GA-APEXNET: GENETIC ALGORITHM IN APEX FRAME NETWORK FOR MICRO-EXPRESSION RECOGNITION SYSTEM

Qiu-Shi Jin¹, Huang-Chao Xu¹, Kun-Hong Liu¹, Sze-Teng Liong², Y. S. Gan³, Shu-Wen Su¹

¹ Xiamen University, School of Informatics, Xiamen University, Xiamen, China
² Feng Chia University, Department of Electronic Engineering, Feng Chia University, Taichung, Taiwan
³ National Taipei University of Nursing and Health Sciences, Taipei, Taiwan

ABSTRACT This paper introduces a novel method to recognize the facial micro-expressions by directly adopting state-of-the-arts deep learning algorithms. Specifically, genetic algorithm (GA) is applied based on the principles of evolution in searching an optimal solution to generate discriminant features. Prior to that, the feature vectors of each video are first encoded using existing apex-based feature descriptors, viz, Bi-WOOF, OFF-ApexNet and STSTNet. Then, GA is utilized to enrich the features by eliminating irrelevant information that do not contribute to the expression prediction. It is acknowledged that the overfitting phenomena can be avoided in the feature selection process, as GA employs tournament selection and deterministic mutation procedures during the evolution procedure. As a result, Genetic Algorithm in Apex Frame Network (GA-ApexNet) is introduced and significant improvement of the recognition performance has been achieved. A standard experimental configuration is designed to evaluate the robustness of the proposed framework. Following the suggestion in Facial Micro-Expressions Grand Challenge (MEGC 2019), the UF1 and UAR obtained are 78.85% and 78.28%, for the composite dataset that comprises CASME II, SMIC and SAMM. We also note that this is the first work that fuses GA in micro-expression recognition system.

1. INTRODUCTION

The research in automatic facial micro-expression (ME) recognition has received great attention in these recent five years from computer vision and machine learning field. However, recognizing ME is always a challenging task and difficult to realize in realistic scenario, due to its subtlety and low intensity characteristics. Concisely, ME normally occurs within 0.04s to 0.2s on the face [1]. ME is involuntary muscular reactions to psychological reactions and thus it potentially reveals a person’s genuine or concealed feeling. In contrast, macro-expression, or also known as the normal facial expression, usually sustains longer compared to that of ME. The common definition of the duration of macro-expression is between 2 〜 3s [2]. Thus, the appearance is always noticeable and obvious.
However, negative emotions can be easily hidden behind a fake smile. Detecting and understanding one’s real emotional state is important as it can be deployed in many applications, such as public surveillance, social communication, clinical diagnosis, advertising and marketing.

So far, the detection and recognition of ME are still performed manually by psychologists or trained experts. There is no commercial products or applications for ME analysis system. This is because of the four main reasons: (1) Lack of ME database - since the ME is uncontrollable and not all the expected ME motion can be captured for everyone by watching the same stimuli (i.e., a triggering film) [3]; (2) Professional video acquisition environment setup - a quiet and noise controlled laboratory is designed to create a flicker-free lighting condition during the video recording; (3) High cost ground-truth labeling - to provide an empirical evidence and desired output for the supervised machine learning algorithm; (4) Data redundancy - the ME videos are elicited using a high frame rate camera (i.e., 100~200fps), thus there will be many frames that shows similar or even same face pose. The repeated frames may confuse the learning algorithm and lead to poor classification accuracy.

The first spontaneous ME database was established in 2011 [3], released by University of Oulu. However, there were six participants' data collected and comprised only of 77 videos. Two years later, the same research group publishes an improved dataset [4] that contains more participants (i.e., 16 participants) that consists of 164 videos. Since then, it has been successful in attracting several researchers from different countries to join this automatic ME analysis and investigation. More recent databases released include CASME /CASME II/ CAS(ME)2 [5, 6, 7] from the Chinese Academy of Sciences, as well as SAMM from Manchester Metropolitan University [8]. Note that, all of the aforementioned databases do not have uniform emotion states and the camera used are different. Among these publicly available databases, the number of participants varies from 5 to 28, whereas the number of videos ranges from 77 to 257. The details of the datasets are summarized in Table 1.

