three orthogonal planes and the last three numbers are radii of these respective connection. The SVM classifier with LIN kernel uses very large penalty factors ($c = 10000$) for regularizing the learning process as differences in facial micro-expressions are often superseded by subject identities.

4.1. Experimental Database

Experiments are done with the CASME II corpus [5], the most comprehensive database of spontaneous subtle emotions so far. It contains a reasonably large number (247) of video samples, elicited from 26 Asian participants with an average age of 22.03 years old. Subjects are instructed to hide their true emotion while being shown emotionally stimulating short movies. At the same time, their facial responses are recorded in videos which are later post-processed and labeled by specialists trained to recognize subtle emotions. These experts separate the recording into video samples with respect to single and complete subtle expressions. Then, they assign each sample into one of five emotional categories: Tense (T), Disgust (D), Happiness (H), Surprise (S) or Repression (R). It is noted that specialists are aided with high frame-rate (200 *fps*) video recording in 290×340 resolution by a Point Grey GRAS-03K2C camera as subjects's responses are elusive and short-lived. To avoid flickering light and to ensure stable illuminating condition during the recording, four LED lamps under umbrella reflectors are directed toward the human subjects. Overall, CASME II is a high-quality database of spontaneous subtle expressions created with a reliable labeling process on a reasonably large number of subjects.

4.2. Experimental Result

Figure 3 shows F1-scores of subtle emotion recognition for both A-EMM and P-EMM over a range of magnification factors $\alpha = [0, 90]$. Their performances increase in accordance to the increment of α ; however, the recognition performance does not improve but declines after the maximum effective magnification factor. As noises are also magnified alongside motions, artificially introduced noise diminishes the benefits

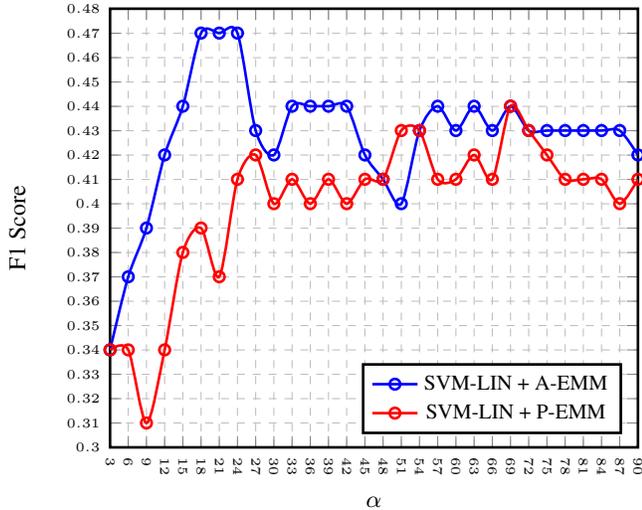


Fig. 3: F1-score of 5-class recognition of A-EMM and P-EMM magnified micro-expressions w.r.t various magnification factors α from 0 to 90

of the magnified motion. Therefore, the recognition performances appear to reach their peaks at the maximum effective magnification factors.

In Sub-section 3, we have theoretically estimated these magnification factors with knowledge about the useful cut-off wavelength of facial features and the duration of a subtle expression. Data in Figure 3 about performances of A-EMM and P-EMM allow to empirically test these estimations. More precisely, the maximum effective magnification factor for A-EMM is estimated to be 23. Figure 3 shows that peak performances of A-EMM happen between $\alpha = 18$ and $\alpha = 24$, which corroborates with the estimated 23. Meanwhile, P-EMM was estimated to reach its peak performance around $\alpha = 47$, and correspondingly, the figure shows two local peaks $\alpha = 48$ and $\alpha = 69$ on the red plot of P-EMM. The estimation of maximum effective magnification factor is more accurate for A-EMM than for P-EMM. Moreover, the F1-scores of A-EMM are almost always better than those of P-EMM for all magnification factors (see Figure 3). It might be due to that P-EMM is more susceptible to noise than A-EMM is. Further studies on signal-to-noise ratio of magnified videos are required but not considered in this paper due to page limitation.

