Multimed Tools Appl
https://doi.org/10.1007/s11042-017-5317-2

A survey: facial micro-expression recognition

Madhumita Takalkar[1] · **Min Xu**[1] · **Qiang Wu**[1] · **Zenon Chaczko**[1]

Received: 29 November 2016 / Revised: 13 June 2017 / Accepted: 17 October 2017

Abstract Facial expression recognition plays a crucial role in a wide range of applications of psychotherapy, security systems, marketing, commerce and much more. Detecting a macro-expression, which is a direct representation of an 'emotion,' is a relatively straight-forward task. Playing a pivotal role as macro-expressions, micro-expressions are more accurate indicators of a train of thought or even subtle, passive or involuntary thoughts. Compared to macro-expressions, identifying micro-expressions is a much more challenging research question because their time spans are narrowed down to a fraction of a second, and can only be defined using a broader classification scale. This paper is an all-inclusive survey-cum-analysis of the various micro-expression recognition techniques. We analyze the general framework for micro-expression recognition system by decomposing the pipeline into fundamental components, namely face detecting, pre-processing, facial feature detection and extraction, datasets, and classification. We discuss the role of these elements and highlight the models and new trends that are followed in their design. Moreover, we provide an extensive analysis of micro-expression recognition systems by comparing their performance. We also discuss the new deep learning features that can, in the near future, replace the hand-crafted features for facial micro-expression recognition. This survey has been developed, focusing on the methodologies applied, databases used, performance regarding recognition accuracy and comparing these to distil the gaps in the efficiencies, future scope, and research potentials. Through this survey, we intend to look into this problem and develop a comprehensive and efficient recognition

✉ Madhumita Takalkar
madhumita.a.takalkar@student.uts.edu.au

Min Xu
min.xu@uts.edu.au

Qiang Wu
qiang.wu@uts.edu.au

Zenon Chaczko
zenon.chaczko@uts.edu.au

[1] School of Computing and Communications, Faculty of Engineering and Information Technology, University of Technology Sydney, 15 Broadway, Ultimo, NSW 2007, Australia

Published online: 13 November 2017

scheme. This study allows us to identify open issues and to determine future directions for designing real-world micro-expression recognition systems.

Keywords Micro-expressions recognition · Feature detection · Feature extraction · Classification · Deep learning features

1 Introduction

Emotions play a very prominent and purposeful role in day-to-day life. There is a high possibility of ambiguity, in guessing the hidden emotion, within an expression that elicits during situations of low or normal stakes. High stake situations provide more probability in predicting the emotion correctly as compared to low and normal stakes. Occurring in high stake situations, micro-expressions are the basis for expressing involuntary feelings. Micro-expressions happen in a fraction of a second and are hard to be recognized in real time especially lacking related expertise.

Macro-expressions are usually displayed for 3/4th of a second to 2 s. Although there are many different categories of emotion, there can be six universal expressions: anger, disgust, fear, happiness, sadness and surprise [14]. Macro-expressions occur over a single or multiple regions of the face depending on the category of expression.

Micro-expressions are described as a habitual pattern of the human face that is observable but too brief to convey an emotion. Micro-expressions are extremely fast facial expressions that usually last for 1/25 s to 1/5 s [22, 66]. They can easily be neglected during the casual conversations.

Since micro-expressions are barely perceptible to humans, a Micro-Expression Training Tool (METT) has been developed by Ekman [15] to teach a human how to spot and respond to micro-expressions. However, micro-expressions can seldom be falsified, and the essential difference between macro- and micro-expressions is the duration instead of the intensity of the expressions [45]. Currently, although experts can identify the existence and recognize micro-expressions, the accuracy is only about 47% [22]. Thus, having a system to improve the micro-expression analyses and help identify and categorize a person's feelings automatically and correctly is desirable.

Automated macro facial expression recognition has an enormous amount of existing research. Researchers have developed many algorithms, which have achieved a recognition accuracy of over 90% [1, 14, 25], for above mentioned six standard posed macro facial expressions. A recent study in [25] proposes a novel technique which when compared with existing state-of-the-art technique indicates a better result.. On the contrary, micro-expressions have not yet been explored extensively due to several challenges.

One of the challenges, most of the researchers face, is the lack of a standard micro-expression database, which makes it difficult to obtain dynamic facial features to train an accurate micro-expression recognition system. There is no significant research on the dynamics of the micro-expressions. Since the appearance of the micro-expression completely resembles the six primary macro-expressions, it is possible for researchers to train a system based on the existing macro facial expression databases by utilizing the appearance information and ignoring the dynamic information. Other challenges include the development of robust methods, which could deal with the short span and low-intensity of micro-expressions.

There is a vast range of applications that can benefit from the study of micro-expressions. A primary reason for the strong interest in micro-expressions is that it proves to be an important clue for lie detection. For example, in situations when the suspects are being questioned, a micro-expression fleeting across the face can tell the Police that the criminal is pretending to be innocent. It can also benefit the border security officers for identifying suspicious behavior of the individuals during usual interviews of checking for potential dangers. In the study of psychotherapy, micro-expressions have been proved very helpful in understanding genuine emotions of the patients. Micro-expression recognition systems are sometimes also used as an additional module for user authentication [44]. In other fields, such as marketing, distance learning, and many more, micro-expressions can be used as recognition to reflect human reactions and feedback to advertisements, products, services and learning materials.

This paper was compiled with the intention of providing a comprehensive survey of the existing micro-expression recognition methods along with their outcomes, to offer researchers a convenient introduction to the recent developments in this domain.

The remaining paper is organized into five sections. Section 2 highlights the factors that affect the recognition accuracy of micro-expressions. Section 3 discusses existing methods of micro-expression recognition. We further discuss the specifications and properties of the current micro-expression databases in Section 4. Section 5 gives a comparison of the existing studies conducted on different databases. Finally, Section 6 points out the challenges, open issues and the future directions in micro-expression recognition.

2 Factors influencing recognition of micro-expressions

Micro-expression is contained in the flow of expressions when individuals are trying to repress their emotions. According to studies in [45, 66], it is observed that certain factors affect recognition of micro-expressions.

2.1 Emotional context

The existing studies have employed neutral expressions before and after the emotion. The research indicates that micro-expressions may be embedded not only in neutral expressions but also in other facial expressions such as sadness and happiness. According to the emotional regulation theory [66], in the priming task, primes presented for longer duration may lead to greater priming effect. Moreover, it is observed that emotional information influences attention [66]. The aims of this research are: 1) to investigate the effect of emotional context on micro-expressions; 2) to explore if the effect of the context was limited to a particular material, and 3) to investigate the reason of the effect. The findings will lead researchers to predict that the emotional context would indeed influence micro-expression recognition.

2.2 Duration of expression

The significant difference between a micro- and a macro- expression is the length for which the expression lasts. There have been many different estimates of the duration of a micro-expression. Thus, there is still a lack of consensus about the time range of the length of a micro-expression. Although the difference in duration might not be significantly noticeable, for micro-expressions it needs to be taken into account.

To verify the effect of duration on micro-expressions recognition, the researchers conducted two experiments [45] asking the participants to recognize the micro-expressions in the images shown to them. In Experiment 1 expression images were shown to participants for 40,120,200 or 300 ms. The researchers employed Brief Affect Recognition Test (BART) for Experiment 1. In Experiment 2 the participants were given the micro-expression recognition training using the Micro Expression Training Tool (METT) paradigm which played a significant role in recognition of the micro-expressions. The outcome of the experiments indicated that the participants could recognize the micro-expression in the images in 200 ms without training and 160 ms after training. The results suggest that the critical time point that differentiates micro-expressions was about 200 ms or less. Thus, in conclusion, the accuracy of the micro-expression recognition is a function of the duration of the expressions.

3 Existing methods on micro-expression recognition

Micro-expression recognition systems are developed by considering many factors and parameters. Many studies have been undertaken and still undergoing in delivering better recognition accuracy.

In this paper, we break down the micro-expression recognition systems into their fundamental components, as shown in Fig. 1, including face detection, pre-processing, facial feature extraction, classification and databases. We discuss the role of each element in detail as below.