Table 1. Detailed information of the three merged ME databases

| Database     | SMIC [4] | CASME II [6] | SAMM [9] |
|--------------|----------|--------------|----------|
| Year         | 2013     | 2014         | 2018     |
| Subject      | 16       | 24           | 28       |
| Sample       | 164      | 145          | 133      |
| Frame rate (fps) | 100     | 200          | 200      |

| Camera (brand) | Point Grey GRAS-03K2C | PixelINK PixelINKPL-B774 | Basler Ace acA2000-340km |
| Camera (model) |                        |                         |                        |
| Cropped image resolution | 170×140 | 170×140 | 170×140 |
| Frame number | Average | 34       | 70      | 73     |
|              | Maximum | 58       | 126     | 101    |
|              | Minimum | 11       | 24      | 30     |
| Video duration (s) | Average | 0.34     | 0.35    | 0.36   |
|                  | Maximum | 0.58     | 0.63    | 0.51   |
|                  | Minimum | 0.11     | 0.12    | 0.15   |
| Expression | Negative | 70       | 88      | 92     |
|             | Positive | 51       | 32      | 26     |
|             | Surprise | 43       | 25      | 15     |
| Ground-truth annotations | Onset index | ✓        | ✓       | ✓       |
|                         | Offset index | ✓        | ✓       | ✓       |
|                         | Apex index | ×        | ✓       | ✓       |
|                         | Action unit | ×        | ✓       | ✓       |

Recently, the Second Facial Micro-Expression Grand Challenge (MEGC 2019) [10] proposed to combine the videos from different into databases into one single composite dataset. Concretely, three
databases are selected for the merging, viz, SMIC, CASME II and SAMM. This is to create a real-world scenario where the subjects are from the diverse backgrounds (ethnicity, gender and culture). By doing so, the composite dataset has been summarized into 442 videos from 68 subjects and comprises three emotion states: positive, negative and surprise. The primary goal of this challenge is to establish a standard for all the new proposed algorithms to evaluate at the same scale, at the meantime to promote ME recognition technology.

In brief, a ME recognition system incorporates three stages: image preprocessing, feature extraction and emotion classification. Particularly for the preprocessing step, the common techniques performed are: (a) facial landmark detection; (b) facial alignment; (c) face subregion selection, and; (d) frame selection. Facial landmark detection is an essential process for facial analysis task. It is to localize the coordinates points of the salient regions, such as: eyes, eyebrow, nose, mouth and face boundary. Popular landmark detection techniques that are utilized in ME analysis include Active Shape Model (ASM) [11], Discriminative Response Maps Fitting (DRMF) [12], Subspace Constrained Mean-Shifts (SCMS) [13], Face++ [14] and Constraint Local Model (CLM) [15]. Conventionally, after obtaining the facial landmark points, it is followed by the alignment step where the faces are canonically normalized based on translation, scale, and rotation. Examples of the method used are Local Weighted Mean (LWM) [16] and Procrustes analysis [17].

Some works had been demonstrated that focusing the features of part of the face is more effective in contributing to higher recognition performance, instead of considering the entire face. This is intuitive as ME usually occurs on certain parts of the face. Therefore, removing the regions that are irrelevant to ME information is preferable. [18] and [19] suggest to select a specific set of facial parts, viz, 16 regions and 3 regions, respectively. To further reduce the redundancy of the input data and suppress the feature dimensionality, several papers propose to select several frames for feature extraction. This is because many similar facial motion patterns will be appeared in multiple consecutive frames, because the frame rate of the high speed camera is up to 200fps.

Succinctly, there are two approaches for the frames selection: sequence-based and apex-based. Sequence-based is to determine several frames to represent the entire video. For instance, the method are Sparsity-promoting Dynamic Mode Decomposition (DMDSP) [20] used by Le Ngo et al [21], Temporal Interpolation Model (TIM) [22] used by Xu et al. [23] and Principal Component Analysis (PCA) used by Wang et al. [18]. On other hand, the apex-based approach utilizes only two frames from the video, viz, onset and apex. An onset frame indicates the beginning of the ME undergoing muscle contraction, whereas the apex frame shows the maximum motion intensity as it is the instant that the ME reaches its climax. This apex-based approach has been adopted by the top four research groups [24, 25, 26, 27] participated in MEGC 2019 and the benchmark ME recognition results are UF1=78.85% and UAR=78.24%.