Table 1 shows performances of previously proposed methods [4][13][14] for automatic recognition of subtle emotions in CASME II corpus with the LOSO validation protocol. At the first line is performance of a baseline method, which includes LBP-TOP features and SVM classifier [5]. Instead of SVM, Le Ngo et al. [4] employ AdaBoost for recognition task of which performance is mentioned at the second line of the table. At the third line is the performance of Liong et al. [13]’s method of where local LBP-TOP features are weighted w.r.t the total amount of optical strain magnitudes in a region. While both baseline and Liong et al.’s method only

use LBP-TOP of image intensities, Oh et al. [14] utilize additional phase and orientation channels from Riesz-Wavelet transform for LBP-TOP feature extraction. Liong et al.’s weighting strategy provides modest improvement, i.e. 3% in F1 score, over the baseline method with additional computational cost of optical strain estimation. Meanwhile, the approach of encoding additional data channels [14] proves to be more successful with 8% improvement in F1 score against the baseline method. However, it is also computationally heavy as the dimensions of input data are tripled. Different from the mentioned approaches on enhancing feature extractions, we choose to magnify facial motions so they become more distinguished in both visual and feature domains. Performance of P-EMM at the magnification factor 69, P-EMM ($\alpha = 69$) is better than Oh et al.’s method marginally by 1% in F1-score, but with less computational complexity. Furthermore, A-EMM provides even better performance i.e. 0.47 in F1-score which outperforms all previously mentioned approaches as shown in Table 1. It is noted that both A-EMM and P-EMM only encode LBP-TOP features with a gray-scale intensity channel while Oh et al. method encode more channels, i.e. amplitude, phase and orientation, into LBP-TOP features. These results show that motion magnification techniques can significantly reduce errors in recognizing subtle emotions without complex and high-dimensional features for suitable choices of magnification factors.

Table 1: Accuracy (ACC), F1-score (F1), precision (Pre) and recall (Rec) results

Method	ACC	F1	Rec	Pre
Baseline	0.38	0.35	0.34	0.36
Le Ngo et al. [4]	0.44	0.33	0.53	0.29
Liong et al. [13]	0.42	0.38	0.36	0.41
Oh et al. [14]	0.46	0.43	0.35	0.55
P-EMM ($\alpha = 69$)	0.50	0.44	0.44	0.45
A-EMM ($\alpha = 24$)	0.51	0.47	0.44	0.49

5. CONCLUSION

Eulerian Motion Magnification allows to exaggerate subtle emotions visually and boost recognition rate of 5-class subtle expression problem by 0.12 in F1-score. The highest recognition rate is 0.47, which is equivalent to the recognition rate (47%) of trained experts and outperforms current state-of-the-art methods. Among the two variations of EMM, A-EMM outperforms P-EMM in the context of recognition rates when facial sequences are exaggerated with increasing magnification factors $\alpha = [0 \dots 90]$.

6. ACKNOWLEDGEMENT

This research was supported by TM (Telekom Malaysia) under the projects UbeAware and 2beAware.