3.1 Face detection

Face detection is the primary stage of the recognition process. Human face(s) is located in the digital images or image sequences. This step is useful for selecting the region(s) of interest (ROI) in the images or selects ROI in the first frame and track the face in the remaining frames in case of image sequences.

There are several face detection methods enforced till date [26, 43, 48–50]. Some of the face detection techniques are summarized here. Mohammad Yeasin et al. [65] used automated face detection method to segment the face region which was based on the work of Rowley et al. [43]. M. Matsugu et al. [38] adapted convolutional neural network for detecting the face, and the rule-based algorithm is used for classification.

Viola et al. [49] introduced the first framework to provide competitive detection rates in real-time since 2001. This framework is capable of processing images rapidly while achieving high detection rates. The algorithm has four stages: 1) Haar feature selection; 2) Creating an integral image; 3) Adaboost training, and 4) Cascade classifiers. Fig. 2. illustrates the four distinct types of features used in the framework. The value of given feature is the sum of the pixels within white rectangles subtracted from the sum of the pixels within grey rectangles.

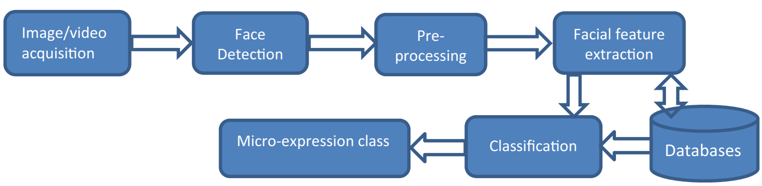

Fig. 1 A framework for micro-expression recognition analysis

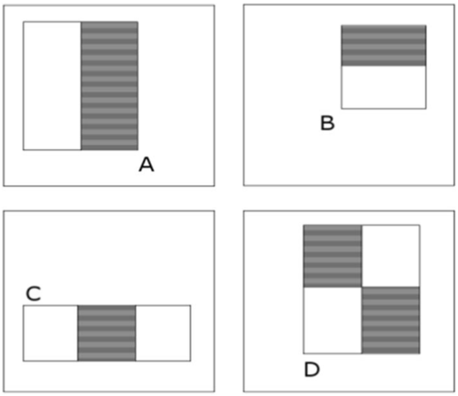

Fig. 2 Feature types used by Viola-Jones [49]

The study [8, 22] shows the use of Viola-Jones pre-processing method. The raw image is pre-processed and cropped using Haar feature. They implemented face detection using Haar features for facial features, which are then classified using a Classification And Regression Tree (CART). The CascadeObjectDetectorSystem defined in the MATLab Computer Vision Toolbox is implemented. This object detector has several built-in object detectors (eye, nose and mouth detectors). This function draws a bounding box around the face in the given image. The image was cropped using this bounding box and then resized to allow for faster processing of obtaining the feature descriptors. Viola-Jones algorithm is also implemented in OpenCV as cvHaarDetectionObjects() or Cascade of Classifiers which is used by researchers in [46].

3.2 Pre-processing

Pre-processing is the common name for operations performed on images at the lowest level. The aim is to achieve improvement of the image data that suppresses unwanted distortions or enhances some features for further processing. The sequences for micro-expressions are of very short duration wherein the intensity of the facial movements is low. There are several methods implemented to normalize the input data so that sufficient details about the micro-expression are extracted for further processing. Some of the novel pre-processing methods are discussed below.

1) *Temporal normalization (TIM)*

Temporal Interpolation Model (TIM) is used to increase short video lengths [23, 31, 40, 41, 54]. TIM uses graph embedding to interpolate images at arbitrary positions within micro-expressions. This interpolation allows inputting a sufficient number of frames to the feature descriptor. TIM is a manifold-based interpolation method that inserts a curve in a low-dimensional space after embedding of an image sequence. In Fig. 3, a micro-expression video is represented as a set of images sampled along the curve creating a low-dimensional manifold by delineating the micro-expression video as a path of the graph with vertices. The interpolated frames are mapped back to a high-dimensional space to form the temporally normalized image sequence [34].

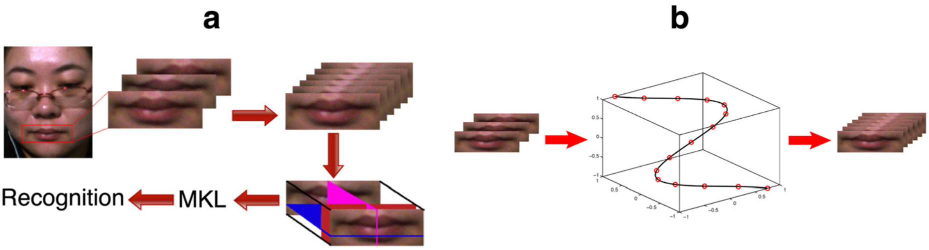

Fig. 3 (**a**) An example of micro-expression being interpolated through graph embedding; (**b**) Temporal interpolation method. The video is represented onto a curve along which a new video is sampled [34]

2) *Integral projection:*

Huang et al. [23] proposed a new framework to obtain the horizontal and vertical projections using the integral projection method based on calculating the difference of images, which helps to preserve the shape attributes of facial images. The integral projection generates a one-dimensional pattern by summing the given set of pixels along a given direction. The integral projections can extract common structure for the same person. In a micro-expression video clip, supposing that a frame is neutral, the difference between neutral face image and the expression image derive new images. The new derived facial images help reduce the influence of face identity on recognition methods. The integral projection itself does not describe the appearance and motion of facial images. It is, therefore combined with feature extraction method, e.g. LBP-TOP (as discussed in Section 3), to get the appearance and motion features. To preserve sufficient information in the process of projection, a new spatiotemporal method based on integral projection is introduced in [23]. Hence, the method is called as Spatiotemporal Local Binary Pattern with Integral Projection (STLBP-IP). Fig. 4 shows the procedure to encode integral projection by using LBP. STLBP-IP achieves state-of-the-art performance compared to TIM.

3) *Colour space model*

Colour is a fundamental aspect of human perception, and its effects on cognition and behaviour have attracted interests of many generations of researchers. Recent research revealed that colour might supply useful data for face recognition.

Wang et al. [51] demonstrated a Tensor Discriminant Color Space (TDCS) model that uses a 3rd-order tensor to represent a color facial image. To make the model

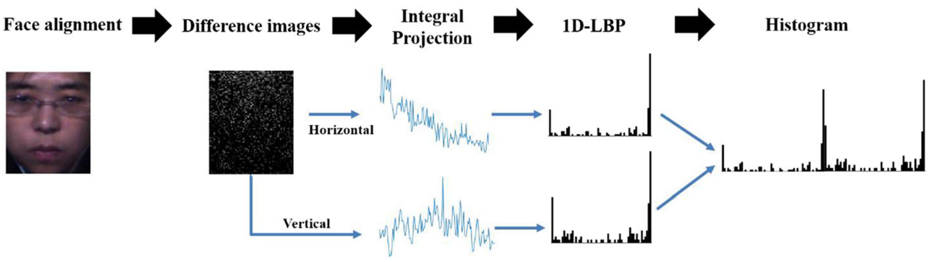

Fig. 4 The procedure of encoding difference-image based integral projection on the spatial domain [23]

robust to noise, they [52] also used an elastic net to propose a Sparse Tensor Discriminant Color Space (STDCS). Lajevardi and Wu [28] also addressed a color facial expression image as a 3rd-order tensor and presented that the perceptual color spaces (CIELab and CIELuv) are better overall than other color spaces for facial expression recognition.

A new color space model called tensor independent color space model (TICS) [55, 57] reveals that a micro-expression color video sequence is conceived as a fourth-order tensor, i.e., a four-dimension array. The first two dimensions cater the spatial information; the third delivers the temporal information, and the fourth is the color information. Wang et al. [57] transformed the fourth dimension from RGB into TICS, in which color components are as independent as possible. In a color micro-expression video clip, the correlated R, G and B components in RGB space are transformed into a series of uncorrelated components T1, T2 and T3, and extract the dynamic texture features from each uncorrelated component to obtain better results.

