

Effective recognition of facial micro-expressions with video motion magnification

Yandan Wang¹ · John See² · Yee-Hui Oh³ ·
Raphael C.-W. Phan³ · Yogachandran Rahulamathavan⁴ ·
Huo-Chong Ling⁵ · Su-Wei Tan³ · Xujie Li¹

Received: 30 March 2016 / Revised: 21 August 2016 / Accepted: 18 October 2016/
Published online: 8 November 2016
© Springer Science+Business Media New York 2016

Abstract Facial expression recognition has been intensively studied for decades, notably by the psychology community and more recently the pattern recognition community. What is more challenging, and the subject of more recent research, is the problem of recognizing subtle emotions exhibited by so-called micro-expressions. Recognizing a micro-expression is substantially more challenging than conventional expression recognition because these micro-expressions are only temporally exhibited in a fraction of a second and involve minute spatial changes. Until now, work in this field is at a nascent stage, with only a few existing micro-expression databases and methods. In this article, we propose a new micro-expression recognition approach based on the Eulerian motion magnification technique, which could reveal the hidden information and accentuate the subtle changes in

Electronic supplementary material The online version of this article (doi:[10.1007/s11042-016-4079-6](https://doi.org/10.1007/s11042-016-4079-6)) contains supplementary material, which is available to authorized users.

✉ Yandan Wang
yandan.wang@wzu.edu.cn

¹ School of Physics and Electronic Information Engineering, Wenzhou University, Wenzhou Zhejiang 325035 China

² Faculty of Computing and Informatics, Multimedia University, Jalan Multimedia 63100 Cyberjaya, Selangor, Malaysia

³ Faculty of Engineering, Multimedia University, Jalan Multimedia 63100 Cyberjaya, Selangor, Malaysia

⁴ Information Security Group, City University London, London, EC1V 0HB, UK

⁵ Department of Electrical and Computer Engineering, Curtin University, CDT 250, 98009 Miri Sarawak, Malaysia

micro-expression motion. Validation of our proposal was done on the recently proposed CASME II dataset in comparison with baseline and state-of-the-art methods. We achieve a good recognition accuracy of up to 75.30 % by using leave-one-out cross validation evaluation protocol. Extensive experiments on various factors at play further demonstrate the effectiveness of our proposed approach.

Keywords Micro-expressions · Motion magnification · EVM · CASME II · Local binary patterns

1 Introduction

Facial micro-expression is a subtle, short, quick (1/3 to 1/25 second [8, 33]) and involuntary expression, which appears when people try to conceal their genuine emotions, especially in high-stake situations [8]. While *macro-expressions* are normal facial expressions that are visibly noticeable and are normally intended, *micro-expressions* are much more rapid and involuntary, or at times, an attempt to hide the true emotion felt. For example, in some situations, people might be very sad, but she showed happiness to disguise her sadness for not worrying people who cares her much. However, involuntarily and unconsciously, there is still a fast expression of sadness revealed in her face which is too fast to be captured by our naked eyes. Only when we examined the recorded video frame by frame, we will find the subtle changes in her face, the so-called micro-expression. Due to the short duration of micro-expressions compared to that of macro-expressions, micro-expressions are difficult to be detected in real-time conversations [8]. Even the experts with professional training can only achieve 47 % recognition accuracy [10] by using their naked eyes to perform the tedious task of frame-by-frame observation. Therefore, an automatic and accurate micro-expression recognition system will be helpful and useful to tackle these issues.

The ability to recognize subtle emotions in humans has vast applications in different domains, ranging from lie detection in the case of crime and public safety to better understand patients with special needs such as autism [39] and schizophrenia [36]. In clinical work, to detect and recognition micro-expression is important to assist psychologists in the diagnosis and remediation of patients with mental diseases. Sometimes patients might not be cooperative towards psychologists, thus the detection of micro-expressions can help psychologists judge whether the patients are telling the truth. Similarly in criminal investigation, the ability to identify micro-expressions also helps the police interrogate suspects by analyzing the truthfulness of their responses as portrayed by their hidden expression [8]. As such, machine automated recognition of facial micro-expressions would be enormously valuable.

While research in the direction of micro-expressions has seen considerable effort in the past decades in the discipline of psychology, research into micro-expressions is only starting to thrive in the discipline of pattern recognition and machine vision. There are currently two promising spontaneous datasets for micro-expressions, notably the Spontaneous Micro-expression (SMIC) [16] and Chinese Academy of Sciences Micro-expression (CASME) II [49] databases. SMIC is of a small sample size taken from a small number of subjects, thereby leaving CASME II to be the most comprehensive dataset for micro-expressions to date. The creators of the CASME II dataset reported accuracy rates of up to 63.41 % using standard techniques of Local Binary Patterns-Three Orthogonal Planes (LBP-TOP) [53] for feature extraction and Support Vector Machine (SVM) for classification [7].

The CASME II database is video based, consisting of an ordered sequence of images; hence, dynamic textures are favored to describe the spatio-temporal information present in videos. Local Binary Pattern - Three Orthogonal Planes (LBP-TOP) [53] is a dynamic texture descriptor that is simple for computing and robust for recognition. The descriptor considers the concatenation of LBP histograms on three orthogonal planes – XY, XT and YT respectively. The XT and YT planes preserve the temporal transition information that explains the facial movement displacement e.g. how eyes, lips, muscle or eyebrows move. In contrast, the XY plane computes only spatial information that explains both expression and identity information of a face appearance.

However, to solve the micro-expression recognition task using only the LBP-TOP technique is not sufficient to fully describe the hidden information and patterns in micro-expression faces because the spatio-temporal changes induced by micro-expressions are not obvious. In this paper, we propose a new technique for micro-expression recognition that adopts Eulerian Video Magnification (EVM) [48] to amplify subtle facial expression motions. The motion magnification process first applies spatial decomposition pyramid and temporal filtering to the video frames before amplifying the resulting signals to reveal hidden motion information. Following that, spatio-temporal feature patterns are then extracted from the motion-amplified face video sequence. Through this process, our best accuracy of 75.30 % is achieved by using LBP-TOP for feature extraction and SVM with Radial Basis Function (RBF) kernel for classification using leave-one-out cross-validation (LOOCV), the same evaluation strategy as the baseline approach. Our novelty and contributions are listed as follows:

- To our best knowledge, this is the first extensive application of EVM for the amplification of micro-expression motions for micro-expression recognition.
- Video motion magnification by EVM is able to extract more discriminative spatio-temporal features, dramatically improving the recognition of facial micro-expressions over state-of-the-art methods.
- We demonstrated the effectiveness of applying EVM on a variety of scenarios – different frame block partitions, plane combinations, local feature neighborhoods and various classifiers.

The rest of the paper is organized as follows: We review the related work in Section 2. Our approach and methods are described in Section 3 while experimental results and in-depth analysis are shown in Section 4. Finally, conclusion and future work are provided in Section 5.

2 Related work

2.1 LBP and its variants

Local Binary Patterns (LBP) which are introduced by Ojala et al., is a class of features originally proposed for texture classification [25]. In recent years, LBP based features have found applications in facial image recognition [13]. The original version of the LBP operator [24] works in a 3×3 pixel block of a gray scale image. The non-center pixels in this block are thresholded by its center pixel value, multiplied by powers of two and then summed to obtain a label for the center pixel. The basic LBP operator was revised by Ojala et al. and a more generic form of LBP operator [25] was presented several years after the

original publication. This generic revised form of LBP operator does not restrict on the size of the neighborhood or the number of sampling points. As shown in Fig. 1, the size of the neighborhood is controlled by the R index; the number of sampling points is set by the P index.

More precisely, given a pixel c located at (x_c, y_c) , then the LBP code is computed as follows:

$$LBP_{P,R}(x_c, y_c) = \sum_{p=0}^{P-1} s(i_p - i_c) \times 2^p \tag{1}$$

where

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

i_c denotes the intensity of that pixel c , i_p indexed by P denotes the intensity of the circular neighbors of c of radius R , and $s(x)$ is a function that outputs 1 if x is non-negative and 0 otherwise, and then proceeds to compute the histogram of all LBP codes for all pixels in an image or region, and this histogram represents the texture feature descriptor of that image or region.

