International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7171
***Article Received**: 22 April 2023 **Revised:** 10 June 2023 **Accepted:** 26 June 2023*

Micro Expression Spotting through Appearance Based Descriptor and Distance Analysis

P.Surekha[1,4], P.Vidya Sagar[2], G.Ramesh[3]
[1]Research Scholar, Department of CSE,
Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P, India
prekha.572@gmail.com
[2]Associate Profressor, Department of CSE
Koneru Lakshmaiah Education Foundation, Vaddeswaram, A.P, India
pvsagar20@gmail.com
[3]Associate Profressor, Department of CSE
[4]Assistant Profressor, Department of CSE
GRIET, Hyderabad, Telangana
ramesh680@gmail.com
Corresponding Author: pvsagar20@gmail.com

Abstract--Micro-Expressions (MEs) are a typical kind of expressions which are subtle and short lived in nature and reveal the hidden emotion of human beings. Due to processing an entire video, the MEs recognition constitutes huge computational burden and also consumes more time. Hence, MEs spotting is required which locates the exact frames at which the movement of ME persists. Spotting is regarded as a primary step for MEs recognition. This paper proposes a new method for ME spotting which comprises three stages; pre-processing, feature extraction and discrimination. Pre-processing aligns the facial region in every frame based on three landmark points derived from three landmark regions. To do alignment, an in-plane rotation matrix is used which rotates the non-aligned coordinates into aligned coordinates. For feature extraction, two texture based descriptors are deployed; they are Local Binary Pattern (LBP) and Local Mean Binary Pattern (LMBP). Finally at discrimination stage, Feature Difference Analysis is employed through Chi-Squared Distance (CSD) and the distance of each frame is compared with a threshold to spot there frames namely Onset, Apex and Offset. Simulation done over a Standard CASME dataset and performance is verified through Feature Difference and F1-Score. The obtained results prove that the proposed method is superior than the state-of-the-art methods.

Keywords- Micro Expressions, Spotting, Local Binary pattern, Facial Alignment, Chi-Squared Distance, F1-Score.

I. INTRODUCTION

Recently, due to the advent of new technologies like computer vision, machine learning, and artificial intelligence, an intelligent Human Computer Interaction (HCI) has become an important part of human lives. The future society will become completely intelligence based and Intelligent HCI will be applied to even the daily activities of human life. In such case an intelligent HCI with emotional attachment not only complete the task successfully but also considers the emotions of users to execute the task. For this purpose, the HCI considers different types of input to analyze the emotional status of a user. Text, speech, and facial expression are the three major sources used for expression synthesis. According to the psychologists, Facial expression is considers as the major source which conveys approximately 55% of expression while speech and text occupied only 38% and 7% respectively [1].

Even though Facial expressions can explore the mental status of people, in some situations, people often deliberately express or disguise particular expressions. In such case, the analysis and prediction of true emotional state becomes tough. Such as kind of expression is called as Facial Micro Expressions (MEs) which are short lived and imperceptible in nature [2]. MEs are voluntarily expressed and can reveal the true emotion of an individual person who willing to conceal, disguise, hide or suppress [3]. Due to the shorter time span, they are difficult to manipulate and reveal the perfect emotional status of a person. In 1966, during the psychotherapy study, Haggard and Isaacs discovered the difficult to identify and short lived facial expressions [4]. Further, Ekman et al. [5] noticed a video showing a conversation between a psychologist and patient in which the patient is in depression and trying to hide that expression with a painful smile. Researchers regard that the people produce MEs as a strong emotions through spontaneous, unconscious and rapid facial movements. Compared with macro or normal facial expressions, MEs reflects emotions in a better way because they consist of true potential emotion information which has great significance in risky situations [6-9]. Different applications of MEs are shown in Figure.1.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7171
***Article Received**: 22 April 2023 **Revised:** 10 June 2023 **Accepted:** 26 June 2023*

(a) (b) (c)

Figure.1 Applications of MEs (a) Lie Detection, (b) Criminal Investigation and (c) Medical Treatment

MEs are short lived and low intensity facial expressions and expressed in such kind of situations where the people tries to hide their feelings. Hence, it is not much easy to detect such kind of emotions. Approximately, the MEs duration ranges from 0.04s to 0.2s [10]. Some of the past studies stated that the duration of ME is less than 0.33s and they won't exceed 0.5s [11]. Next, MEs have very less intensities at the corresponding movements of facial muscles [12] and their involvement lies only in the part of a region. In such constraints the detection of MEs is a big challenging task. At first, Ekman et al. [13] introduced a tool called as "Micro-Expression Training Tool (METT)" which allows the identification of totally seven types of emotions from ME videos. However, Frank et al. [14] determined that the detection performance of METT is very limited and it is approximately 40%. Hence, here is a need to develop an effective MEs recognition model which was main motivation of our work. An automatic MEs recognition model accomplishes two stages; they are ME spotting and ME recognition. In the former stage, key frames (frames with emotion attributes) are extracted from the input video and in the second stage type of emotion is identified. Since ME lies only in few frames, processing an entire video with huge frames introduces computational complexity to the recognition system. Hence, MEs spotting in required in which the precise and accurate identification frames' containing the emotion attributes is carried out.