Subsequent step after the preprocessing step is feature extraction. Most ME works utilize Local Binary Pattern (LBP) [28] and optical flow (OF) [29] families in the feature encoding. In short, LBP is a simple yet powerful visual descriptor that represent the features by comparing each pixel intensity with its surrounding neighborhood. Over the years, a variety of modified LBP methods have been introduced, such as Local Binary Patterns with Six Intersection Points (LBP-SIP) [30], Completed Local Quantization Pattern (STCLQP) [31] and Discriminative Spatiotemporal Local Binary Pattern with Revisited Integral Projection (DiSTLBP-RIP) [32]. As for the OF method, it is to capture the apparent motion of the objects between two frames. In ME recognition system, OF normally used to collocate with the apex-based preprocessing method. Specifically, it describes the motion between the onset and apex frames, and hence utilizes the motion patterns to feed into the machine learning classifier as the input data. Thus far, there are many works [33] [34][35] employ this kind of methods combination (i.e., apex-based preprocessing + OF) in ME analysis mechanism.

Based on some of the observations that have been described above, this paper aims to introduce an alternative solution to improve the ME recognition performance. In this paper, genetic algorithm (GA) [36] is applied to enhance the features extracted by adopting metaheuristic search and optimization concepts. GA is a classical solution that continually evolve the features by inheriting the superior
characteristics of the previous generation. Thus, meaningful features that can lead to high classification accuracy will be remained, while the redundant features will be eliminated.

The objective of this paper is to present an optimal design of GA algorithm that can be applied on the state-of-the-art approaches to boost the ME recognition performance. The three objectives are explicitly stated as follows:

1. Extension of some of the existing feature descriptors by adopting the genetic algorithm.
2. Evaluation of the proposed approach on the standard composite dataset that composes three recent spontaneous micro-expression databases.
3. Graphical and statistical trend analysis to highlight the effectiveness of the proposed method.

2. PROPOSED ALGORITHM
The proposed approach consists of seven steps: (1) Face landmark detection and alignment - to standardize the position of the facial fiducial coordinates, by transforming and mapping each landmark points to a reference face; (2) Apex frame estimation - to obtain the indices of the apex, especially for those database that do not provide this groundtruth information; (3) OF components computation - to describe the motion patterns between the onset and apex frames, which represents the features in the spatio-temporal perspective. (4) OF features concatenation - to concatenate the OF derived component along the feature dimension and produce a 3D feature; (5) Feature extraction - three existing feature descriptors are utilized to extract the expression features; (6) Feature enhancement using GA - introduction of GA to enrich the features encoded through evolutionary biological process. (7) Expression classification - to distinguish the emotion class based on the features extracted.

![Figure 1. The flow chart of our proposed micro-expression recognition framework.](image)

2.1 Face Landmark Detection and Alignment
The original papers of the two micro-expression databases [4, 6] have gone through the face registration and alignment before releasing the databases publicly. Therefore, in order to improve data set consistency, we follow their suggestions in standardizing all faces by rescaling all the faces' landmark points based on a template face. In brief, the datasets in our experiment are being processed by the following four steps:

1. A DRMF landmark face detector [12] is employed to annotate the 68 feature points.
2. One of the faces that portrays neutral expression is selected as the template.
3. LWM [16] is applied to transform the face shapes of the first frame of each video into the template face, by referring to the positions of the 68 landmark points.
4. Since the transformation matrix of each video is different, the remaining frames in the same video re-use the transformation matrix derived from its first frame.

2.2 Apex Frame Estimation

Inspired by [37, 38], the apex spotting algorithm is adopted on the databases that do not provide the apex frame information, i.e., SMIC. Concretely, the apex spotting is acquired by first identifying the salient regions that contain interesting expression details. Liong et al. [19] summarize that only three parts (i.e., eye, eyebrow and mouth) on the face portrays the expression information frequently, based on the data observation in CASME II. Then, three rectangle bounding boxes that cover the three facial regions are extracted as the regions of interest (RoIs). Note that RoIs for each video is different, as the bounding boxes are defined according to the landmark points detected.

Next, LBP features are encoded to describe the local texture information of each RoI. To determine the frame that poses the maximum expression intensity, the similarity index score is computed between the onset (assumed as neutral expression) and the remaining frames. It is obtained by employing the correlation coefficient method. Suppose, the higher the difference obtained, the higher the possibility of that frame as the apex frame. Lastly, to avoid the misjudgment due to the spikes that may be caused by background noises, a divide-and-conquer strategy is applied to identify the apex frame index.

In the following sections, we define the video sequence in a dataset as:

$$s_i = \{f_{i,j} | i=1,...,n; j=1,...,F_i\}$$  \hspace{1cm} (1)

where $i$-th refers the number of video and $F$ is the total number of frames in that video. Note that, each video only comprises one onset (starting), apex (climax) and offset (ending) frames, that can be expressed as $f_{i,1}, f_{i,m} \in \{f_{i,1}, f_{i,F_i}\}$ and $f_{i,F_i}$, respectively.