Different variants of LBP have been subsequently proposed in the literature. The most basic extension of LBP descriptors would be rotation invariant uniform pattern LBP [25]. To be insensitive to rotation, $P - 1$ bitwise circular shifts are performed on the binary pattern and the smallest value is chosen to be the texture feature for that block. A binary pattern is defined as a uniform pattern if the total number of bitwise (0-1 or 1-0) transitions is at most 2 on the binary string traversed in circular form. To consider both the spatial and temporal information of a dynamic texture, [53] proposed a 3D variant of the LBP. So-called LBP-TOP, this considers the co-occurrences on three types of orthogonal planes: the XY plane, the XT plane and the YT plane.

Another 3D LBP variant was proposed by [9], extending the original LBP from 2D images to 3D volumetric data while maintaining full rotational invariance. To further reduce the vector dimension, [11] proposed a variant of LBP, called Center-Symmetric LBP (CS-LBP), where only center-symmetric pairs of pixels are compared instead of comparing each pixel with the center pixel. To overcome the problem of sensitivity to noise in uniform image regions, another variant called Local Ternary Patterns (LTP) was proposed by [43]. This method suggested a 3-valued coding scheme that is quantized to 0, +1 and -1 based on whether the neighborhood pixel intensity is equal, higher or less than that of the center pixel intensity, i_c .

In another direction, instead of the original circular neighborhood definition, an elliptical neighborhood definition was proposed for LBP, namely the Elliptical Local Binary

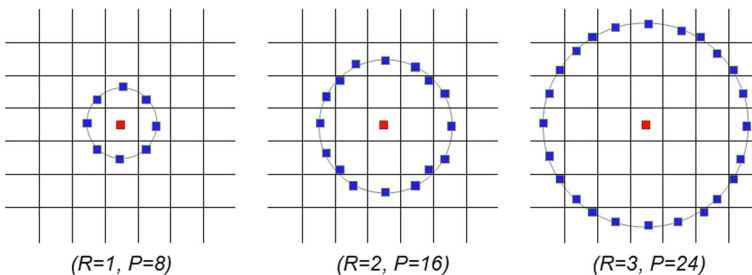


Fig. 1 Circularly symmetric neighbor sets for different (P;R)

Pattern (ELBP) [23], which was applied in face recognition to exploit the anisotropic structural information. The concept of applying of elliptical patterns was originally proposed by [17].

Besides the extensions or variants of LBP, researchers have tried to combine LBP with other feature extraction methods for better performance. Wang et al. [45] proposed a new method combining Histogram of Oriented Gradients (HOG) and LBP, called HOG-LBP. This method is notably robust to partial occlusion as the HOG-LBP features contain both shape/edge information and texture information. Ahonen et al. [1] proposed a rotation invariant image descriptor - LBP Histogram Fourier features (LBP-HF). It is a descriptor computed from discrete Fourier transforms of LBP histogram. The LBP-HF features are rotation invariant (globally for the whole image to be described) and have high discriminative power. Azmi and Yegane [2] proposed a method called Local Gabor Binary Pattern (LGBP), which is the combination of Gabor and LBP. LGBP is robust to the appearance variations due to misalignment and lighting.

2.2 Emotion and micro-expression recognition

In this section, we will review some popular techniques which have been applied in facial macro-expression recognition (emotion recognition) and micro-expression recognition. To some extent, macro-expressions and micro-expressions are in common, thus techniques applied for macro-expressions could potentially work for micro-expression recognition as well. However, micro-expressions are more difficult to analyze than macro-expressions, since micro-expressions last only 1/3 to 1/25 seconds and involve minute spatial variations that can hardly be detected by naked eyes.

2.2.1 Emotion recognition

Emotion recognition has been researched for decades primarily by the psychology community [8]. In the pattern recognition and machine vision community, research started on posed datasets [14, 20, 21, 28] i.e. where subjects were asked to show how they are feeling; it was subsequently felt that spontaneous datasets [3, 22] capture more realistically the actual emotional situations experienced by subjects i.e. subjects were induced to naturally express their feelings rather than being asked to show a specific feeling. Stimuli included carefully designed video clips to trigger different emotions.

A facial expression recognition method involves two different phases: facial feature extraction and classification. Facial feature extraction involves extracting features of facial images; the resulting feature vectors can then be used to project the facial image from the higher dimensional image space into a lower dimensional feature space while preserving the discriminative features. Discriminative features separate the facial images of one class from facial images of other classes in the lower dimensional feature space [35]. Better separation among classes in the lower dimensional space leads to higher classification rate.

Generally, the features that are commonly adopted can be categorized into two major types: geometric features and appearance features. Geometric features measure the displacement of face components such as eyes while appearance features capture the texture changes on a face when an action is performed such as smiling. Geometry based techniques include Haar-like feature extraction, LBP, and Gabor wavelets. Bashyal and Venayagamoorthy [4] proposed recognition of facial expressions using Gabor wavelets and Learning Vector Quantization (LVQ). This method was used to recognize seven different facial

expressions from still pictures of the human face. Shan et al. [40] formulated Boosted-LBP to extract the most discriminant LBP features. These selected features were then classified by SVM. A better performance is obtained by using the Boosted-LBP features instead of LBP features. Satiyan et al. [38] attempted to recognize facial expressions by using Haar-like features. Namely, six statistical features mean, variance, standard deviation, power, energy and entropy which were derived from the approximation coefficients of Haar-like decomposition, and then use the neural network for classification.

Appearance based techniques consist of two fundamental methods i.e., Principle Component Analysis and Independent Component Analysis (ICA). Oliveira et al. [26] suggested 2D PCA (2DPCA) for facial expression recognition. A multiobjective genetic algorithm called the Nondominated Sorting Genetic Algorithm (NSGA) [27] was applied on the 2DPCA features to perform further feature selection. The proposed approach was evaluated on the Japanese Female Facial Expression (JAFPE) database. Both classifiers, kNN and SVM were used to classify the 2DPCA features. Uddin et al. [44] suggested to utilize Enhanced Independent Component Analysis (EICA) to extract independent features which will then be classified by Fisher Linear Discriminant Analysis (FLDA) [5]. Using these extracted features, Hidden Markov Models (HMMs) is then utilized to model various expressions like joy, sad, anger, surprise, fear, and disgusting. The proposed approach was evaluated on Cohn-Kanade database.

Besides these two types of techniques for feature extraction, there is another type of technique called model based techniques. In the model-based methods, the statistical model is used to recognize facial expressions and this statistical model is actually constructed from training images. Lucey et al. [20] suggested to adopt holistic model based approach called the Active Appearance Model (AAM) and a linear SVM classifier for detecting both Action Units (AUs) and emotions on the Extended Cohn-Kanade (CK+) database. Samarawickrame et al. [37] presented the promising accuracy and effectiveness of applying Active Shape Models (ASM) and SVM. The facial coordinates which are located by ASM, were fed into SVM for classification. The proposed method was evaluated on JAFPE Database. Once the feature extraction is performed then classification can be done by kNN classifier or SVM classifier.

2.2.2 *Micro-expression recognition*

More recently, the research community has realized that while emotion recognition has seen considerable research by both the psychology and vision community, challenges remain in terms of being able to recognize more subtle emotions.

In the literature, there are not many papers working on facial micro-expression recognition due to the lack of the existing micro-expression based high-speed video databases. Shreve et al. [41, 42] computed the strain magnitude of optical flow to discriminate micro-expressions from macro-expressions by observing the interval flow in a given threshold. They evaluated their methods on BU [52], USF and USF-HD databases. however, the work is contributed to the spotting the micro and macro expression in video only rather than the recognition. Polikovskiy et al. [33] extracted 3D-Gradient descriptors from the widely used face region (FACS), and the approach was evaluated on their own 200 fps high-speed camera recorded dataset. However, automatic FACS region detection still remains a challenge. To mark face region manually would be tedious and professional training is required. Park and Kim [29] used motion magnification to detect subtle facial expression. They first marked facial shape points by using AAM model, and then aligned the face followed by motion vector of 27 feature points magnified. The framework was evaluated on the SFED2007 [30]

dataset. However, the SFED2007 dataset is actually based on static images rather than video and it is not popularly used.