This paper proposes a new ME spotting framework in three stages; they are Pre-processing, feature extraction and Identification. At pre-processing, the non-aligned or non-frontal faces are aligned into frontal views. For this purpose, a new approach is proposed which extracts landmark regions and computes new tilts the faces through an In-plane rotation matrix. At feature extraction, we applied two texture descriptors namely Local Binary Pattern (LBP) and Local Mean Binary Pattern (LMBP). Finally for key frames identification, we used Feature Difference Analysis (FDA) and applied over the texture descriptors of each frame.

Remaining paper is organized as follows; the details of literature survey abut ME spotting are discussed in 2nd section. Next the details about the proposed three stage ME spotting mechanism is discussed in 3rd section. The discussion about the simulation experiments is done in 4th section and last section provides concluding remarks.

II. LITERATURE SURVEY

For a given video sequences, MEs spotting mechanisms finds the key frames in which the temporal dynamics of micro movements exists. The key frames include three frames namely onset, Apex, and Offset. The onset frame is defined as the frame at which the facial appearance becomes stronger and the contraction of facial muscle is observed. Next, the Apex is the frame in which peak emotion lies. Finally, the offset frame is defined as the frame at which the facial muscles starts relaxation mode and expression becomes neutral. In summary, onset frame is an initiator; apex is the peak and offset of end of MEs [15]. Hence, the identification onset, apex and offset frames are called as ME spotting. In such regard, different authors proposed different methods and they are broadly categorized as Apex spotting and Movement Spotting. In the case of Apex spotting, only apex frame spotted while in the movement spotting, the frames are extracted in the order of "neutral-onset-apex-offset-neutral".

Polikovsky et al. [16], [17] considers the ME spotting problem as a classification and classified each frame into four classes, they are Offset, Apex, Onset and Neutral. For this purpose, they applied 3D Histogram oriented Gradients (3D-HOG) to extract features from every frame and K-means clustering for classifying them. Even though they attained 68% spotting performance, it was tested only on the posed videos which are limited for exact ME. Further, the accomplishment of ME spotting as a classification induces hug complexity.

Moilanen A, Zhao G, Pietikäinen M. [18] proposed a simple method for spotting the rapid facial movements from videos. They analyzed the difference between appearance based features of consecutive frames. Local Binary Pattern (LBP) is adapted to describe appearance based features and frames are extracted based on thresholding of differences. Additionally, they employed spatial information about facial movements to find the temporal locations.

A. K. Davison et al. [19] focused on the detection micro movements from videos and applied Histogram of Oriented Gradients (HOGs) to extract features from each frame. Initially, they pre-processed each frame by cropping and aligning followed by removing noise. Next, each frame is divided into blocks and for each block HOGs are calculated. Then, chi-squared distance is used to find the dissimilarity between spatial appearances between consecutive frames. Then the obtained distances are normalized and peak is detected and used to spot the key frames. D. Patel et al. [20] captured the direction continuity of motion features and used them to detect ME frames. They computed optical Flow Vectors for small sized

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7171
***Article Received**: 22 April 2023 **Revised:** 10 June 2023 **Accepted:** 26 June 2023*

local spatial regions and integrated them to form features of temporal regions. Further, they applied Heuristics to remove the non MEs from videos and appropriately determined the ME frames.

Z. Xia et al. [21] proposed a random walk based probabilistic model to find the ME frames by determining the probability of each frame to be a key frame. The random walk model utilized Adaboost algorithm initially to find the probability and then used correlation between frames. Each frame is described through a geometric descriptor based on procrustes analysis and active shape model. S. J. Wang et al. [22] proposed a Mean Directional Maximal Difference (MDMD) analysis for spotting the MEs. MDMD computes the Optical flow features and determined the maximal magnitude difference in the main direction. Instead of complete frames, MDMD considers block structured facial region to sport MEs.