2.3 Optical Flow Components Computation and Features Concatenation

OF is adopted to compute the motion changes between the onset and apex frames. In general, the estimated OF field from two frames can be formulated as:

$$O_i = \{(u(x, y), v(x, y)) | x = 1, 2, ..., X; y = 1, 2, ..., Y\}$$  \hspace{1cm} (2)

where $u$ and $v$ represent the horizontal and vertical components of OF, respectively. $X$ is the width and $Y$ is the height of the image frame. Alternatively, OF can be converted from this vector form (i.e., $(u,v)$) to polar form, viz., magnitude ($\rho$) and orientation ($\theta$):

$$\rho(u,v) = \sqrt{(u(x, y))^2 + (v(x, y))^2}$$  \hspace{1cm} (3)

$$\theta(u,v) = \tan^{-1}\left(\frac{v(x, y)}{u(x, y)}\right)$$  \hspace{1cm} (4)

In addition, the optical strain information can be derived to describe the facial deformation, and can be estimated as follows:

$$\varepsilon = \frac{1}{2}[\nabla u + (\nabla u)^T]$$

$$= \begin{bmatrix}
\varepsilon_{xx} & \varepsilon_{xy} & \frac{1}{2}(\varepsilon_{uv} + \varepsilon_{vu}) \\
\varepsilon_{yx} & \varepsilon_{yy} & \frac{1}{2}(\varepsilon_{uv} + \varepsilon_{vu}) \\
\frac{1}{2}(\varepsilon_{uv} + \varepsilon_{vu}) & \frac{1}{2}(\varepsilon_{uv} + \varepsilon_{vu}) & \varepsilon_{zz}
\end{bmatrix}$$  \hspace{1cm} (5)

where the diagonal terms, ($\varepsilon_{xx}, \varepsilon_{yy}$) are normal strain components and ($\varepsilon_{xy}, \varepsilon_{yx}$) are shear strain components.

Thus, there are five OF derived components: $u, v, \rho, \theta$ and $\varepsilon$. New feature maps can be created by concatenating any three of them along the third dimension, mimicking the RGB-like vectorial field.
For instance, the combinations of \((u,v,\rho)\) [39] or \((u,v,e)\) [25][40]. These features will be served as the input data to the feature extractors in the next processing step.

2.4 Feature Extraction using Existing Architecture
Motivated by the OF derived input features with the apex-based approaches (i.e., Bi-WOOF [40], OFF-ApexNet [35] and STSTNet [25]) that had been outperformed others conventional ME recognition methods, we directly exploit their feature extraction architectures in our experiment. The maximum length of the feature vectors produced by Bi-WOOF, OFF-ApexNet and STSTNet are 512, 1024 and 1024, respectively.

2.4.1 Bi-WOOF
(Bi-Weighted Oriented Optical Flow) [40] is the first work at recognizing ME using only the apex frame and impressive had been achieved. Succinctly, it utilizes the three OF components (i.e., u, v and e). For each video, the three features maps are first partitioned into \(N \times N\) non-overlapping blocks. Then, the orientation histogram is generated on each block. The magnitude and optical strain values are treated as the local and global weighting schemes, respectively.

2.4.2 OFF-ApexNet
(Optical Flow Features from Apex frame Network) [35] is a deep learning architecture that receives u and v images as the input data. The two images are processed individually through two parallel streams of two convolutional and two maximum pooling layers. Then, the output of both the streams are concatenated to serve as the input vector to the fully connected layers.

2.4.3 STSTNet
(Shallow Triple Stream Three-dimensional CNN) [25] is one of the computationally light deep learning model, that is capable to produce outstanding recognition performance. In addition, this approach wins second place in MEGC 2019. In brief, a 3D image that composes of u, v and e feeds into this architecture. Then, three parallel streams that comprises one convolutional layer and one pooling layer in each stream are designed to extract different meaningful features.

2.5 Features Enhancement using Genetic Algorithm
GA is a feature integration stage that highlights important expression details whilst removing the redundant data. It is a search heuristic approach that gradually enhances the feature selection process across each evolutionary process. Among the coding methods, we opt for binary encoding, as it is easier to implement the crossover and mutation operations. In short, the crossover operator selects some of the feature candidates and splice them together to generate new features. On the other hand, the mutation is to invert some feature candidates at random.