4) *Eulerian video magnification (EVM)*

EVM [8, 29, 31, 59], magnifies small and subtle motion which is impossible to be identified with naked eyes. EVM technique not only magnifies motion but also amplifies color. In EVM, certain spatial locations are selected to expand the variations in the temporal domain. Thus, EVM increases non-periodic movements with smaller magnitudes that are exhibited by the face. The extraction of features becomes easier with the magnified motion and colour videos.

Apart from the above mentioned pre-processing methods, the traditional methods such as Gaussian filter [32], Gaussian pyramid [56] are also widely used.

3.3 Facial feature extraction

For micro-expression recognition, feature extraction is an important critical issue. Recent studies show that spontaneous facial micro-expression analysis has been receiving attention from numerous researchers [30, 41] since involuntary micro-expressions can reveal genuine emotions which people try to conceal.

Face representations can be categorized as spatial and spatio-temporal. Spatial information encodes image sequences frame-by-frame, whereas spatio-temporal information considers a sequence of frames within a temporal window as a single parameter and enable modeling temporal variation to represent subtle expressions more efficiently. Another classification is based on the type of information encoded in space: shape and appearance. Geometry-based and appearance-based features have been commonly used to examine facial expression recognition.

Facial feature extraction is a two-step process: 1) Feature detection and 2) Feature extraction.

3.3.1 Feature detection

A feature is defined as an “interesting” component of an image. Feature detection is a low-level image processing operation which it is usually performed as the first operation on an image and analyzes every pixel to see if there is a feature present at that pixel.

Some of the feature detection methods are discussed as follows.

1) *Facial Action Coding System (FACS)*

The Facial Action Coding System (FACS) is an anatomically based system for comprehensively describing all facial movements. FACS devised by Ekman and Friesen [17], provides an objective means for measuring the facial muscle contractions involved in a facial expression. FACS was developed to allow researchers to measure the activity of facial muscles from video images of faces. Each noticeable component of facial movement is called an Action Unit (AU). Ekman and Friesen [17] defined 46 distinct action units, each of which corresponds to displacement in a specific muscle or muscle group, and produces facial feature deformations which can be identified in the images.

2) *Active Appearance Models (AAM)*

Active Appearance Models (AAM) is a statistically based template matching method, where a representative training set takes the variability of shape and texture. A group of images with landmark coordinates that appear in all of the images is given to the training supervisor. Edwards, Cootes, and Taylor [16] were the first to introduce the model in the context of face analysis. The method is widely used for matching and tracking faces and for medical image interpretation [10, 11].

The algorithm applies the difference between the current estimate of appearance and the target image to derive an optimization process. To match an image, the current residuals are measured and use the framework to anticipate changes to the present parameters, leading to a better match. A good overall match is obtained in a few iterations, even from poor starting estimates. AAMs, learn what are the valid shapes and intensity variations from their training set.

3) *Active Shape Models (ASM)*

Active Shape Model (ASM) algorithm is a fast and robust method of matching a set of points controlled by a shape model to a new image. Cootes et al. [9] proposed the active shape model where shape variability is learned through observation. ASM is again a statistical model of the shape of objects which iteratively deform to fit an example of the object in a new image.

The technique relies on each object or image structure being represented by a set of points. The points can represent a boundary, internal features, or even external ones, such as the center of a concave section of the border. Points are placed in the same way on each of a training set of examples of the object. The sets of points are aligned automatically to minimize the variance in the distance between similar points. By analyzing the statistics of the positions of the labeled points a “Point Distribution Model (PDM)” is derived. The model gives the average positions of the points and has some parameters which control the main modes of variation found in the training set [9].

4) *Discriminative Response Map Fitting (DRMF)*

Registering and tracking a non-rigid object has significant variations in shape and appearance. DRMF is one of holistic texture based methods, which relies on shape initialization. Moreover, as a discriminative regression-based approach, DRMF performs impressively well in the generic face fitting scenario [3]. DRMF is used to detect a set of facial feature points in the facial region of the first frame in each micro- expression video clip. DRMF located 68 feature points in a facial region. With the help of Facial Action Coding System (FACS), 36 Regions of Interest (ROI) are marked, and the face region is partitioned [33, 39] as shown in Fig. 5.

5) *Optical flow vectors*

Optical flow infers the motion of objects by detecting the changing intensity of pixels between two image frames over time. Lucas-Kanade [35] method assumes the displacement of the pixels between two nearby frames is small and nearly constant. Horn-Schunck introduces a global constraint of smoothness to solve the aperture problem. The method assumes smoothness in the flow over the whole image, trying to minimize distortions in the flow [21]. Optical flow method can be used for the face alignment provide better results than the image-domain-based method [33]. Usually, optical flow is extracted and analysed for cropped and pre-processed images to identify pose and face variations [46]. The research in [39] uses raw images as the input and Total Variation (TV-L1) optical flow estimation to analyse optical flow and discards the head movements. The integral of L1 norm of the gradient is referred as a Total variation (TV). Hence, the name is TV-L1.

3.3.2 Feature extraction

Feature extraction involves scaling down the amount of data required to represent a large set of data. Extraction of facial features is the most important step in recognition of micro-expressions. Many researchers have introduced different feasible features to represent facial characteristics. These features are classified as Geometric based features and Appearance-based features.

Geometric based features Geometric-based features represent face geometry, such as shapes and location of facial landmarks. These representations ignore skin texture. The facial elements or facial feature points are extracted to form a feature vector that represents the face geometry.

Some of the recently introduced geometry-based approaches are explained below.

1) *Delaunay-based temporal coding model (DTCM)*

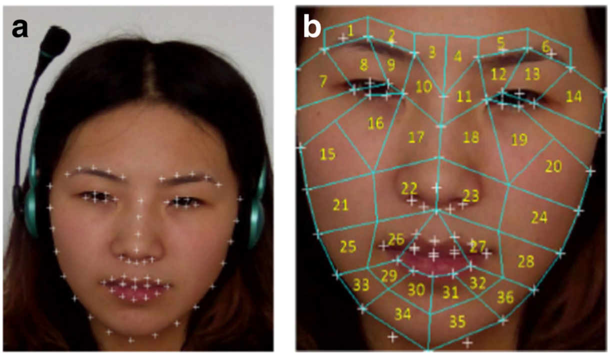

Fig. 5 (**a**) 66 feature points using DRMF; (**b**) 36 regions-of-interest (ROIs) [33]

Lu et al. [34] proposes the Delaunay-based temporal coding model (DTCM) in which the image sequences containing micro-expressions are normalized temporally as well as spatially based on Delaunay triangulation to remove the influence of personal appearance on micro-expression recognition. They applied Delaunay triangulation and standard deviation analysis to locate facial sub-regions related to micro-expressions. The variations of textures are converted instead of the feature points movements as is the case in most of the other methods.

2) *Main Directional Mean Optical Flow (MDMO)*

MDMO is ROI-based, a normalized statistic feature that considers both local static motion information and spatial location [33]. MDMO also reduces feature dimension. In MDMO, it selects the strongest component of ROI i.e. the main direction of the optical flow. The face region is divided into 36 regions and spatial coordinates and converted to polar coordinates (2 components). Therefore, the dimension is $36 \times 2 = 72$, which is far less then HOOF (36x8x2 = 576). The HOOF dimensions are obtained by trivially applying the HOOF feature in each ROI (36 ROIs, 8 bins, and 2 components). Histogram of Oriented Optical Flow (HOOF) [7] is used to model the activity profile in each video frame. It captures the optical flow orientation of the features and provides the histogram. An advantage of MDMO is that it does not depend on the number of frames in the image sequence. In comparison to LBP-TOP, MDMO showcased a better result in micro-expression recognition.

3) *Facial Dynamics Map (FDM)*

Another novel feature extraction method is called Facial Dynamics Map [62]. FDM is performed after the pre-processing and feature detection steps. Optical flow estimation finely aligns the cropped face image. To detect micro-expressions, a more compact representation of dynamics is required. Two assumptions, i.e. pixel-level description, and very high frame rate lead to the introduction of FDM. As shown in Fig. 6, there are three major steps. Firstly, facial landmarks points are located and used for face cropping and alignment. Secondly, an optical flow map is extracted for finer alignment. Finally, Facial Dynamics Maps are calculated for each clip for classification.