In the first attempt at recognizing spontaneous micro-expressions, Pfister et al. [32] used temporal interpolation was used to count the length of short videos and the spatio-temporal local texture descriptors to handle dynamic features with SVM, MKL and RF for classifier training. Their approach was evaluated on the Spontaneous MICro-expression (SMIC) database [16] which was recorded using a 100 fps high-speed camera. They reported that temporal interpolation could help achieve equivalent micro-expression detection performance to a standard 25 fps camera. Later on, Yan et al. [49] proposed CASME II, the most comprehensive micro-expression dataset up to date, which comprises a total of 247 videos from 26 subjects, captured at 200fps and coded into 5 class labels. More recently, more approaches have been proposed for micro-expression recognition on this dataset; Liong et al. [18, 19] introduced features derived from optical strain information, Wang et al. [46, 47] reinvented the popular LBP-TOP into efficient variants that retain essential information, while Park et al. [31] attempted to improve recognition by leveraging on an adaptive motion magnification approach.

With the exception of [31], most works in literature focused on the extraction of features from micro-expression samples, and does not attempt to magnify these subtle facial motions before feature extraction. We hypothesize that the subtlety of these motions is a bottleneck to better performance as the minute change of intensities at these areas cannot contribute better representation. Our work differs from the work of Park et al. [31] in that we do not subscribe to the idea of adaptively selecting the most discriminative frequency band when it is not constrained to AU-specific locations, i.e. a chosen frequency band may not be suitable for different parts of the face. Moreover, their work considers only the highest frequency cutoff of 10 Hz while micro-expressions of shorter durations can potentially reach the upper limit of 15–25 Hz [8] (though Yan et al. [50] found that these are usually rare). Our work also emphasizes on the choice of magnification factor, which is a crucial factor in achieving good performance.

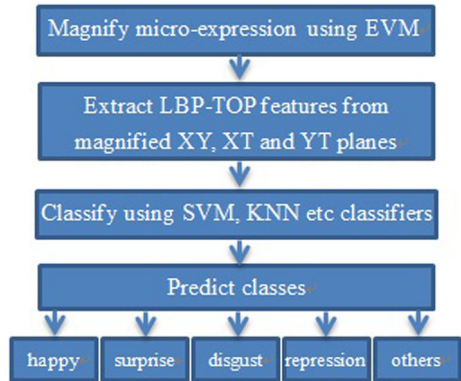
3 Proposed approach

The workflow (Fig. 2) of our proposed approach is succinctly summarized with the following three main steps: 1) Video motions are pre-processed and amplified with Eulerian Video Magnification (EVM); 2) Spatio-temporal feature patterns are extracted from the motion-amplified data by LBP-TOP; 3) SVM classification is performed on the features to recognize the facial micro-expression present in the video. The following sub-sections will elaborate these steps in detail.

3.1 Eulerian video magnification

EVM was proposed by [48] to reveal subtle motion changes in videos, which can be almost impossible to see with the naked eye. In general, there are 4 steps to amplify the subtle motion: 1) compute the full Laplacian pyramid [6], which decomposes the frame sequences of the video into different spatial frequency bands. 2) A band-pass filter, e.g. Ideal, Butterworth, Second-order IIR, is applied to extract the frequency bands of interest. 3) Amplify the motion by multiplying the extracted bandpassed signal at different spatial level by a magnification factor α . 4) The amplified signal is added back to the original to obtain the final motion magnified video.

Fig. 2 The flow diagram of the proposed approach, with three main steps: 1) micro-expression signals are amplified using EVM; 2) LBP-TOP features are extracted from motion-magnified XY, XT and YT planes; 3) SVM is used to predict classes



3.1.1 Motion magnification using first-order Taylor series expansions

In [48], the mathematical inference of EVM is presented as the first-order expansion of the Taylor series. The magnification is produced by analyzing the temporal motion based on the first-order Taylor series expansions rather than directly tracking the motion. We will briefly mathematically describe the relationship between temporal processing and motion magnification in depth next. For further reference, the details of EVM can be found in [48].

Let $I(x, t)$ be the image intensity at position x and time t , and $\delta(t)$ denote the motion displacement function with respect to the observed pixel intensity at position x , such that $I(x, t) = f(x + \delta(t))$ and $I(x, 0) = f(x)$. By equation, we will have the synthesized signal that is magnified by the factor α

$$\widehat{I}(x, t) = f(x + (1 + \alpha)\delta(t)) \tag{3}$$

The first-order Taylor series expansion is used to approximately expand the image displaced intensity $f(x + \delta(t))$ about x at time t , as

$$I(x, t) \approx f(x) + \delta(t) \frac{\partial f(x)}{\partial x} \tag{4}$$

Formula (4) is applicable only if the image is approximately expandable by the first-order Taylor series.

Moving on to the band-pass filtered signal. Let $B(x, t)$ be the result of applying a broad-band temporal band-pass filter to $I(x, t)$ as shown in equation (5), which later will be amplified with the factor α and added back to the original signal $I(x, t)$.

$$B(x, t) = \delta(t) \frac{\partial f(x)}{\partial x} \tag{5}$$

To amplify the band-pass filtering processed signal by the factor α , and synthesize the signal, the result of adding original and band-pass filtering processed signal, will be

$$\widehat{I}(x, t) = I(x, t) + \alpha B(x, t) \tag{6}$$

Replacing $I(x, t)$ and $B(x, t)$ with (4) and (5), we have

$$\begin{aligned} \widehat{I}(x, t) &\approx f(x) + \delta(t) \frac{\partial f(x)}{\partial x} + \alpha \delta(t) \frac{\partial f(x)}{\partial x} \\ &\approx f(x) + (1 + \alpha) \delta(t) \frac{\partial f(x)}{\partial x} \\ &\approx f(x + (1 + \alpha)\delta(t)) \end{aligned} \tag{7}$$

Through the final inference equation, we can see that the displacement $\delta(t)$ is amplified by the magnitude of $(1 + \alpha)$ via the first-order Taylor series expansion. However, in practice, the assumptions presented above have some limitations that hold for smooth images with small motions only.

$$(1 + \alpha)\delta(t) < \frac{\lambda}{8} \quad (8)$$

where the spatial wavelength, $\lambda = 2\pi/\omega$ of the moving signal of frequency ω . The bounds shown in (8) provides a guideline that gives the largest motion amplification factor, α , compatible with accurate motion magnification of a given video motion, $\delta(t)$ and image structure spatial wavelength, λ . However, in certain video scenarios detailed in [48], violating the approximation limit can be perceptually preferred.

3.2 LBP-TOP based Spatio-temporal feature extraction

LBP-TOP is a robust dynamic texture descriptor proposed in [53], which has been popularly applied to facial expressions. To consider both spatial and temporal information of the video, LBP-TOP extends the LBP in that three types of planes (XY, XT, YT) are considered instead of one plane (spatial XY) only. Given a video sequence, it can be viewed as a stack of XY, XT, YT planes along time T axis, spatial Y axis and spatial X axis respectively. The XY plane mainly reveals the spatial information, while XT and YT planes contain rich information of how pixel grayscale values transit temporally.