To provide a clear understanding regarding the process of GA, a flow diagram of GA is shown in Figure 2. The input data of the GA is the output vector of the three aforementioned feature descriptors, that are performed separately. Since the length of the feature vectors are not fixed, we apply Principal component analysis (PCA) [41] statistical technique to standardize the length size to 400. The algorithm first generates random columns are at random positions. Those selected column are marked as 1, whereas the others are annotated as 0. An \(N\) columns is generated from the 400 columns, where the first individual is having the matrix size of \(S \times N\), where \(S\) is the number of videos in the dataset. Assume that the first generation has four individuals, there will be four \((0 - 1)\) matrix as the original chromosomes, and the quantity of individuals are the same in every generation.
Then, each individual is evaluated and the recognition results are compared. "Good" individuals are chosen to reproduce while "bad" individuals are removed from the next generation. Therefore, the new generation inherits discriminant properties of the previous generation, thus leading to evolve or highlight meaningful features for expression classification.

3. RESULTS AND ANALYSIS

The performance metric to examine the robustness of the proposed method is accuracy, F1-score, UF1 and UAR. This is following the suggestion of MEGC 2019 to make a fair comparison to the state-of-the-arts results. All the results presented are validated based on leave-one-subject-out (LOSO) cross validation protocol. A comprehensive recognition performance comparison is tabulated in Table 2. Concretely, methods #1 and #2 are the conventional feature extraction approaches that exploited the LBP and OF families. For methods #3 ~ #7 adopt deep learning models to encode the features as well as to perform expression classification. Specifically for #3 ~ #6, they are well-known architectures that have been demonstrated that their capabilities in many computer vision applications. Methods #8 ~ #11 are the top four submitting teams in MEGC 2019, while methods #12 ~ #15 are the proposed approaches in this paper.
### Table 2. Comparison of micro-expression recognition performance in terms of Accuracy (Acc), F1-score, Unweighted F1-score (UF1) and Unweighted Average Recall (UAR) on the composite (Full), CASME II, SMIC and SAMM databases

| No. Method                  | Composite | SMIC | CASME II | SAMM |
|-----------------------------|-----------|------|----------|------|
|                            | Acc       | F1   | UF1      | UAR  |
| Handcrafted                 |           |      |          |      |
| 1 BIP-TPS [6, 8]            | 0.5882    | 0.5785 | 0.5280  | 0.7429 | 0.3954 | 0.5102 |
| 2 BIP-WOOF [40]             | 0.6304    | 0.6296 | 0.6227  | 0.7305 | 0.8026 | 0.5211 | 0.5139 |
| 3 AlexNet [41]              | 0.6933    | 0.7154 | 0.5727  | 0.7805 | 0.8026 | 0.5211 | 0.5139 |
| Deep learning               |           |      |          |      |
| 4 SqueezeNet [43]           | 0.5950    | 0.6166 | 0.5361  | 0.6894 | 0.7278 | 0.5039 | 0.5362 |
| 5 Goolanet [44]             | 0.5573    | 0.6049 | 0.5123  | 0.5989 | 0.6414 | 0.5124 | 0.5992 |
| 6 VGG16 [45]                | 0.6437    | 0.6516 | 0.5800  | 0.8106 | 0.8202 | 0.4870 | 0.4793 |
| 7 OFF-ApeXNet [35]          | 0.7196    | 0.7096 | 0.6817  | 0.8674 | 0.8681 | 0.5409 | 0.5392 |
| MEGC 2019                   |           |      |          |      |
| 8 Quan et al. [27]          | 0.6520    | 0.6506 | 0.5820  | 0.7068 | 0.7018 | 0.6209 | 0.5989 |
| 9 Zhou et al. [26]          | 0.7322    | 0.7278 | 0.6645  | 0.8621 | 0.8569 | 0.5868 | 0.5663 |
| 10 STSTNet [25]             | 0.7389    | 0.7353 | 0.6801  | 0.8703 | 0.8382 | 0.6868 | 0.6610 |
| 11 Liu et al. [24]          | 0.7885    | 0.7824 | 0.7461  | 0.7530 | 0.8293 | 0.8209 | 0.7754 | 0.7152 |
| Proposed                    |           |      |          |      |
| 12 BI-WOOF (5-5) + GA       | 0.7021    | 0.7097 | 0.6758  | 0.8711 | 0.8654 | 0.6039 | 0.6035 |
| 13 BI-WOOF (5-5) + GS       | 0.7738    | 0.7970 | 0.6784  | 0.8711 | 0.8654 | 0.6039 | 0.6035 |
| 14 OFF-ApeXNet + GA         | 0.7195    | 0.6557 | 0.6331  | 0.7313 | 0.8547 | 0.5112 | 0.6405 |
| 15 STSTNet + GA             | 0.7688    | 0.7663 | 0.7169  | 0.8772 | 0.8872 | 0.6678 | 0.6995 |