A. Davison et al. [23] referred a 26 regions based Facial Action Coding System (FACS) to spot the micro movements. For each unit, they extracted temporal features based on 3D HOG model and computed Chis-Square distance to find the subtle motion from local regions. Finally, an adaptive baseline threshold is deigned and determined an automatic peak detector for micro movement's detection. J. Li et al. [24] employed two appearance based methods such as LBP and Local Temporal Pattern (LTP) to describe each frame through its appearance. Here, the LTP features are determined through Principal Component Analysis (PCA) in temporal windows over several local facial regions. Then they classified MEs are spotted through a local classification and global fusion. In the second model, the LDP features are processed through Chi-squared distance and MEs are spotted base on a baseline threshold.

Duque et al. [25] employed a video magnification and Reisz Pyramid methods to sport ME frames. Additionally, to suppress artifacts and delays, they employed a masking and filtering mechanisms to segment the motion of interest. It had shown significant impact on the detection of eye blink movements in MEs. Z. Zhang et al. [26] applied a Composite deep learning model called as SMEConvnet to extract Spatio-temporal features from lengthy videos. Then the feature matrix is processed through a sliding window to spot apex frame. Additionally, at pre-processing, the frames are subjected to alignment and cropping. V. Burni and D. Vitulano [27] introduced new frames called as Frozen Frames which occur just before or immediately after a ME. The frozen frames indicate that the speaker is trying to hide something. These frozen frames can be detected through a simplified version of Adelson and Bergen Energy model [28] for motion perception. The authors identified the ME frames based on frozen frames which consist of group of frames.

Y. Han et al. [29] proposed a ME spotting method called as Feature Difference Analysis (FDA). FDA depends on the partitioning of a face image into several uniform Region of Interests (ROIs) and computing features. An evaluation method is proposed based on Fisher Linear Discriminant Analysis (LDA) which assigns a weight for each ROI. Next, FDA considered only two features namely LBP and Histogram of Optical Flow (HOOF) independently. Further, they introduced MDMO into FDA [31] and proposed a simple collaborative strategy called as Collaborative Feature Difference (CDA) based on two complementary features such as LBP and MDMO. Here LBP characterize texture information while MDMO characterize Motion information. V. Esmaeili & S. O. Shahdi [30] proposed a new LBP based Texture descriptor called as Cubic LBP which computes the LBP on totally 15 introduced planes. They demonstrated the effectiveness of 15 planes to find the apex frame where the maximum facial movements occur.

III. PROPOSED METHOD

A. Overview

Here in the current section, we explore the complete details of proposed ME spotting mechanism. This method aims at the detection of ME frames in which the maximum emotion persists. This method is simple and effective as it won't consider the spotting as a classification problem. In summary, the proposed method works as; for a given ME video, initially this method preprocesses each frame and aligns the facial region. The alignment is done with the help of landmark points identified through Viola Jones algorithm. Next, appearance based features are extracted from each frame and fed to a feature difference analysis technique to spot the ME frames. Figure.2 shows the overall working schematic of proposed ME spotting mechanism.

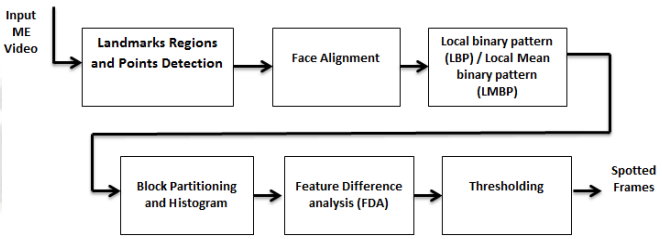


Figure 1. Overall architecture of Proposed ME spotting method

B. Landmarks Detection

In computer vision, facial landmarks have great significance due to their necessity in different applications including Face recognition, face expression recognition or emotion recognition etc. Moreover, some of the facial landmarks can be used as landmarks to align the facial images which are generally non-aligned in nature. After alignment of face through Landmark regions, they provide a uniform and more information for face

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7171
***Article Received**: 22 April 2023 **Revised:** 10 June 2023 **Accepted:** 26 June 2023*

related applications. From several past studies, it was proved that the system trained with aligned face exhibits better performance. Mainly three regions are determined on the face which can carry most of the emotion related information. They are; two eye's regions and one mouth region. Consider a face with happy emotion, the maximum emotion can be observed at the upper muscles of the mouth. Similarly the disgust can explore the muscle movement at lips. Hence this work considered three Landmark regions (Left Eye Region, Right Eye Region and Mouth) to perform facial alignment.