Appearance-based features

1) *Local Binary Pattern- Three Orthogonal Planes (LBP-TOP)*

The basic idea behind Local Binary Pattern (LBP) is to compare the center pixel value with the neighborhood pixel values. A binary code is generated by assigning the value one to the greater neighbor pixel value and assigning zero to the rest. The obtained binary code is converted to decimal to get the local binary pattern value of the center pixel. The calculation process of a basic LBP operator can be understood from Fig. 7.

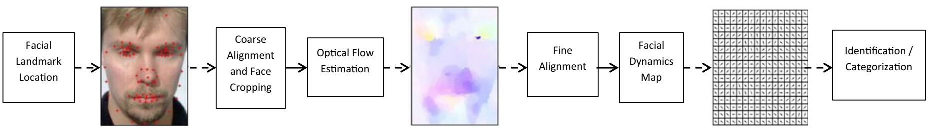

Fig. 6 Facial Dynamics Map: Adapted from [62]

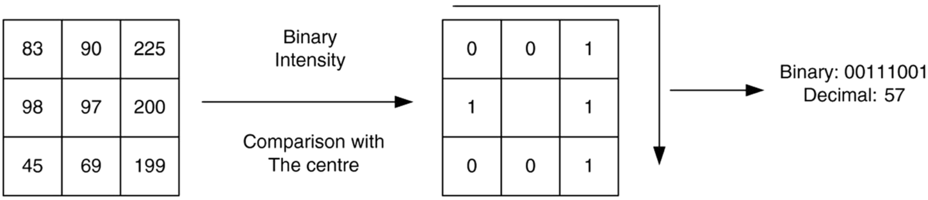

Fig. 7 The calculation process of a basic LBP operator

Local Binary Pattern-Three Orthogonal Planes (LBP-TOP) combines the temporal features along with the spatial features from LBP of the image sequence. A video is a chronological succession of frames with three dimensions, i.e. two dimensions (X and Y) are the spatial information and the third dimension (T) is time. The three orthogonal plane corresponds to the combination of spatial and temporal planes: XY, YT, and XT. LBP is firstly computed on these three planes. After that, histograms corresponding to each plane are obtained, which are concatenated to describe the dynamic texture of the micro-expression video [58, 67].

A feature vector is generated by calculating a histogram of LBPs over a whole image. The LBP method is effective for describing 2D textures of static images, but to analyze time-dependent textures (i.e. changing expressions in the video), LBP method needs to be broadened. To extend the LBP method, computation of LBP histograms is done in three orthogonal planes. For a video with time duration T, LBP is calculated for the XY-, XT-, and YT- planes. The XY- plane describes the spatial changes, while the XT- and YT-planes describe the spatial-temporal change in each respective dimension. The calculated histograms are then concatenated to form the final LBP-TOP feature vector. The frequency of patterns at each of the three planes is counted to prevail the corresponding histograms, which are then integrated to describe the dynamic texture of the video (Fig. 8).

LBP-TOP has been by far the most common method combined with various learning techniques for database evaluation and classification [30, 64].

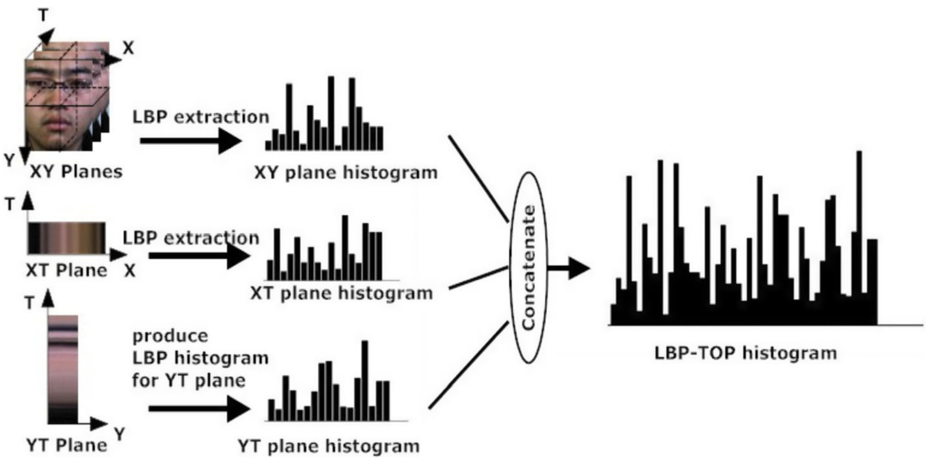

Fig. 8 LBP-TOP example: scan all the pixels to calculate their LBP histograms on the XY, XT, and YT plane respectively. The data of the pattern frequency is counted in each corresponding histogram and then concatenated as one [58]

2) *Histogram of Oriented Gradients (HOG)*

The histogram of oriented gradients is a feature descriptor used in computer vision and image processing for the use of object detection. The procedure counts occurrences of gradient orientation in localized portions of an image. According to the study of Dalal et al. [12], the local object appearance and shape within an image can be depicted by the distribution of intensity gradients or edge directions.

The image is divided into small spatial regions called cells, and for the pixels within each cell, a histogram of gradient direction is compiled. The descriptor is a concatenation of these histograms. For better precision, the local histograms can be contrast-normalized by calculating a measure of the intensity across a larger spatial region of the image, called a block, and then using this value to normalize all cells within the block. These normalized blocks are referred to as Histogram of Oriented Gradient (HOG) descriptors [12].

3) *Gabor wavelet filter*

In appearance-based methods, image filters like Gabor wavelets are used in all or part of the face to extract the feature vector. Gabor filter is a linear filter applied for edge detection. Frequency and orientation illustrations of Gabor filters are like those of the human visual system, and they have been determined to be particularly appropriate for texture representation and discrimination. Gabor filters are directly related to Gabor wavelets since they can be directly designed for some dilations and rotations. A filter bank comprised of Gabor filters with various scales and rotations is created. This set of Gabor filters with different frequencies and orientations is helpful for extracting useful features from an image. These filters remove most of the variability in the image caused by changes in lighting conditions.

Shape and texture changes characterize the deformation of facial features and lead to high spatial gradients that are good indicators of facial actions and may be analyzed either in the image or the spatial frequency domain. Gabor wavelet-based filters can compute the latter, which allows detecting line endings and edge borders over multiple scales and with different orientations.

The study by Wu et al. [61], proposes a new approach for automatic micro-expression recognition by ignoring the dynamical information and analyzing the videos frame by frame. They used the Gabor filters with nine scales and eight orientations. Fig. 9 shows the Gabor filters used for feature extraction on the pre-processed face image.

Deep learning features – Convolutional neural networks (CNN) Convolutional neural networks (ConvNets or CNNs) have been very effectively used in the area of image classification and understanding [24]. ConvNets have also been applied to identify faces and objects. Usually, there are four main steps in the ConvNet: 1) Convolution; 2) Non-Linearity (ReLU); 3) Pooling or Sub-Sampling; and 4) Classification (Fully connected layer). The primary purpose of the *Convolution* step is to extract features from an input image. Convolution maintains the spatial relationship between pixels by learning image features using small squares of input data. An additional function called *ReLU* has been used after Convolution operation. ReLU stands for Rectified Linear Unit and is a non-linear process. ReLU is an element-wise operation (applied per pixel) and substitutes all negative pixel values in the feature map by zero. The intention of ReLU is to introduce non-linearity in ConvNet. *Spatial Pooling* (also called as subsampling or downsampling) reduces the dimensionality of each

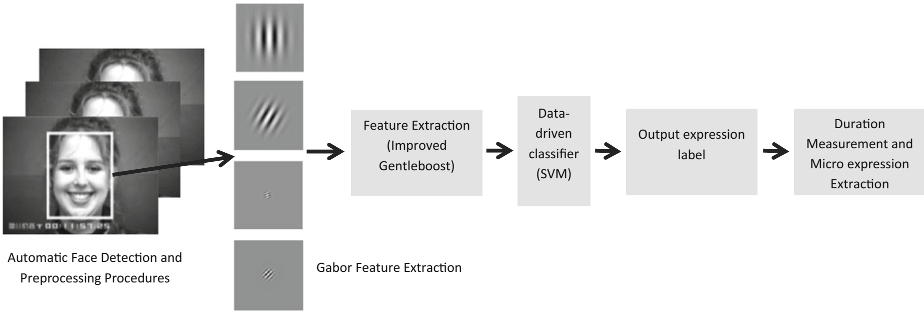

Fig. 9 Gabor feature extraction [61]

feature map but retains the most important information. Unitedly these layers extract the powerful features from the images, introduce non-linearity and reduce feature dimension. The *Fully connected layer* is a conventional Multi Layer Perceptron that utilizes a softmax activation function in the output layer. The term "Fully Connected" means that every neuron in the previous layer is associated with every neuron in the next tier. The aim of the fully connected layer is to apply these features for classifying the input image into respective classes based on the training dataset.