7. REFERENCES

- [1] MG 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.
- [2] James A Russell, “Is there universal recognition of emotion from facial expressions? a review of the cross-cultural studies.,” *Psychological bulletin*, vol. 115, no. 1, pp. 102, 1994.
- [3] Michel François Valstar, Bihan Jiang, Marc Mehu, Maja Pantic, and Klaus Scherer, “The first facial expression recognition and analysis challenge,” in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*. IEEE, 2011, pp. 921–926.
- [4] Anh Cat Le Ngo, Raphael C.-W. Phan, and John See, “Spontaneous subtle expression recognition: Imbalanced databases & solutions,” in *Asian Conference on Computer Vision (ACCV)*, 2014.
- [5] Wen-Jing Yan, Xiaobai Li, Su-Jing Wang, Guoying Zhao, Yong-Jin Liu, Yu-Hsin Chen, and Xiaolan Fu, “Casmie ii: An improved spontaneous micro-expression database and the baseline evaluation,” *PLoS one*, vol. 9, no. 1, pp. e86041, 2014.
- [6] Evangelos Sariyanidi, Hatice Gunes, and Andrea Cavallaro, “Automatic analysis of facial affect: A survey of registration, representation and recognition,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2014.
- [7] Sungsoo Park and Daijin Kim, “Subtle facial expression recognition using motion magnification,” *Pattern Recognition Letters*, vol. 30, no. 7, pp. 708–716, 2009.
- [8] Sungsoo Park, Hyung-Soo Lee, Jongju Shin, and Daijin Kim, “The postech subtle facial expression database 2007,” in *Proc. of the 8th POSTECH-KYUTECH Joint Workshop On Neuroinformatics*, 2008, pp. 653–54.
- [9] Xiaobai Li, Xiaopeng Hong, Antti Moilanen, Xiaohua Huang, Tomas Pfister, Guoying Zhao, and Matti Pietikäinen, “Reading hidden emotions: Spontaneous micro-expression spotting and recognition,” *arXiv preprint arXiv:1511.00423*, 2015.
- [10] SungYeong Park, SeungHo Lee, and YongMan Ro, “Subtle facial expression recognition using adaptive magnification of discriminative facial motion,” in *Multimedia (MM), 2015 23rd ACM International Conference on*. ACM, 2015, pp. 911–914.
- [11] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John V Guttag, Frédo Durand, and William T Freeman, “Eulerian video magnification for revealing subtle changes in the world.,” *ACM Trans. Graph.*, vol. 31, no. 4, pp. 65, 2012.
- [12] Neal Wadhwa, Michael Rubinstein, Frédo Durand, and William T Freeman, “Riesz pyramid for fast phase-based video magnification,” in *Computational Photography (ICCP), 2014 IEEE International Conference on*. IEEE, 2014.
- [13] Sze-Teng Liong, John See, Raphael C-W Phan, Anh Cat Le Ngo, Yee-Hui Oh, and KokSheik Wong, “Subtle expression recognition using optical strain weighted features,” in *Computer Vision-ACCV 2014 Workshops*. Springer, 2014, pp. 644–657.
- [14] Yee-Hui Oh, Anh Cat Le Ngo, John See, Sze-Teng Liong, Raphael C-W Phan, and Huo-Chong Ling, “Monogenic riesz wavelet representation for micro-expression recognition,” in *Digital Signal Processing (DSP), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1237–1241.
- [15] Ce Liu, Antonio Torralba, William T Freeman, Frédo Durand, and Edward H Adelson, “Motion magnification,” in *ACM Transactions on Graphics (TOG)*. ACM, 2005, vol. 24, pp. 519–526.
- [16] Neal Wadhwa, Michael Rubinstein, Frédo Durand, and William T Freeman, “Phase-based video motion processing,” *ACM Transactions on Graphics (TOG)*, vol. 32, no. 4, pp. 80, 2013.
- [17] Michael Rubinstein, *Analysis and Visualization of Temporal Variations in Video*, Ph.D. thesis, Massachusetts Institute of Technology, Feb 2014.
- [18] Ardeshir Goshtasby, “Image registration by local approximation methods,” *Image and Vision Computing*, vol. 6, no. 4, pp. 255–261, 1988.
- [19] Guoying Zhao and Matti Pietikäinen, “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.
- [20] Martial Mermillod, Patrick Bonin, Laurie Mondillon, David Alleysson, and Nicolas Vermeulen, “Coarse scales are sufficient for efficient categorization of emotional facial expressions: Evidence from neural computation,” *Neurocomputing*, vol. 73, no. 13, pp. 2522–2531, 2010.