The computation of LBP-TOP is the same as LBP in (1) for each independent plane, i.e. in the case of LBP, one plane XY is considered only for LBP computation. Each pixel of the plane will be indexed by the local binary pattern, which computes the spatial relationship between the observed pixel and neighbor pixels. The length of the binary code is the same as the number of the neighbor pixels. Then one histogram is used to statistically compute how those binary codes are distributed. The size of the histogram is determined by the number of neighbor pixels used. For n number of neighbor pixels, the length of the histogram is 2^n . However, to reduce the number of histogram bins, uniform patterns were introduced for the LBP [25]. For example, for 8 neighbor pixels, the number of histogram bins are reduced from 256 to 59. Essentially, much of the computation of LBP-TOP is directly inherited from LBP. The difference is that the XT and YT planes have to be computed the same way as the XY plane such that three histograms are derived from the three types of planes respectively and concatenated to form one single histogram as the dynamic video texture descriptor, as shown in Fig. 3.

3.3 Classification

For classification we consider the SVM technique due to its strong mathematical foundations and high reliability in many practical applications. Depending on the number of classes that the data samples are to be from, the SVM can be divided into two problems: two-class problem and multi-class problem [34]. However, in this paper we consider the multi-class problem since the number of emotions considered is more than two, which are ‘Happiness (HAP)’, ‘Surprise (SUR)’, ‘Disgust (DIS)’, ‘Repression (REP)’, and ‘Others (OTH)’. There are two classical approaches to solve the multi-class problem: one-against-all and one-against-one. Hsu and Lin [12] reported that the latter approach is quicker and more practical than a *one-against-all* approach. In the *one-against-one* approach $k(k - 1)/2$ classifiers are constructed for k classes, with each classifier training data from two different classes in a pairwise fashion. However, the matching class can be found by majority voting

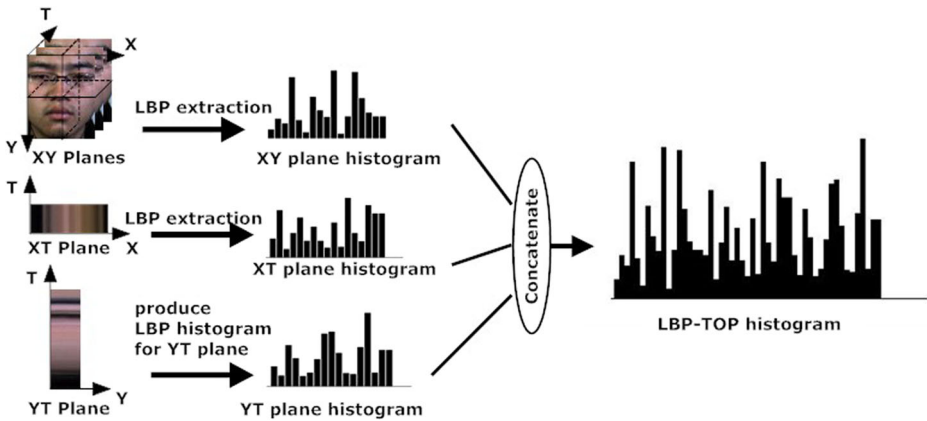


Fig. 3 The process of LBP extraction from XY, XT, and YT planes to form the LBP-TOP feature histogram

technique [12]: each binary classification performed on an input pattern contributes a vote to either class of the classifier. At the end, we predict the input pattern as belonging to the class with the largest vote.

4 Experiments

4.1 Dataset

There are only a few micro-expression databases available due to the difficulty in eliciting these micro-expressions and annotating them into relevant ground truth categories. In the aspect of acquisition, high speed recording devices are required to capture the subtle expressions since micro-expressions by definition only exist in fractions of a second. Furthermore, to be realistically useful, naturally induced expressions are favored rather than posed ones. However, how these spontaneously expressed micro-expressions are collected is tricky, as the emotional stimuli used to naturally induce a micro-expression has to be carefully designed. When all the videos are collected, to label the ground truth requires professional expertise, which can only be done by trained professionals through careful observation of the video, frame by frame. Thus, to create a spontaneous micro-expression video database is a costly effort. SMIC [16], CASME [51] and CASME II [49] are the most current micro-expression databases to the best of our knowledge. However, the SMIC dataset is much smaller in size (with only 164 samples from 16 subjects) with only 3 valid labels for recognition (positive, negative, and surprised) while the CASME is simply a preliminary subset of the newer CASME II.

Hence, we consider the most comprehensive dataset up to date – CASME II [49], which was created by the Chinese Academy of Sciences and publicly available for research use, to validate the performance of our proposed technique. There are a total of 247 videos from 26 Asian subjects with average age of 22.03 years old. All videos are labeled by two professional coders (to an acceptable reliability of 0.846) into 5 micro-expression class labels, i.e. ‘Happiness (HAP)’, ‘Surprise (SUR)’, ‘Disgust (DIS)’, ‘Repression (REP)’, and ‘Others

(OTH)’ which include other emotions such as anger, sadness and tense). To avoid flickering light from the high-speed recording, they strictly selected and placed four LED lamps under umbrella reflectors to make sure the illumination is steady and of high-intensity. The participants’ faces were captured by using a Grey GRAS-03K2C camera with resolution 640×480 at 200fps through “Raw 8” mode. The recordings were then saved in MJPEG format without any interframe compression. For further details, refer to [49]. Figure 4 shows a sample video of ‘*subject6*’, which is labeled as the ‘Surprise’ micro-expression. Visually by the naked eye, it is very difficult to spot the subtle or “hidden” changes present in the video. Hence, to tackle this issue, the amplification of these subtle changes is desirable.

4.2 Experimental results and discussions

In this section, we first discuss the parameter settings in Section 4.2.1, particularly the choice of band-pass filter and amplification factor. In Section 4.2.2, the performance of our proposed approach is reported, including a benchmark comparison against current state-of-the-art approaches. Besides, we also analyze how EVM improves the recognition capability for each micro-expression class as shown by confusion tables. The following sections further verify the effectiveness of our approach by an array of extensive tests that investigates the impact on performance under various conditions: different combination of LBP-TOP planes (Section 4.2.3), LBP-TOP spatial block partitions (Section 4.2.4), LBP-TOP feature neighborhood size and classifiers (Section 4.2.5). In all these experiments, we compared the EVM based and non-EVM based approach fairly with all other parameters and conditions held unchanged.



Fig. 4 A sample video of ‘Surprise’ class (“EP15_02”) from the CASME II dataset with subtle changes in the eye and eyebrow regions. The micro-expression exhibited by the subject is almost unnoticeable to the naked eye. Refer to animation in Online Resource 1

4.2.1 Motion magnification: filter and parameter selection

The selection of parameters for motion magnification is a difficult and manual task, but the original EVM paper [48] offers a heuristical guide that is largely empirical in nature, in order to arrive at some sufficiently good values. The authors note that “the choice of filter is generally application dependent”. For motion magnification, filters with a broad passband such as low-order infinite impulse response (IIR) filters are the preferred choice. In experiments from the original paper, second-order IIR filters are found to be capable of revealing small head movements and subtle body motions of a sleeping baby. The slow roll-off of the IIR filter introduces a broad-band that tapers off indefinitely towards zero, making it suitable for motion sweeps instead of short pulses (where ideal filters with sharp roll-offs are more suitable).

Therefore, in our work, we opt for a second-order IIR filter to best magnify the subtle motions present in facial micro-expressions. The parameters used for the ‘baby’ video in the original paper [48] are taken as our initial choice of parameters. This is intuitively chosen as the breathing motion of the baby is somewhat similar to that exhibited by subtle facial motions.

By visual inspection, the resulting output video is generally more appealing, as the effect of motion magnification is clearly visible (Fig. 5) on selected frames of the sample video (Fig. 4). At a closer look, the motion between frames, particularly in the eye region, appears more pronounced in Fig. 5 than Fig. 4. For better comparison of the same sample sequence before and after motion magnification, refer to the animations given in Online Resource 1.