In recent research [27], Kim et al. introduced deep learning features for micro-expression recognition. They proposed a new method consisting of two parts. First, the spatial features of micro-expressions at different expression states (i.e., onset, onset to apex transition, apex, apex to offset transition and offset) are encoded using convolutional neural networks (CNN). Next, the learned spatial features with expression-state constraints are transferred to learn temporal features of micro-expression. The temporal feature learning encodes the temporal characteristics of the different states of the micro-expression using long short-term memory (LSTM) recurrent neural networks. The time scale dependent information that resides along the video sequences is consequently learned by using LSTM.

3.4 Classification

Image classification analyzes the statistical properties of various image features and organizes data into categories. Classification is typically a two-step process: training and testing. Most of the research on micro-expression recognition applies existing classification methods as discussed below.

1) *Support Vector Machines (SVM)*

SVMs [6, 22, 23, 41, 64] are based on the concept of decision planes that define decision boundaries, which primarily perform classification tasks by constructing hyperplanes and a multidimensional space that separate samples of different class labels. SVMs correlate to the general category of kernel methods, which can operate in a high-dimensional, implicit feature space, through applying kernel functions.

SVM has two advantages: Firstly, SVM can generate non-linear decision boundaries using methods intended for linear classifiers. Secondly, the use of kernel functions grants the user to implement a classifier to data that have no demonstrable fixed-dimensional vector space representation [4]. Some kernels can be used in SVMs, e.g., linear, polynomial, Radial Basis Function (RBF) and sigmoid. The RBF is one of the most widely used kernel types in SVMs mainly because of their localized and limited responses across the entire range of the real x-axis.

2) *Extreme Learning Machine (ELM)*

ELM is a single hidden layer feed-forward neural network which has extremely fast learning speed [53]. ELM has better generalization performance. ELM classifier provides a unified learning platform with popular features mappings and can be directly applied in the regression and multiclass classification [20]. According to the research, ELM is proved to have better generalization performance, much faster training and learning speed than the traditional SVM. This characteristic of ELM proves vital in the micro-expression recognition.

3) *Nearest Neighbor Algorithm (NNA)*

NNA depends on limited adjacent samples, so as compared to other methods, NNA is more efficient for the sample set with class fields cross [36]. In this paper [36], the system integrated the gradient magnitude weighted into Nearest Neighbour Algorithm for classification. The idea of the nearest neighbor method is to compare the distances between unknown samples with the entire known sample set and to judge the distances between samples. Euclidean distance is one of the measurements of similarity among samples [19]. If the distance of two samples in the feature space is close, then the samples may have the same label. The fine calculation ability of NNA can help the micro-expression recognition system to classify accurately.

4) *Multiple Kernel Learning (MKL)*

MKL is developed for supervised, semi-supervised and unsupervised learning. The fundamental idea behind MKL is to add an extra parameter to the minimization problem of the learning algorithm. MKL determines weights for linear/non-linear combinations of kernels over different domains by optimizing a cost function [41]. Compared to SVM, MKL can provide better micro-expression recognition in some cases.

5) *Random Forest (RF)*

Random Forest, also known as Random Decision Forest, is an ensemble learning method that is used for classification and can be thought of as a form of the nearest neighbor predictor. Ensemble learning is a divide-and-conquer method used to improve performance. The basic principle of ensemble methods is that a set of "weak learners" can come together to form a "strong learner" [5]. Several researchers [13, 34, 40, 41] opted to use RF classifier for facial micro-expression recognition.

4 Existing micro-expression databases

The success in macro- facial expression recognition primarily relies on sufficient facial expression databases, such as the popularly used CK+, MUG, MMI, JAFFE, Multi-PIE, and also several 3-D facial expression databases. In contrast, there are very few well-developed micro-expression databases, which have hindered the development of micro-expression recognition research. Some databases are developed by asking the participants to pose facial expressions quickly. However, these posed "micro-expressions" are different from the natural ones. The need for spontaneous facial micro-expression databases arises for academic research and practical applications. Recently, to capture involuntary micro-expressions, the participants were asked to watch some videos while their expressions were captured. Later, the captured expressions were verified by the participants.

Table 1 gives a brief description of the datasets used to evaluate micro-expression recognition systems. The table lists the statistics and properties of the available datasets. There exists posed and non-posed datasets for micro-expression recognition including six emotion categories (sadness, fear, happiness, disgust, anger, surprise) and/or three emotion categories (positive, negative, surprise). The datasets with spontaneous micro-expressions can potentially be useful for validating the systems' performance in detecting subtle expressions. The available micro-expression datasets are Spontameous Micro-Expression Database (SMIC), The Chinese Academy of Sciences Micro-Expression (CASME and CASME II), Micro-Expression Training Toolkit (METT), Unsupervised Segmentation Fusion-High Definition (USF-HD), York Deception Detection Test (York DDT).

Apart from the databases listed in Table 1, there are a few macro-expression databases, which have been incorporated in the micro-expression recognition system. Table 2 shows the list of such databases. The database names are Extended Cohn- Kanade (CK+), Audio-Visual Emotion recognition Challenges (AVEC), Japanese Female Facial Expression (JAFFE), Multimedia Understanding Group (MUG). Since some samples of these databases are identified as micro-expressions, they are usually used along with the micro-expression databases.

5 Comparison of different studies

In a comparison Table 3, we list different studies using different methods, databases, and their outcomes. The research in [61], implements Gabor filters for feature extraction and employs GentleSVM for recognition of micro-expression which achieves 85.42% recognition accuracy. A micro-expression dataset collected and used for training and Micro-expression videos from METT was used in [61] as the test set. The micro-expression data in METT are in their simplest form, which might make the recognition less challenging than using other datasets. The research in [31], applies SMIC micro- expression dataset along with a temporal normalization and feature detection method, i.e. Eulerian video magnification, which gives 81.69% recognition accuracy. The approach in [41] achieves 71.5% recognition accuracy through implementing LBP-TOP on YorkDDT dataset with MKL as the classifier. The most challenging datasets may be CASME and CASME II. Existing research on these datasets can only achieve 40% to 70% recognition accuracy. From the comparison, we can find that features development and classification method selection can significantly affect recognition accuracy; therefore, these two issues attract most of the research efforts.