The results of the original feature extractors Bi-WOOF, OFF-ApeXNet and STSTNet are listed in methods #2, #7 and #10, respectively. It is obvious the results have been improved when the features are integrated with GA. For example, when comparing Bi-WOOF method (i.e., #2 and #13), UF1 and UAR have been increased by 10% and 15%. Notably, STSTNet + GA (i.e., method #15) obtains the highest accuracy and F1-scores among the methods, which are 0.8009 and 0.7688. However, it can be observed that #11 outperforms in most of the cases. Besides, it is worth mentioning that BiWOOF + GA (i.e., method #12) exhibits the best UAR, with UAR=0.9232. Note that, this is a significant breakthrough whereby the recognition result reaches more than 90% in CASME II dataset.

To further verify the proposed approach, the confusion matrix of STSTNet + GA is reported in Table 3. It can be seen that CASMEII exhibits 100% accuracy rate in recognizing the surprise emotion class. The overall results in SMIC are lower compared to CASME II and SAMM. This is because the camera used to capture the SMIC samples is having the frame rate of 100fps, whereas CASME II and SAMM have double the number, viz, 200fps. Thus, it provides more precise apex frames and hence leads to better recognition results. In addition, the video elicitation environment and the acquisition setup for CASME II and SAMM are more careful to compare with SMIC, especially to avoid various background noises being captured. Moreover, the spotting step may introduce potential errors because of the reason of inaccurate in spotting the apex frames.

### Table 3. The confusion matrix of STSTNet+GA on composite, SMIC, CASME II and SAMM databases (measured by recognition rate %)

|                | (a) Composite | (b) SMIC | (c) CASME II | (d) SAMM |
|----------------|--------------|----------|--------------|----------|
|                | Neg | Pos | Sur | Neg | Pos | Sur | Neg | Pos | Sur | Neg | Pos | Sur |
| Neg            | 84.40 | 10.00 | 5.60 | 75.71 | 12.86 | 11.43 | 79.45 | 12.72 | 7.83 | 83.70 | 10.87 | 5.43 |
| Pos            | 19.27 | 73.39 | 7.34 | 15.69 | 72.55 | 11.76 | 18.60 | 9.30 | 72.09 |
| Sur            | 13.25 | 6.02 | 80.72 | 7.60 | 6.02 | 86.38 |
| Neg            | 92.00 | 6.82 | 1.14 | 83.70 | 10.87 | 5.43 |
| Pos            | 18.75 | 78.13 | 3.13 | 26.92 | 69.23 | 3.85 |
| Sur            | 0.00 | 0.00 | 100.00 | 0.00 | 0.00 | 100.00 |
4. CONCLUSION
In a nutshell, this paper presents the first attempt that utilizes genetic algorithm in the micro-expression recognition system. Promising recognition results are attained when integrating the genetic algorithm on existing feature descriptors (i.e., Bi-WOOF, OFF-ApexNet and STSTNet). The best accuracy and F1-score attained in the composite database are 80.09% and 76.88%, respectively. The composite database merges the three spontaneous publicly available micro-expression databases that comprises the three expressions, viz, positive, negative and surprise emotions. In the future, there will be more analysis to explore the integration of genetic algorithms with existing architectures to enhance identification performance.

ACKNOWLEDGMENTS
This work was funded by Ministry of Science and Technology (MOST) (Grant Number: MOST108-2221-E-035-066- and Grant Number: MOST 108-2218-E-227-002- and MOST-108-2115-M-227-001-MY2), National Natural Science Foundation of China (No. 61772023) and National Key R&D Program of China (No. 2019QY1803).