In Fig. 6, we further analyze the amplified motion effect on the XT and YT planes of the same sample sequence. Although the changes on the XT plane (Fig. 6a) are not obvious, the visual difference in the texture of the motion-amplified image (on the right image) is noticeable compared to that of the original sequence (on the left image). This verifies the



Fig. 5 The sample video (from Fig. 4) after motion magnification. The motion at the eye and eyebrow regions is much more obvious now. Refer to animation in Online Resource 1

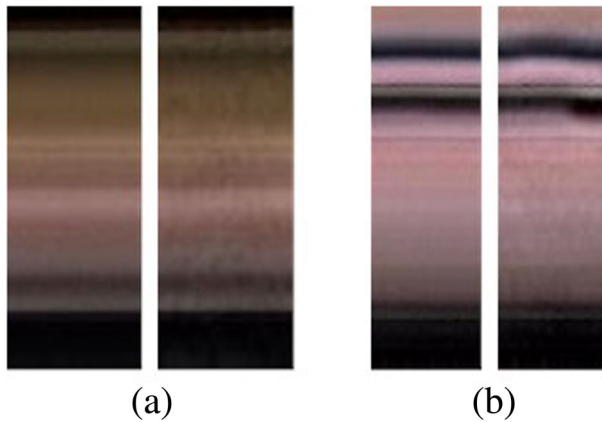


Fig. 6 **a** Visualizing the effect on the XT plane without (*left*) and with (*right*) motion magnification, and **b** visualizing the effect on the YT plane without (*left*) and with (*right*) motion magnification

augmentation of noise after magnification. The changes in the YT plane are much more obvious. From Fig. 6b, there are three dark horizontal strips. In the first and second dark strips of the motion-amplified image (on the right), we can clearly see that the vertical displacement along these strips swell more than that of the original sequence (on the left); the first two black areas correspond to the regions of eyebrows and eyes respectively.

The subsequent steps after motion magnification are as follows. We apply LBP-TOP [53] to extract spatio-temporal features from each video sequence, with 4 neighbor pixels for all three planes, using 5x5 block partitions. Videos are not normalized into equal-sized volumes (width (X), height (Y) and length (T)), since LBP-TOP does not require this for different videos. Instead, the resultant histogram is normalized, providing a statistical representation of the distribution of textures within the video. We set the radii of X, Y and T to 1, 1 and 2 respectively, which is our default setting that achieves the best result. We perform

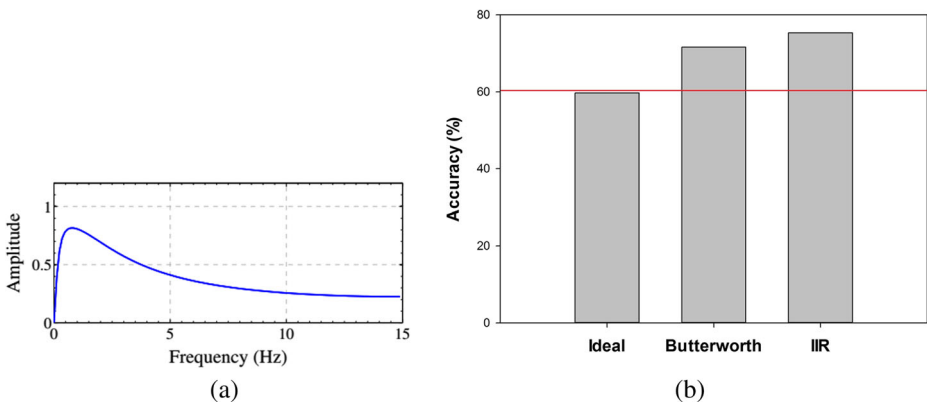


Fig. 7 **a** Second-order IIR filter ($w_l = 0.4, w_h = 3$); **b** Comparison of micro-expression recognition with EVM using different filter types (Ideal, Butterworth, IIR) with $\alpha = 10$ and $\lambda_c = 16$. The red horizontal line indicates the performance of the baseline method [49] without motion magnification

Table 1 Comparison of micro-expression recognition with EVM using second-order IIR filters of different α and λ_c values

α	λ_c	IIR filter (%)
10	16	72.47
20	16	75.30
30	16	74.49
20	26	74.49

classification of videos from the CASME II dataset on 5 micro-expression classes (*i.e.* happiness, disgust, tense, surprise, repression), using SVM classifier with a leave-one-out cross validation strategy. This is conducted in the same manner as that performed by the original authors [49].

The choice of amplification factor α is crucial, but yet sensitive. This is because as the value of α increases, the noise present in the image is also amplified. Also, a large value of α violates the bounds in (8) that sufficiently amplifies the motion to a translation of $\delta(1) = \pi/8$. With the knowledge of these parameter constraints and the author's suggestions, we experimented with various filter types and refine its related parameters to obtain the best results.

Figure 7b shows the recognition performance using the ideal, Butterworth and second-order IIR filters with magnification factor, $\alpha = 10$ and spatial frequency cutoff (wavelength) $\lambda_c = 16$ (following settings used by the 'baby' video from [48]). We note that the second-order IIR filter (7a) clearly outperforms the other two filters, thus verifying the strengths of the IIR filter mentioned earlier. As expected, the ideal filter performed the worst based on suggestions in [48] since the narrow passband is more suited for color or intensity amplification. Based on this observation, we proceed to test the recognition performance using a second-order IIR filter with various α and λ_c settings, as shown in Table 1. By first constraining λ_c while increasing α , it can be observed that the accuracy rates start to drop when α increases beyond 20. Meanwhile, the value of λ_c does not appear to influence the magnification output.

Table 2 Confusion tables of the proposed and baseline methods using SVM classifier with linear kernel (LOSOCV protocol)

Expression	OTH	DIS	HAP	SUR	REP
Proposed method (with EVM)					
OTH	0.735	0.147	0.059	0.020	0.039
DIS	0.433	0.433	0.133	0.000	0.000
HAP	0.182	0.091	0.606	0.000	0.121
SUR	0.320	0.200	0.040	0.400	0.040
REP	0.370	0.074	0.333	0.000	0.222
Baseline (without EVM)					
OTH	0.559	0.245	0.137	0.020	0.039
DIS	0.467	0.283	0.167	0.050	0.033
HAP	0.303	0.030	0.394	0.061	0.212
SUR	0.240	0.080	0.240	0.320	0.120
REP	0.407	0.037	0.444	0.074	0.037

Table 3 Confusion tables of the proposed and baseline methods using SVM classifier with RBF kernel (LOSOCV protocol)

Expression	OTH	DIS	HAP	SUR	REP
Proposed method (with EVM)					
OTH	0.735	0.167	0.049	0.010	0.039
DIS	0.550	0.367	0.083	0.000	0.000
HAP	0.212	0.091	0.576	0.000	0.121
SUR	0.400	0.200	0.040	0.320	0.040
REP	0.407	0.037	0.370	0.000	0.185
Baseline (without EVM)					
OTH	0.784	0.176	0.000	0.020	0.020
DIS	0.850	0.083	0.067	0.000	0.000
HAP	0.727	0.121	0.091	0.000	0.061
SUR	0.560	0.360	0.080	0.000	0.000
REP	0.704	0.037	0.222	0.037	0.000

4.2.2 Performance of proposed technique

With a sufficiently good parameters ($\alpha = 20$ and $\lambda_c = 16$), we achieve the best recognition accuracy of 75.30 %. Despite the empirical nature of determining these parameters, it is important to highlight that all recognition results in Table 1 are significantly better than the baseline result of 60.32 % without motion magnification.¹ This singles out the importance of amplifying subtle motions found in facial micro-expressions for the purpose of recognition.