Table 1 Statistics and properties of existing micro-expression databases

Dataset		Frame rate (Fps)	No. of participants	No. of samples	Elicitation	Emotion class	Remarks
SMIC [41]		100	6	76	Spontaneous	3	Positive, Negative and Surprise
SMIC2 [30]	HS	100	20	164	Spontaneous	3	Positive, Negative and Surprise
	VIR	25	10	71			
	NIS	25	10	71			
CASME [63]		60	35	195	Spontaneous	7	Tension and Repression in addition to primary emotions
CASME II [64]		200	35	247	Spontaneous		
USF-HD [46]		29.7	/	77	Posed	6	Their criteria set for micro-expressions (2/3 s) is longer than most accepted durations
Polikovsky database [42]		200	10	/	Posed	7	7 basic emotions with low facial muscle intensity
York DDT [60]		25	9	18	Spontaneous	4	Truth/lie, emotional/ unemotional
METT [18, 45, 60, 61]		–	12	–	Posed	7	–

Table 2 Macro-expression databases

Database	References	No. of samples	No. of subjects	Frame rate (Fps)	Elicitation	Emotion class
Extended Cohn- Kanade	[1, 8, 25]	16,128	28	–	Posed	9
Yale	[36]	165	15	30	Posed	5
YaleB	[53]	5850	10	30	Posed	5
AVEC	[47]	95		49.979	Posed	4 (Arousal, Power, Expectancy, and Intensity)
JAFFE	[25, 37]	213	10	–	Posed	7 (6 basic +1 neutral)
MUG	[1, 2]	Posed – 103,307 Spontaneous– 100,935	86	19	Posed and Spontaneous	6

Table 3 The recognition accuracy obtained from different studies using different methods

Database	Reference	Pre-processing + Feature Detection	Feature Extraction	Classifier	Label	Performance
YorkDDT	[41]	TIM + ASM	LBP-TOP	MKL	2 Emotions	71.5%
SMIC	[34]	TIM + AAM	Delaunay Triangulation	SVM	2 Emotions	82.86%
	[31]	Eulerian video magnification	HOG	SVM	3 Emotions	81.69%
	[33]	DRMF + optical flow	MDMO	SVM	4 Emotions	80.0%
	[62]	ASM + Local Weight Mean	Facial Dynamics Map	SVM	3 Emotions	71.43%
	[41]	TIM + ASM	LBP-TOP	MKL	2 Emotions	71. 4%
	[19]	–	LBP-TOP	NNA	3 Emotions	65.83%
CASME	[20]	Adaboost + bilinear interpolation	LBP-TOP	ELM	5 Emotions	73.82%
	[33]	DRMF + optical flow	MDMO	SVM	4 Emotions	68.86%
	[34]	TIM + AAM	Delaunay Triangulation	RF	4 Emotions	64.95%
	[55]	TIM + TICS	LBP-TOP	SVM	4 Emotions	61.85%
	[53]	TSA	DTSA	ELM	5 Emotions	46.90%
	[62]	ASM + Local Weight Mean	Facial Dynamics Map	SVM	3 Emotions	42.02%
CASME II	[59]	Eulerian video magnification	LBP-TOP	SVM	5 Emotions	75.30%
	[33]	DRMF + optical flow	MDMO	SVM	4 Emotions	67.37%
	[31]	Eulerian video magnification	HOG	SVM	3 Emotions	67.21%
	[34]	TIM + AAM	Delaunay Triangulation	RF	4 Emotions	64.19%
	[27]	Convolutional Neural Networks		LSTM	5 Emotions	60.98%
	[23]	TIM + integral projection	LBP-TOP	SVM	5 Emotions	59.51%
	[55]	TIM + TICS	LBP-TOP	SVM	4 Emotions	58.54%
	[29]	Eulerian video magnification	LBP-TOP	SVM	5 Emotions	47.0%

2 Emotions: Positive, Negative (emotional vs. unemotional in [41] for YorkDDT); 3 Emotions: Positive, Negative, Surprise; 4 Emotions: Positive, Negative, Surprise, and Tense; 5 Emotions: Happy, Disgust, Repression, Surprise and Others

6 Challenges, open issues, and future directions

In this section, we will discuss the challenges faced, current issues to be worked on and future scope of the research.

6.1 Challenges

The major challenges in recognizing micro-expressions derive from their short span and spontaneous nature. Below we will address what the authors consider to be the most pressing challenges in automatic micro-expression recognition. This includes the existing limited datasets, issues around obtaining ground truth and developing efficient recognition algorithms, among others.

1) *Temporal information*

- There are many spatial normalization methods but very few for normalizing temporal information. Due to the variation in the micro-expressions duration, it is necessary to normalize temporal dimension for micro-expressions. One example of normalization method is Temporal Interpolation Model (TIM) based on the Laplacian graph.
- Due to low-intensity facial movements in existing database, the computer may mark micro-expressions as a neutral face. Without temporal information, it is hard, even for a human eye, to detect and recognize micro-expressions. Thus, to better identify and distinguish micro-expressions, researchers should take temporal information into account

2) *Dimensionality reduction*

Very high dimensional data are generated when high-speed (200fps) and high-resolution (800 × 600) cameras are used for recording micro-expression video sequences. For such high dimensional data, it may be inappropriate to use simple dimensionality reduction methods, e.g. PCA, because it may not preserve the required feature information. Thus, it is necessary to develop a more efficient algorithm.

3) *Micro-expression detection*

To boost the recognition accuracy, it is important to spot the onset, peak and offset frames of a micro-expression. Currently, the research focuses on detecting and recognizing micro-expressions rather than identifying the onset, peak and offset frames. The software is highly demanded to help capture onset, peak and offset frames for building spontaneous micro-expression databases.

6.2 Open issues

Since this is a new and developing research field, the researchers are coming across many challenges and finding a solution for these challenges they have identified some issues that need to be addressed. Some of these open issues suggest increasing research efforts should be allocated into micro-expression database collection and recognition method development.

1) *Databases*

- Due to different cultural backgrounds and life experiences, people may have different responses to the same video. This factor makes labeling of spontaneous micro-expressions a difficult task. For example, the scene of chewing a worm is not always reported as "Disgusting." Some participants also reported it as "Funny" or "Interesting." A proper database can be considered to include information related to action units in regarding different expression categories as action units are widely used to describe facial expression. Categorisation of the micro-expressions requires further in-depth study.
- Micro-expression databases are usually created in a controlled environment, e.g. in laboratories, and may not cover the micro-expressions in different situations. Further studies should be conducted to explore and investigate variations of micro-expressions in various circumstances.
- It would be more helpful if the databases could include expressions of participants from different age groups from different cultural backgrounds to make micro-expression recognition more practical.

2) *Micro-expression recognition methods*

- Current micro-expression systems could not separate the micro-expressions from the non-emotional rapid movements such as eye blinks.
- Even the state-of-the-art methods are not exceptions in recognizing the micro-expressions incorrectly. The false positive rate has to be reduced to increase the efficiency of the system.

6.3 Future directions

Although feasible solutions have been proposed in the field of micro-expression recognition, there might still be a couple of possible research directions that we can suggest. The motivations for the below-mentioned suggestions come from the challenges and the open issues discussed in earlier sections.

- Estimating the pixel-level movement can help to refine the micro-expression recognition efficiency in real time.
- Expanding the databases to more participants will create a better training set to the recognition systems. Participants of all age groups and different cultural backgrounds could be included in the same database to make it more real-time.
- A micro-expression recognition system should work for less constrained and uncontrolled situation micro-expression videos.
- Some more research into deep learning features, e.g. CNN features, is required to be conducted on additional datasets to improve the micro-expression recognition results.

7 Conclusion

In this paper, we gave a far-reaching overview of cutting edge facial micro-expression recognition approaches including both handcrafted and learning-based methods that structure the key components of the micro-expression recognition system. The handcrafted micro-expression recognition approaches have been there for a significantly long time and accomplished surprising outcome on available benchmark datasets. However, most successful handcrafted recognition methods are based on the local densely-sampled descriptors. In these methodologies, the necessary, features are extracted from a sequence of image frames to generate the feature vector using human engineered feature detectors and descriptors. Later, the classification is performed by training a generic classifier. These approaches include space-time, appearance, geometry, and local binary patterns based methods.

On the other hand, learning-based micro-expression recognition method uses trainable feature extractors followed by the trainable classifier, which prompts the idea of end-to-end learning or learning from pixel level to micro-expression class identification. This rules out the need for handcrafted feature detectors and descriptors utilized for micro-expression recognition. These approaches include deep-learning-based approaches. Recently, the research community has given careful consideration to these methodologies. This is primarily due to their high performance as compared to their handcrafted counterparts on micro-expression datasets. However, fully data-driven deep learning models have a restriction: A few of the best-performing deep-learning-based methods are still dependent on handcrafted features. This is principally due to the unavailability of the major high-quality datasets for micro-expressions, unlike macro-facial expressions.