REFERENCES
[1] P. Ekman and W. V. Friesen, “Nonverbal leakage and clues to deception,” Journal for the Study of Interpersonal Processes, vol. 32, pp. 88-106, 1969.
[2] Matthew Shreve, Sridhar Godavarthy, Vasant Manohar, Dmitry Goldgof, and Sudeep Sarkar, “Towards macro- and micro-expression spotting in video using strain patterns,” in 2009 Workshop on Applications of Computer Vision (WACV). IEEE, 2009, pp. 1-6.
[3] Tomas Pfister, Xiaobai Li, Guoying Zhao, and Matti Pietikainen, “Recognising spontaneous facial microexpressions,” in 2011 international conference on computer vision. IEEE, 2011, pp. 1449-1456.
[4] Xiaobai Li, T Pfister, Xiaohua Huang, and Guoying Zhao, “A spontaneous micro-expression database: Inducement, collection and baseline,” in IEEE International Conference Workshops on Automatic Face Gesture Recognition, 2013.
[5] Wen Jing Yan, Qi Wu, Yong Jin Liu, Su Jing Wang, and Xiaolan Fu, “Casme database: A dataset of spontaneous micro-expressions collected from neutralized faces,” in IEEE International Conference Workshops on Automatic Face Gesture Recognition, 2013.
[6] W. J. Yan, X. Li, S. J. Wang, G. Zhao, Y. J. Liu, Y. H. Chen, and X. Fu, “Casme ii: an improved spontaneous micro-expression database and the baseline evaluation,” Plos One, vol. 9, no. 1, pp. e86041, 2014.
[7] Fangbing Qu, Su-Jing Wang, Wen-Jing Yan, He Li, Shuhang Wu, and Xiaolan Fu, “Cas (me) 2 A database for spontaneous macro-expression and microexpression spotting and recognition,” IEEE Transactions on Affective Computing, vol. 9, no. 4, pp. 424-436, 2017.
[8] Adrian K. Davison, Cliff Lansley, Nicolas Costen, Kevin Tan, and Moi Hoon Yap, “Samm: A spontaneous micro-facial movement dataset,” IEEE Transactions on Affective Computing, vol. 9, no. 1, pp. 116-129, 2018.
[9] Adrian K. Davison, Cliff Lansley, Nicolas Costen, Kevin Tan, and Moi Hoon Yap, “Samm: A spontaneous micro-facial movement dataset,” IEEE Transactions on Affective Computing, vol. 9, no. 1, pp. 116-129, 2018.
[10] John See, Moi Hoon Yap, Jingting Li, Xiaopeng Hong, and Su-Jing Wang, “Megc 2019-the second facial micro-expressions grand challenge,” in 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019). IEEE, 2019, pp. 1-5.
[11] S Milborrow and F Nicolls, “Active shape models with sift descriptors and mars. visapp, 2014,” .
[12] Akshay Asthana, Stefanos Zafeiriou, Shiyang Cheng, and Maja Pantic, “Robust discriminative response map fitting with constrained local models,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2013, pp. 3444-3451.
[13] Jason M Saragih, Simon Lucey, and Jeffrey F Cohn, “Face alignment through subspace constrained mean shifts,” in 2009 IEEE 12th International Conference on Computer Vision. IEEE, 2009, pp. 1034-1041.
[14] I Megvii, “Face++ research toolkit,” 2013.
[15] David Cristinacce and Timothy F Cootes, “Feature detection and tracking with constrained local models,” in Bmvc. Citeseer, 2006, vol. 1, p. 3.
[16] Ardeshir Goshtasby, “Image registration by local approximation methods,” Image and Vision Computing, vol. 6, no. 4, pp. 255-261, 1988.
[17] Zhaoqiang Xia, Xiaoyi Feng, Jinye Peng, Xianlin Peng, and Guoying Zhao, “Spontaneous micro-expression spotting via geometric deformation modeling,” Computer Vision and Image Understanding, vol. 147, pp. 87-94, 2016.
[18] Su-Jing Wang, Wen-Jing Yan, Guoying Zhao, Xiaolan Fu, and Chun-Guang Zhou, “Micro-expression recognition using robust principal component analysis and local spatiotemporal directional features,” in European Conference on Computer Vision. Springer, 2014, pp. 325-338.
[19] Sze-Teng Liong, John See, Raphael C-W Phan, Kok-Sheik Wong, and Su-Wei Tan, “Hybrid facial regions extraction for micro-expression recognition system,” Journal of Signal Processing Systems, vol. 90, no. 4, pp. 601-617, 2018.
[20] Mihailo R Jovanovic, Peter J Schmid, and Joseph W Nichols, “Sparsity-promoting dynamic mode decomposition,” Physics of Fluids, vol. 26, no. 2, pp. 024103, 2014.
[21] Anh Cat Le Ngo, John See, and Raphael C-W Phan, “Sparsity in dynamics of spontaneous subtle emotions: analysis and application,” IEEE Transactions on Affective Computing, vol. 8, no. 3, pp. 396-411, 2016.
[22] Ziheng Zhou, Guoying Zhao, Yimo Guo, and Matti Pietikainen, “An image-based visual speech animation system,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 10, pp. 1420-1432, 2012.
[23] Feng Xu, Junping Zhang, and James Z Wang, “Microexpression identification and categorization using a facial dynamics map,” IEEE Transactions on Affective Computing, vol. 8, no. 2, pp. 254-267, 2017.
[24] Yuchi Liu, Heming Du, Liang Zheng, and Tom Gedeon, “A neural micro-expression recognizer,” in 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019). IEEE, 2019, pp. 1-4.
[25] Sze-Teng Liong, YS Gan, John See, Huai-Qian Khor, and Yen-Chang Huang, “Shallow triple stream threedimensional cnn (ststnet) for micro-expression recognition,” in 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019). IEEE, 2019, pp. 1-5.
[26] Ling Zhou, Qirong Mao, and Luoyang Xue, “Dualinception network for cross-database micro-expression recognition,” in 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019). IEEE, 2019, pp. 1-5.
[27] Nguyen Van Quang, Jinhee Chun, and Takeshi Tokuyama, “Capsulenet for micro-expression recognition,” in 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019). IEEE, 2019, pp. 1-7.
[28] Guoying Zhao and Matti Pietikainen, “Dynamic texture recognition using local binary patterns with an application to facial expressions,” IEEE Transactions on Pattern Analysis & Machine Intelligence, , no. 6, pp. 915-928, 2007.
[29] B Horn, B Schunck, and JJ Pearson, “Artif. intell," Determining optical flow, vol. 17, pp. 185-203, 1981.
[30] Yandan Wang, John See, C. W. Phan, and Yee Hui Oh, “Lbp with six intersection points: Reducing redundant information in lbp-top for micro-expression recognition,” 2014.
[31] Xiaohua Huang, Guoying Zhao, Xiaopeng Hong, Wen-ming Zheng, and Matti Pietikainen, “Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns,” Neurocomputing, vol. 175, pp. 564-578, 2016.