For a closer look at the performance of individual expression classes, we compare the confusion tables of the proposed method (with EVM) to that of the baseline (without EVM) with respect to two different kernel functions (linear, RBF) of the SVM classifier. Tables 2 and 3 compares the confusion tables with and without EVM based on the linear and RBF kernel respectively by using leave-one-subject-out² cross validation protocol. This is arguably less bias for datasets with an imbalanced distribution of samples across subjects [15], providing a fairer measure of class-specific performance. Overall, the proposed technique greatly improves the recognition accuracy for each micro-expression class. Using SVM linear kernel (Table 2), there are marked improvements of up to 50 % for the ‘OTH’, ‘DIS’, ‘HAP’ and ‘SUR’ classes, while the ‘REP’ class is five-fold better than the baseline performance. A similar trend is observed when the RBF kernel is used instead. In Table 3, we see that the performance of four of five classes (‘DIS’, ‘HAP’, ‘SUR’, ‘REP’) vastly improved using the proposed technique, with the exception of the ‘OTH’ class which dropped slightly. At this juncture, it is important to first point out that the RBF kernel seems to be inappropriate for the baseline method as it tends to over-fit the ‘OTH’ class. This is obvious from the high number of false positives matching to this class. We postulate that with motion magnification, the samples become more distinguishable, providing larger

¹Original paper [49] reported the best possible result of 63.41 % using different set of settings.

²Leave-one-subject-out cross validation (LOSOCV) is a subject-independent evaluation where the videos of one object are held out as testing set while the remaining videos form the training set. This is then repeated for all subjects in dataset, and the average accuracy rate across all folds is taken.

Table 4 Benchmarking against current state-of-the-art approaches for micro-expression recognition on the CASME II dataset

Methods	Recognition rate (%)
Baseline [49]	63.41
OSW [19]	66.40
LBP-MOP [47]	66.80
LBP-SIP [46]	67.21
Adaptive MM [31]	69.63
Proposed method	75.30

between-class margins within the underlying distribution, hence reducing the effect of over-fitting. This observation of the effect of applying motion magnification was also noted by Park et al. [31].

Our best result reported in this paper (75.30 %) uses the RBF kernel for SVM classification. Table 4 summarizes how our proposed method fare against current state-of-the-art works on the benchmark CASME II dataset. The recent work of Park et al. [31] is much similar in nature to our approach. Although they have proposed an adaptive method for motion magnification which predicts the frequency band using intensity variation features, our approach is still superior in terms of recognition capability. Our conjecture is that our method simply uses a fixed bandwidth across all samples instead of relying entirely on the predictor to return a frequency band range (which was ad-hocly fixed as well), for each different test sample.

With EVM, the subtle expressions that are hardly detected by naked eyes now are more visually notable as shown in Fig. 5. Besides, more discriminative features are extracted with useful motion signals amplified. However, the limitation of this technique exists as well. With amplifying the expected signal, the noise signal might be also magnified which will interfere the extracted feature with less discriminative. To decrease this limitation, we can either improve the filter for filtering out noise signals or increase the resolution of our recorded videos.

4.2.3 Analysis on LBP-TOP plane selection

To further show the effectiveness of our proposed method for micro-expression recognition, we tested it on individual LBP-TOP planes and their possible combinations. The results showed the effectiveness of EVM on all possible planes combinations.

Table 5 Comparison of micro-expression recognition accuracy with different combination of LBP-TOP planes (5x5 block partitioning), classified using SVM with linear kernel

	Baseline (%)	Proposed method (%)
XY	55.47	68.83
XT	51.42	63.97
YT	55.47	72.87
XY-XT	58.70	75.71
XY-YT	59.51	74.09
XT-YT	59.92	73.68
XY-XT-YT	60.32	74.90

Table 6 Comparison of micro-expression recognition accuracy with different combination of LBP-TOP planes (5x5 block partitioning), classified using SVM with RBF kernel

	Baseline(%)	Proposed method(%)
XY	55.06	69.64
XT	53.04	65.99
YT	60.32	72.47
XY-XT	59.41	75.30
XY-YT	59.92	74.09
XT-YT	59.51	75.30
XY-XT-YT	59.11	75.30

Tables 5 and 6 reports the recognition accuracy on different feature planes (XY, XT, YT) of LBP-TOP and their combinations, using SVM with linear and RBF kernel respectively. This analysis assumes a 5x5 block partitioning for LBP-TOP.

Firstly, we observe that even with motion magnification, the classification accuracy when only two planes are used (XY-XT, XY-YT, XT-YT), does not differ much compared to when all three planes (XY-XT-YT) are used. This strengthens the results in [53] that using two planes is comparable to using all three planes, albeit with some insignificant decrease of accuracy in most cases. More importantly, Table 5 shows that the recognition accuracy of our approach is significantly higher than that of the baseline method. In fact, the increase in accuracy is consistently above 10 % regardless of the type or combination of planes used. Likewise, the same experiment conducted using SVM classifier with RBF kernel (shown in Table 6) also yielded similar conclusions with regards to the use of motion magnification. The improvement observed with the RBF kernel is slightly more pronounced, averaging (across all plane combinations) an increase of 12.45 % compared to 12.15 % for the linear kernel.

In short, the recognition performance of our proposed technique is consistently better than that of the baseline technique on different types (and combinations) of LBP-TOP planes (XY, XT, YT).

4.2.4 Analysis on feature block partitions

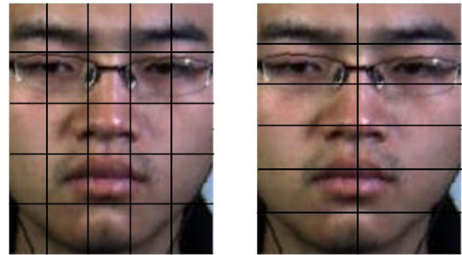
Similarly, we also examine the performance of micro-expression recognition with respect to different block partition sizes to further evaluate the robustness of our proposed method. Block partitioning is a method of representation in which the image is divided into several non-overlapping or overlapping blocks to obtain region-concatenated descriptors. This was

Table 7 Comparison of micro-expression recognition using different $n \times 2$ symmetrical block partitioning

	Baseline(%)	Proposed method(%)
5x2	55.87	64.78
6x2	58.70	65.18
7x2	57.89	66.40
8x2	53.44	62.35

All three LBP-TOP planes are used while linear kernel is used for SVM

Fig. 8 Illustration of the 5×5 (left) and 6×2 (right) block partitions



primarily employed in [53] for the LBP-TOP feature descriptor. For clarity, we consider the division of an image on a certain plane into $n \times m$ blocks; n rows and m columns. With the exception of this analysis, we fixed the block partitioning for all of our experiments to 5×5 blocks, in accordance with that chosen in the original baseline work [49].

Table 7 shows the recognition performance (SVM linear kernel) of a variety of symmetrical block partitions of the $n \times 2$ form. The reason why the number of columns is fixed to 2 is that the face is to some extent, symmetric in nature. Using such a partition, the facial components *e.g.* eyes, eyebrows, etc. could be preserved as a whole in one single block compared to that of $n \times n$ partitions (see the illustration of partitions in Fig. 8), which could divide the image into two or more blocks, both row-wise and column-wise. Interestingly, from the experimental results (in Table 7), we see that the 6×2 partition (baseline) achieved an accuracy of 58.7 % compared to the 5×5 partition (baseline) that managed a 60.32 % accuracy. Although the partition of 6×2 performs slightly worse than the 5×5 partition, the feature dimension drops dramatically by more than half, from $(16 \times 3 \times 25) = 1200$ for 5×5 blocks to $(16 \times 3 \times 12) = 576$ for 6×2 blocks. This reduction of feature dimensionality comes at the expense of recognition accuracy.

Furthermore, we also investigated the impact of block partitioning on the recognition accuracy with RBF kernel for SVM classification (shown in Table 8). We observe that the best performance of the baseline method is 59.92 % with 6×2 partition compared to 59.11 % with 5×5 partitions (reported in Table 6). Contrary to that of the linear kernel, the 6×2 partition performs marginally better than the 5×5 partition on the RBF kernel, it does not appear to be statistically significant.

Nevertheless, it turns out that regardless of the choice of block partitioning and SVM kernel used, the role of EVM is crucial for better recognition of facial micro-expressions.