To give further insight into the field, we have classified the public micro-expression datasets. These datasets include CASME, CASME II, SMIC, USF-HD, Polikovsky, YorkDDT and Micro-Expression Training Toolkit (METT). Our paper also highlights the factors influencing the micro-expression recognition systems. We compared the different handcrafted and deep learning methods, based on the datasets used, to give a better view of their recognition performance. The biggest challenges the researchers have to deal with are a short span of micro-expressions which makes it difficult to detect a micro-expression and the lack of availability of large public datasets. As a consequence of all of the above micro-expression analysis, possible directions towards which research should focus on progressing beyond state of the art to become more robust and be able to adapt to more complex states.

References

1. Agarwal S, Mukherjee DP (2017) Facial expression recognition through adaptive learning of local motion descriptor. Multimed Tools Appl 76:1073–1099
2. Aifanti N, Papachristou C, Delopoulos A (2010) The MUG facial expression database. In: 11th international workshop on image analysis for multimedia interactive services WIAMIS 10, pp 1–4
3. Asthana A, Zafeiriou S, Cheng S, Pantic M (2013) Robust discriminative response map fitting with constrained local models. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3444–3451
4. Ben-Hur A, Weston J (2010) A user's guide to support vector machines. In: Carugo O, Eisenhaber F (eds) Data mining techniques for the life sciences. Methods in Molecular Biology (Methods and Protocols), vol 609. Humana Press, New York
5. Benyamin D (2012) A gentle introduction to random forests, ensembles, and performance metrics in a commercial system | Blog & Press. [online] Blog & Press. Available at: https://citizennet.com/blog/2012/11/10/random-forests-ensembles-and-performance-metrics/. Accessed 2 Nov 2015
6. Burges CJ (1998) A tutorial on support vector machines for pattern recognition. Data Min Knowl Disc 2:121–167
7. Chaudhry R, Ravichandran A, Hager G, Vidal R (2009) Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions. In: Computer vision and pattern recognition, 2009. CVPR 2009. IEEE conference on, pp 1932–1939
8. Chavali GK, Bhavaraju SKNV, Adusumilli T, Puripanda V (2014) Micro-expression extraction for lie detection using Eulerian video (motion and color) magnification. Dissertation

9. Cootes TF, Taylor CJ, Cooper DH, Graham J (1995) Active shape models-their training and application. Comput Vis Image Underst 61:38–59
10. Cootes TF, Edwards GJ, Taylor CJ (1998) Active appearance models. In: European conference on computer vision, pp 484–498
11. Cootes TF, Edwards GJ, Taylor CJ (2001) Active appearance models. IEEE Trans Pattern Anal Mach Intell 23:681–685
12. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: 2005 I.E. computer society conference on computer vision and pattern recognition (CVPR'05), pp 886–893
13. Davison AK, Yap MH, Costen N, Tan K, Lansley C, Leightley D (2014) Micro-facial movements: an investigation on spatio-temporal descriptors. In: European conference on computer vision, pp 111–123
14. Donia MM, Youssif AA, Hashad A (2014) Spontaneous facial expression recognition based on histogram of oriented gradients descriptor. Comput Inform Sci 7:31
15. Eckman P (2003) Emotions revealed. St. Martin's Griffin, New York
16. Edwards GJ, Taylor CJ, Cootes TF (1998) Interpreting face images using active appearance models. In: Automatic face and gesture recognition, 1998. Proceedings. Third IEEE international conference on, pp 300–305
17. Ekman P, Friesen W (1978) Facial action coding system: a technique for the measurement of facial movement. Consulting Psychologists, Palo Alto
18. Endres J, Laidlaw A (2009) Micro-expression recognition training in medical students: a pilot study. BMC Med Educ 9:47
19. Guo Y, Tian Y, Gao X, Zhang X (2014) Micro-expression recognition based on local binary patterns from three orthogonal planes and nearest neighbor method. In: 2014 international joint conference on neural networks (IJCNN), pp 3473–3479
20. Guo Y, Xue C, Wang Y, Yu M (2015) Micro-expression recognition based on CBP-TOP feature with ELM. Optik Int J Light Electron Opt 126:4446–4451
21. Horn BK, Schunck BG (1981) Determining optical flow. Artif Intell 17:185–203
22. House C, Meyer R (2015) Preprocessing and descriptor features for facial micro-expression recognition [Online]. Available: https://web.stanford.edu/class/ee368/Project_Spring_1415/Reports/House_Meyer.pdf. Accessed 15 Oct 2016
23. Huang X, Wang S-J, Zhao G, Piteikainen M (2015) Facial micro-expression recognition using spatiotemporal local binary pattern with integral projection. In: Proceedings of the IEEE international conference on computer vision workshops, pp 1–9
24. Karn U (2016) An intutive explanation of convolutional neural networks [Online]. Available: https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/. Accessed 2 Nov 2016
25. Khan SA, Hussain A, Usman M (2017) Reliable facial expression recognition for multi-scale images using weber local binary image based cosine transform features. Multimed Tools Appl. https://doi.org/10.1007/s11042-016-4324-z
26. Kim I, Shim JH, Yang J (2010) Face detection. Comput Inf Sci 3:71–90
27. Kim DH, Baddar WJ, Ro YM (2016) Micro-expression recognition with expression-state constrained spatio-temporal feature representations. In: Proceedings of the 2016 ACM on multimedia conference, pp 382–386
28. Lajevardi SM, Wu HR (2012) Facial expression recognition in perceptual color space. IEEE Trans Image Process 21:3721–3733
29. Le NgoAC, Oh Y-H, Phan RC-W, See J (2016) Eulerian emotion magnification for subtle expression recognition. In: acoustics, speech and signal processing (ICASSP), 2016 I.E. international conference on, pp 1243–1247
30. Li X, Pfister T, Huang X, Zhao G, Pietikäinen M (2013) A spontaneous micro-expression database: inducement, collection and baseline. In: Automatic face and gesture recognition (FG), 2013 10th IEEE international conference and workshops on, pp 1–6
31. Li X, Hong X, Moilanen A, Huang X, Pfister T, Zhao G et al (2015) Reading hidden emotions: spontaneous micro-expression spotting and recognition. arXiv preprint arXiv:1511.00423
32. Liong S-T, See J, Phan RC-W, Le Ngo AC, Oh Y-H, Wong K (2014) Subtle expression recognition using optical strain weighted features. In: Asian Conference on Computer Vision, pp 644–657
33. Liu YJ, Zhang JK, Yan WJ, Wang SJ, Zhao G, Fu X (2015) A main directional mean optical flow feature for spontaneous micro-expression recognition. IEEE Trans Affect Comput PP:1–1
34. Lu Z, Luo Z, Zheng H, Chen J, Li W (2014) A Delaunay-based temporal coding model for micro-expression recognition. In: Asian conference on computer vision, pp 698–711
35. Lucas BD (1985) Generalized image matching by the method of differences
36. Lun X, Xin L, Xiujun Y, Zhiliang W (2016) Cognitive regulation and emotion modeling for micro-expression. Int J Control Autom 9:361–372
37. Lyons M, Kamachi M, Gyoba J (1997) Japanese female facial expressions (JAFFE). Database of Digital Images
38. Matsugu M, Mori K, Mitari Y, Kaneda Y (2003) Subject independent facial expression recognition with robust face detection using a convolutional neural network. Neural Netw 16:555–559