[32] Xiaohua Huang, Su-Jing Wang, Xin Liu, Guoying Zhao, Xiaoyi Feng, and Matti Pietikainen, “Discriminative spatiotemporal local binary pattern with revisited integral projection for spontaneous facial micro-expression recognition,” IEEE Transactions on Affective Computing, pp. 1-1, 2018.

[33] Min Peng, Chongyang Wang, Tong Chen, Guangyuan Liu, and Xiaolan Fu, “Dual temporal scale convolutional neural network for micro-expression recognition,” Frontiers in psychology, vol. 8, pp. 1745, 2017.

[34] Haoyuan Ma, Gaoyun An, Shengjie Wu, and Feng Yang, “A region histogram of oriented optical flow (roof) feature for apex frame spotting in micro-expression,” in 2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), IEEE, 2017, pp. 281-286.

[35] YS Gan, Sze-Teng Liong, Wei-Chuen Yau, Yen-Chang Huang, and Lit-Ken Tan, “Off-apexnet on microexpression recognition system,” Signal Processing: Image Communication, vol. 74, pp. 129-139, 2019.

[36] John Daniel Bagley, “The behavior of adaptive systems which employ genetic and correlation algorithms: technical report,” 1967.

[37] Sze-Teng Liong, John See, KokSheik Wong, Anh Cat Le Ngo, Yee-Hui Oh, and Raphael Phan, “Automatic apex frame spotting in micro-expression database,” in 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR). IEEE, 2015, pp. 665-669.

[38] Sze-Teng Liong, John See, KokSheik Wong, and Raphael Chung-Wei Phan, “Automatic micro-expression recognition from long video using a single spotted apex,” in Asian Conference on Computer Vision. Springer, 2016, pp. 345-360.

[39] Huai-Qian Khor, John See, Sze-Teng Liong, Raphael CW Phan, and Weiyao Lin, “Dual-stream shallow networks for facial micro-expression recognition,” in 2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019, pp. 36-40.

[40] Sze-Teng Liong, John See, KokSheik Wong, and Raphael C-W Phan, “Less is more: Micro-expression recognition from video using apex frame,” Signal Processing: Image Communication, vol. 62, pp. 82-92, 2018.

[41] Ian Jolliffe, Principal component analysis, Springer, 2011.

[42] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in neural information processing systems, 2012, pp. 1097-1105.

[43] Forrest N Iandola, Song Han, Matthew W Moskewicz, Khalid Ashraf, William J Dally, and Kurt Keutzer, “SqueezeNet: Alexnet-level accuracy with 50x fewer parameters andd 0.5 mb model size,” arXiv preprint arXiv:1602.07360, 2016.

[44] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, “Going deeper with convolutions,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1-9.

[45] Karen Simonyan and Andrew Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.