Table 8 Comparison of micro-expression recognition using different $n \times 2$ symmetrical block partitioning

	Baseline (%)	Proposed method (%)	
5x2	57.09	68.02	
6x2	59.92	65.59	
All three LBP-TOP planes are used while RBF kernel is used for SVM	7x2	59.11	66.80
	8x2	57.09	64.78

Table 9 Micro-expression recognition performance using kNN and SVM classifiers with different kernels, with and without EVM, based on 5x5 blocks partition

Classifier	Baseline (%)	Proposed method (%)
KNN (k=5)	52.03	61.79
SVM (linear kernel)	60.32	74.90
SVM (polynomial kernel, 5th order)	60.73	72.87
SVM (RBF kernel))	59.11	75.30

4.2.5 Analysis on classifiers and feature neighborhood

In our final analysis, we further compare the recognition performance of the proposed method against the baseline in Table 9, classified by the k-nearest neighbor (kNN)³ and SVM with different kernels. We see that across all classifiers, the recognition performance of our proposed technique is superior to that of the baseline method [49]. Our best result, the SVM classifier with RBF kernel posed an improvement of more than 14 % over the baseline. As expected, the SVM is a more effective classifier than the kNN owing to its maximum margin properties that is able to separate high-dimensional data.

We also changed the number of neighbor pixels that construct the LBP-TOP feature pattern to 8 for all three planes. This resulted in a baseline recognition accuracy of 62.35 % and 69.23 % for the proposed method. Hence, the positive effect of motion magnification on facial micro-expressions remains telling.

4.2.6 Summary

Overall, the in-depth experimentation presented in this section demonstrated the reliability and robustness of our proposed micro-expression recognition approach by amplifying subtle facial motions for better feature representation. We uncovered the following points:

- The second order IIR band-pass filter performed best for magnifying facial micro-expressions, with the most visually obvious results.
- The extracted LBP-TOP features were more discriminative on EVM-magnified frames.
- EVM is found to be effective on all possible combination of LBP-TOP feature planes (XY, XT, YT).
- EVM is also found to be effective on various LBP-TOP spatial block partitions and neighborhood sizes.

5 Conclusion and future work

In this paper, we introduced a new approach that incorporates video motion magnification to accentuate subtle motions in facial micro-expressions, resulting in a greatly improved recognition performance. In our extensive experiments, we demonstrate that the proposed approach which adopts Eulerian Video Magnification (EVM) consistently performs better

³We empirically found that $k = 5$ yielded the best possible result for kNN.

than baseline and current state-of-the-art methods. This was also noticeable across various conditions; notably, different EVM parameters, LBP-TOP with different combination of planes, different frame partitions, different SVM kernels as well as different neighboring pixels for LBP-TOP feature. Hence, we elucidate in this paper, the vital role of motion magnification in amplifying subtle micro-expression changes in video. This uncovers a promising new direction towards effective machine recognition of facial micro-expressions.

With this framework in place, there are many future directions for further improvements. One possible work is to make use of facial landmarks of Action Units (AU) to improve overall recognition accuracy; or in that very sense, attempt to magnify only selective regions that correspond to these landmarks instead of the entire face area. Also, LBP-based feature descriptors have their intrinsic limitations. Thus, the introduction of more robust spatio-temporal texture descriptors may potentially unlock further discrimination between the micro-expressions.

Acknowledgments This work is supported by Zhejiang Provincial Natural Science Foundation of China (Grant Nos. LQ17F020002 and LQ14F020006), TM Grant under project UbeAware and 2beAware, and MOHE Grant FRGS/1/2016/ICT02/MMU/02/2. The authors would like to thank the Chinese Academy of Sciences for the CASME II micro-expression database and Su-Jing Wang for providing more details of their CASME II work in [49].

References

1. Ahonen T, Matas J, He C, Pietikainen M (2009) Rotation invariant image description with local binary pattern histogram fourier. In: Image analysis, pp 61–70
2. Azmi R, Yegane S (2012) Facial expression recognition in the presence of occlusion using local gabor binary patterns. In: Iranian Conf. on Elect Eng. (ICEE), pp 742–747
3. Bartlett MS, Littlewort GC, Frank MG, Lainscsek C, Fasel IR, Movellan JR (2006) Automatic recognition of facial actions in spontaneous expressions. *J Multimed* 1(6):22–35
4. Bashyal S, Venayagamoorthy GK (2008) Recognition of facial expressions using gabor wavelets and learning vector quantization. *Eng Appl Artif Intell* 21(7):1056–1064
5. Belhumeur P, Hespanha J, Kriegman D (1997) Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Trans Pattern Anal Mach Intell* 19(7):711–720
6. Burt P, Adelson E (1983) The laplacian pyramid as a compact image code. *IEEE Trans Commun* 31(4):532–540
7. Chang CC, Lin CJ (2011) A library for support vector machines. *ACM Trans Intell Syst Tech*:1–39
8. Ekman P (2009) Lie catching and microexpressions. *Philos Deception*:118–133
9. Fehr J, Burkhardt H (2008) 3D rotation invariant local binary patterns. *ICPR* 2008:1–4
10. Frank M, Herbasz M, Sinuk K, Keller A, Nolan C (2009) I see how you feel: training laypeople and professionals to recognize fleeting emotions, presented at the Annual Meeting of the International Communication Association, New York Unpublished manuscript
11. Heikkilä M, Pietikainen M, Schmid C (2009) Description of interest regions with center-symmetric local binary patterns. *Pattern Recognit* 42(3):425–436
12. Hsu CW, Lin CJ (2002) A comparison of methods for multiclass support vector machines. *IEEE Trans Neural Netw* 13(2):415–425
13. Huang D, Shan C, Ardabilian M, Wang Y, Chen L (2011) Local binary pattern and its application to facial image analysis: a survey. *IEEE Trans Syst Man Cybern Part C Appl Rev* 41(6):765–781
14. Kanade T, Cohn JF, Tian Y (2000) Comprehensive database for facial expression analysis. In: Proceedings Fourth IEEE international conference on automatic face and gesture recognition, 2000. IEEE, pp 46–53
15. Le Ngo AC, Phan RCW, See J (2014) Spontaneous subtle expression recognition: imbalanced databases and solutions. In: Asian conference on computer vision, pp 33–48
16. Li X, Pfister T, Huang X, Zhao G, Pietikainen M (2013) A spontaneous micro-expression database: inducement, collection and baseline. In: IEEE Int. Conf. and workshops on automatic face and gesture recognition (FG), 10th, pp 1–6