39. Patel D, Zhao G, Pietikäinen M (2015) Spatiotemporal integration of optical flow vectors for micro-expression detection. In: International conference on advanced concepts for intelligent vision systems, pp 369–380
40. Pfister T, Li X, Zhao G, Pietikäinen M (2011) Differentiating spontaneous from posed facial expressions within a generic facial expression recognition framework. In: Computer vision workshops (ICCV workshops), 2011 I.E. international conference on, pp 868–875
41. Pfister T, Li X, Zhao G, Pietikäinen M (2011) Recognising spontaneous facial micro-expressions. In: 2011 international conference on computer vision, pp 1449–1456
42. Polikovsky S, Kameda Y (2013) Facial micro-expression detection in hi-speed video based on facial action coding system (FACS). IEICE Trans Inf Syst 96:81–92
43. Rowley HA, Baluja S, Kanade T (1998) Neural network-based face detection. IEEE Trans Pattern Anal Mach Intell 20:23–38
44. Sampangi RV, Hawkey K (2014) Can't) Lie to ME: Using Micro Expressions for User Authentication. In: Who are you?! Adventures in Authentication: WAY Workshop, Symposium on Usable Privacy and Security (SOUPS) 2014
45. Shen X-b, Wu Q, Fu X-l (2012) Effects of the duration of expressions on the recognition of microexpressions. J Zhejiang Univ Sci B 13:221–230
46. Shreve M, Godavarthy S, Goldgof D, Sarkar S (2011) Macro-and micro-expression spotting in long videos using spatio-temporal strain. In: Automatic Face & Gesture Recognition and workshops (FG 2011), 2011 I.E. international conference on, pp 51–56
47. Song Y, Morency L-P, Davis R (2013) Learning a sparse codebook of facial and body microexpressions for emotion recognition. In: proceedings of the 15th ACM on international conference on multimodal interaction, pp 237–244
48. Sung K-K, Poggio T (1998) Example-based learning for view-based human face detection. IEEE Trans Pattern Anal Mach Intell 20:39–51
49. Viola P, Jones MJ (2004) Robust real-time face detection. Int J Comput Vis 57:137–154
50. Wang P, Ji Q (2004) Multi-view face detection under complex scene based on combined SVMs. In: ICPR (4), pp 179–182
51. Wang S-J, Yang J, Zhang N, Zhou C-G (2011) Tensor discriminant color space for face recognition. IEEE Trans Image Process 20:2490–2501
52. Wang S-J, Yang J, Sun M-F, Peng X-J, Sun M-M, Zhou C-G (2012) Sparse tensor discriminant color space for face verification. IEEE Trans Neural Netw Learn Syst 23:876–888
53. Wang S-J, Chen H-L, Yan W-J, Chen Y-H, Fu X (2014) Face recognition and micro-expression recognition based on discriminant tensor subspace analysis plus extreme learning machine. Neural Process Lett 39:25–43
54. Wang S-J, Yan W-J, Zhao G, Fu X, Zhou C-G (2014) Micro-expression recognition using robust principal component analysis and local spatiotemporal directional features. In: Workshop at the European conference on computer vision, pp 325–338
55. Wang S, Yan W-J, Li X, Zhao G, Fu X (2014) Micro-expression recognition using dynamic textures on tensor independent color space. In: ICPR, pp 4678–4683
56. Wang Y, See J, Phan RC-W, Oh Y-H (2014) Lbp with six intersection points: reducing redundant information in lbp-top for micro-expression recognition. In: Asian Conference on Computer Vision, pp 525–537
57. Wang S-J, Yan W-J, Li X, Zhao G, Zhou C-G, Fu X et al (2015) Micro-expression recognition using color spaces. IEEE Trans Image Process 24:6034–6047
58. Wang Y, See J, Phan RC-W, Oh Y-H (2015) Efficient Spatio-temporal local binary patterns for spontaneous facial micro-expression recognition. PLoS One 10:e0124674
59. Wang Y, See J, Oh YH et al (2017) Effective recognition of facial micro-expressions with video motion magnification. Multimed Tools Appl 76: 21665. https://doi.org/10.1007/s11042-016-4079-6
60. Warren G, Schertler E, Bull P (2009) Detecting deception from emotional and unemotional cues. J Nonverbal Behav 33:59–69
61. Wu Q, Shen X, Fu X (2011) The machine knows what you are hiding: an automatic micro-expression recognition system. In: international conference on affective computing and intelligent Interaction, pp 152–162
62. Xu F, Zhang J, Wang JZ (2017) Microexpression identification and categorization using a facial dynamics map. IEEE Trans Affect Comput 8(2):254–267
63. Yan W-J, Wu Q, Liu Y-J, Wang S-J, Fu X (2013) CASME database: a dataset of spontaneous micro-expressions collected from neutralized faces. In: Automatic face and gesture recognition (FG), 2013 10th IEEE international conference and workshops on, pp 1–7
64. Yan W-J, Li X, Wang S-J, Zhao G, Liu Y-J, Chen Y-H et al (2014) CASME II: an improved spontaneous micro-expression database and the baseline evaluation. PLoS One 9:e86041

65. Yeasin M, Bullot B, Sharma R (2006) Recognition of facial expressions and measurement of levels of interest from video. IEEE Trans Multimedia 8:500–508
66. Zhang M, Fu Q, Chen Y-H, Fu X (2014) Emotional context influences micro-expression recognition. PLoS One 9:e95018
67. Zhao G, Pietikainen M (2007) Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Trans Pattern Anal Mach Intell 29:915–928

Madhumita Takalkar is a Ph.D. candidate at University of Technology Sydney with particular interests in image processing, computer vision and multimedia applications. She holds a M.E. degree in Computer Science and Engineering from Government Engineering College, Aurangabad (MS), India in 2013 and B.E. degree in Information Technology from P.E.S. College of Engineering, Aurangabad (MS), India (Affiliated to Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS)) in 2007.

Min Xu received the B.E. degree from University of Science and Technology of China, in 2000, M.S. degree from National University of Singapore in 2004 and Ph.D degree from University of Newcastle, Australia in 2010. She is currently a Senior Lecturer at University of Technology, Sydney. Her research interests include multimedia data analytics, pattern recognition and computer vision. She has published over 100 research papers in high quality international journals and conferences.

Qiang Wu received the B.Eng. and M.Eng. degrees from the Harbin Institute of Technology, Harbin, China, in 1996 and 1998 respectively and the Ph.D. degree from the University of Technology Sydney, Australia in 2004. He is an associate professor and a core member of Global Big Data Technologies Centre at University of Technology Sydney. Dr. Qiang Wu's research interests include computer vision, image processing, pattern recognition, machine learning, and multimedia processing. His research outcomes have been published in many premier international conferences including ECCV, CVPR, ICIP, and ICPR and the major international journals such as IEEE TIP, IEEE TSMC-B, IEEE TCSVT, IEEE TIFS, PR, PRL, Signal Processing, Signal Processing Letter. Dr. Qiang Wu also serves several journals and conferences as reviewer including TPAMI, PR, TIP, TCSVT, TSMC-B, CVIU, IVC, PRL, Neurocomputing and EURASIP Journal on Image and Video Processing.

Dr. Zenon Chaczko is a Senior Lecturer and Director of ICT Programs, Faculty of Engineering and Information Technology, the University of Technology, Sydney (UTS). He is also the core member of CRIN - UTS Research Centre for Real-time Information Networks**.** Dr. Chaczko's research interests include Artificial Intelligence, Software Systems Engineering, Cloud and Ambient Computing (Wireless Sensor/Actuators Networks, Internet of Things (IoT), Body/Personal Area Networks), Augmented Reality and Complex Software Systems (Laparoscopic Simulation, Early Detection of Environmental/Medical Anomalies, Morphotronics). Dr. Zenon Chaczko received his Bachelor in Computer Science (Honours, 1st class) and his PhD in Computer and Software Engineering. Dr. Chaczko has been actively conducting research in his fields and has published numerous (~ 200) research papers. Before UTS, he was full-time working for over 20 years in the software and system engineering industry. He held various visiting positions: several Visiting Professorships at the University of Arizona, Tucson, USA; IPN in Mexico City, Mexico; the University of Las Palmas, Canary Islands, Spain; Wroclaw University of Technology, Poland; and University of Applied Science and Engineering, Hagenberg/ Linz, Austria. While on sabbatical at UA in Tucson, he participated in several projects including 24–7 Virtual Student Exchange Software Framework, Laparoscopic Simulation Systems, and Wireless Sensor/Actuators Networks Simulators. He is currently leading research into establishing IoT infrastructure for fire detection and emergency support systems through innovative use of AI/Augmented Reality computation, Middleware, Cloud, and advanced sensing/monitoring/control technologies. He has been developing software and hardware infrastructure based on Sensor/Actuator Grid Cloud for emergency management and remote monitoring. He was a Keynote speaker at multiple conferences.