17. Liao S, Chung AC (2007) Face recognition by using elongated local binary patterns with average maximum distance gradient magnitude. In: Asian conf. on computer vision (ACCV), pp 672–679
18. Liong ST, Phan RCW, See J, Oh YH, Wong K (2014) Optical strain based recognition of subtle emotions. In: 2014 international symposium on Intelligent signal processing and communication systems (ISPACS). IEEE, pp 180–184
19. Liong ST, See J, Phan RCW, Le Ngo A, Oh YH, Wong K (2014) Subtle expression recognition using optical strain weighted features. In: 1st workshop on computer vision for affective computing, asian conf. on computer vision (ACCV), pp 644–657
20. Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z, Matthews I (2010) The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression. In: IEEE conference on computer vision and pattern recognition workshops (CVPRW), pp 94–101
21. Lyons M, Akamatsu S, Kamachi M, Gyoba J (1998) Coding facial expressions with gabor wavelets. In: Proceedings Third IEEE international conference on automatic face and gesture recognition, 1998. IEEE, pp 200–205
22. McDuff D, El Kaliouby R, Senechal T, Amr M, Cohn JF, Picard R (2013) Affectiva-MIT facial expression dataset (AM-FED): naturalistic and spontaneous facial expressions collected “in-the-wild”. In: 2013 IEEE conference on computer vision and pattern recognition workshops (CVPRW), pp 881–888
23. Nguyen HT, Caplier A (2013) Elliptical local binary patterns for face recognition. In: Computer vision-ACCV 2012 workshops, pp 85–96
24. Ojala T, Pietikainen M, Harwood D (1996) A comparative study of texture measures with classification on featured distribution. *Pattern Recogn* 29:51–59
25. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans Pattern Anal Mach Intell* 24(7):971–987
26. Oliveira L, Mansano M, Koerich A, De Souza Britto Jr A (2011) 2d principal component analysis for face and facial-expression recognition. *Comput Sci Eng* 13(3):9–13
27. Oliveira LS, Sabourin R, Bortolozzi F, Suen CY (2003) A methodology for feature selection using multiobjective genetic algorithms for handwritten digit string recognition. *Int J Patt Recog Artif Intell* 17(06):903–929
28. Pantic M, Valstar M, Rademaker R, Maat L (2005) Web-based database for facial expression analysis. In: IEEE international conference on multimedia and expo, 2005. ICME 2005. IEEE, p 5
29. Park S, Kim D (2009) Subtle facial expression recognition using motion magnification. *Pattern Recogn Lett* 30(7):708–716
30. Park S, Lee H, Shin J, Kim D (2008) The POSTECH subtle facial expression database 2007 (SFED07). In: Proceedings of the 8th POSTECH-KYUTECH joint workshop on neuroinformatics, pp 653–654
31. Park SY, Lee SH, Ro YM (2015) Subtle facial expression recognition using adaptive magnification of discriminative facial motion. In: Proceedings of the 23rd annual ACM conference on multimedia, pp 911–914
32. Pfister T, Li X, Zhao G, Pietikainen M (2011) Recognising spontaneous facial micro-expressions. In: 2011 IEEE international conference on computer vision (ICCV). IEEE, pp 1449–1456
33. Polikovskiy S, Kameda Y, Ohta Y (2009) Facial micro-expressions recognition using high speed camera and 3d-gradient descriptor. In: 3rd international conference on crime detection and prevention (ICDP 2009). IET, pp 1–6
34. Rahulamathavan Y, Phan R, Veluru S, Cumanan K, Rajarajan M (2013) Privacy-preserving multi-class support vector machine for outsourcing the data classification in cloud. *IEEE Trans Dependable and Secure Computing* 99:467–479
35. Rahulamathavan Y, Phan RW, Chambers J, Parish D (2013) Facial expression recognition in the encrypted domain based on local fisher discriminant analysis. *IEEE Trans on Affect Comput* 4(1):83–92
36. Russell TA, Chu E, Phillips ML (2006) A pilot study to investigate the effectiveness of emotion recognition remediation in schizophrenia using the micro-expression training tool. *Br J Clin Psychol* 45(4):579–583
37. Sarnarawickrame K, Mindya S (2013) Facial expression recognition using active shape models and support vector machines. In: 2013 International Conference on advances in ICT for emerging regions (ICTer), pp 51–55
38. Satiyan M, Nagarajan R (2010) Recognition of facial expression using haar-like feature extraction method. In: 2010 International Conference on intelligent and advanced systems (ICIAS). IEEE, pp 1–4
39. Schopler E, Mesibov GB (1998) Kunc. L.J., Asperger syndrome or high-functioning autism?
40. Shan C, Gong S, McOwan PW (2009) Facial expression recognition based on local binary patterns: a comprehensive study. *Image Vis Comput* 27(6):803–816

41. Shreve M, Godavarthy S, Goldgof D, Sarkar S (2011) Macro- and micro-expression spotting in long videos using spatio-temporal strain. In: IEEE International conference on automatic face & gesture recognition and workshops (FG 2011), pp 51–56
42. Shreve M, Godavarthy S, Manohar V, Goldgof D, Sarkar S (2009) Towards macro- and micro-expression spotting in video using strain patterns. In: Workshop on applications of computer vision (WACV), pp 1–6
43. Tan X, Triggs B (2010) Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans Image Process* 19:1635–1650
44. Uddin MZ, Lee J, Kim TS (2009) An enhanced independent component-based human facial expression recognition from video. *IEEE Trans. on Consum Electron* 55(4):2216–2224
45. Wang X, Han TX, Yan S (2009) An hog-lbp human detector with partial occlusion handling. In: 2009 IEEE 12th international conference on computer vision, pp 32–39
46. Wang Y, See J, Phan RCW, Oh YH (2014) Lbp with six intersection points: Reducing redundant information in lbp-top for micro-expression recognition. In: Asian conference on computer vision (ACCV), pp 525–537
47. Wang Y, See J, Phan RCW, Oh YH (2015) Efficient spatio-temporal local binary patterns for spontaneous facial micro-expression recognition. *PLoS One* 10:e0124674
48. Wu HY, Rubinstein M, Shih E, Guttag J, Durand F, Freeman W (2012) Eulerian video magnification for revealing subtle changes in the world. *ACM Trans Graph (TOG)* 31(4):65:1–65:8
49. Yan WJ, Wang SJ, Zhao G, Li X, Liu YJ, Chen YH, Fu X (2014) Casme ii: an improved spontaneous micro-expression database and the baseline evaluation. *PLoS ONE* 9:e86041
50. Yan WJ, Wu Q, Liang J, Chen YH, Fu X (2013) How fast are the leaked facial expressions: the duration of micro-expressions. *J Nonverbal Behav* 37(4):217–230
51. Yan WJ, Wu Q, Liu YJ, Wang SJ (2013) Casme database: a dataset of spontaneous micro-expressions collected from neutralized faces. In: 2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG), pp 1–7
52. Yin L, Wei X, Sun Y, Wang J, Rosato MJ (2006) A 3d facial expression database for facial behavior research. In: 7th international conference on automatic face and gesture recognition, pp 211–216
53. Zhao G, Pietikainen M (2007) Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Trans Pattern Anal Mach Intell* 29(6):915–928



Yandan Wang received her Ph.D in Information Technology from Monash University, Malaysia. She is currently working as a lecturer at Wenzhou University, China. Her research interests are in the area of multimedia indexing and retrieval, machine vision and pattern recognition.



John See received his PhD in Computer Science, MEngSc and BEng degrees from Multimedia University, Malaysia. He is currently working as a Senior Lecturer and leader of the Pattern Recognition sub-cluster of the Centre for Visual Computing at Multimedia University, Malaysia. His research interests is in the area of computer vision and pattern recognition, particularly in the domain of video-based biometrics, surveillance and aesthetics.



Yee-Hui Oh received his B.Eng. degree in Multimedia University, Malaysia. He is currently pursuing his Master degree in Engineering at Multimedia University, Malaysia.



Raphael C.-W. Phan received his B.Eng. (Hons), M.Eng.Sc. and Ph.D degrees from Multimedia University, Malaysia. He is currently working as a professor in Faculty of Engineering and director for University-Industry Engagement at Multimedia University, Malaysia. His research interests include security & cryptography, emotion recognition, secure signal processing, game theoretic security and/or privacy, critical infrastructure protection, forensics, surveillance & machine vision and resilience.



Yogachandran Rahulamathavan received the PhD degree in Signal Processing from Loughborough University, UK in 2011. He is currently working as a Research Fellow with the Information Security Group at City University London, UK. His research interests include signal processing, machine learning and information security and privacy.



Huo-Chong Ling received his B.Eng. (Hons), M.Eng.Sc. and Ph.D. degrees from Multimedia University, Malaysia. He is currently a senior lecturer in the Department of Electrical and Computer Engineering, Faculty of Engineering and Science, Curtin University, Sarawak, Malaysia. His research interests include steganography, watermarking, access control system and micro-expression.



Su-Wei Tan receives his B.Eng. (Hons), M.Eng.Sc. Ph.D from University Malaya, Multimedia University, Malaysia and university of Kent, UK respectively. He is currently a associate professor in Faculty of Engineering, Multimedia University, Cyberjaya, Malaysia. His research interests include application/network layer multicast, peer-to-peer, traffic engineering, routing protocols, wireless networking and micro-expression recognition.



Xujie Li is a lecturer at Intelligent Information Systems Institute, Wenzhou University, Wenzhou, China. He received his BEng degree in communication engineering from Author Biographies Click here to download Author Biographies biography.docx Ningbo University in 2005 and his MEng degree in communications and information systems from Ningbo University in 2008. His research interests include computer graphics, image processing and GPU parallel computing.