To find out Landmark regions from facial images we used the most popular Viola Jones algorithm [32]. This algorithm uses pixel analysis in front view facial images. Majorly there are four advantages with Viola zones, they are; 1) Great significance for real time images, 2) Larger true positives and smaller false positives, 3) Can extract face regions even from non-frontal images and 4) Larger detection rate. Majorly, the Viola Jones algorithm is executed in four stages, they are; 1) features election through Haar filter, 2) integral image generation, 3) Adaboost training and 4) Cascading of classifiers. The landmark regions identified through Viola zones algorithm is depicted in the following figures.

(a) (b)

Figure 2. Land mark regions identified through VJ algorithm (a) Face Region and (b) Landmark regions (Left Eye, Right Eye and Mouth)

From the above figure we can see that every Landmark region is marked with a rectangle which was constructed with four values they are x-, y-, coordinates, Length (L) and width (W). Consider one corner of the rectangle is represented with (x_1, y_1, L, W) then the remaining coordinates can be calculated through the following mathematical expressions

$$x_2 = x_1, x_3 = x_1 + W \text{ and } x_4 = x_3 \quad (1)$$

$$y_2 = y_1 + L, y_3 = y_1 \text{ and } y_4 = y_2 \quad (2)$$

Even though there exist several Landmark regions in the facial image, we consider only 3 because the facial alignment depends on eye and mouth regions only.

C. Face Alignment

Facial alignment has a significant role in the MEs spotting. Generally, the MEs are subtitle and spontaneous in nature the frames at maximum emotion information may be non-frontal. In such case, the MEs analysis and determination becomes less effective. Hence they need to be aligned properly. Moreover there are several constraints exist in facial images like head moments and head postures which can affect the performance of MEs analysis. Here we used the above obtained Landmark regions 4 co-ordinates to align the facial region. Here we apply an in-plane rotation over the coordinates to do the facial alignment. Since the three Landmark regions are less affected with the movements of the face, they are used here as a baseline regions. For each region, we derived three Landmark points such as Left Eye Inner Corner (LEIC), Right Eye Inner Corner (REIC), and Nasal Spine Point above the Mouth (NSPM). These three points are computer based on the four coordinates of three Landmark regions. These three Landmark points are almost stable and hence they are chosen for facial alignment. Consider the four coordinates of four corner points of a region $rr(r \in \{LE, RE, M\})$ is defined as (x_1^r, y_1^r), (x_2^r, y_2^2), (x_3^r, y_3^r) and (x_4^r, y_4^r). Based on these four co-ordinate points, the landmark points are measured as follows.

$$(x_{LEIC}, y_{LEIC}) = \left(\frac{x_2^{LE}+x_4^{LE}}{2}, \frac{y_2^{LE}+y_4^{LE}}{2}\right) \quad (3)$$

$$(x_{REIC}, y_{REIC}) = \left(\frac{x_1^{RE}+x_3^{RE}}{2}, \frac{y_1^{RE}+y_3^{RE}}{2}\right) \quad (4)$$

$$(x_{NSPM}, y_{NSPM}) = \left(\frac{x_1^{M}+x_2^{M}}{2}, \frac{y_1^{M}+y_2^{M}}{2}\right) \quad (5)$$

Where (x_2^{LE}, y_2^{LE}) and (x_4^{LE}, y_4^{LE}) are the 2nd and 4th corner coordinates of left eye, (x_1^{RE}, y_1^{RE}) and (x_3^{RE}, y_3^{RE}) are 1st and 3rd corner coordinates of right eye and (x_1^{M}, y_1^{M}) and (x_2^{M}, y_2^{M}) are the 1st and 2nd corner coordinates of mouth region. The landmark points obtained from above equation are used for facial alignment after transforming them through an in-plane rotation matrix. Here the In-plane rotation matrix is defined as a Square Matrix with consists of 4 rotation elements such as E_1, E_2, E_3 and E_4. Mathematically the in-plane rotation matrix (IR) is expressed as

$$IR = \begin{bmatrix} E_1 & E_2 \\ E_3 & E_4 \end{bmatrix} \quad (6)$$

where

$$E_1 = \frac{x_{LEIC}-x_{REIC}}{\sqrt{(y_{LEIC}-y_{REIC})^2+(x_{LEIC}-x_{REIC})^2}} \quad (7)$$

$$E_2 = \frac{y_{REIC}-y_{LEIC}}{\sqrt{(y_{LEIC}-y_{REIC})^2+(x_{LEIC}-x_{REIC})^2}} \quad (8)$$

$$E_3 = \frac{y_{LEIC}-y_{REIC}}{\sqrt{(y_{LEIC}-y_{REIC})^2+(x_{LEIC}-x_{REIC})^2}} \quad (9)$$

$$E_4 = \frac{x_{REIC}-x_{LEIC}}{\sqrt{(y_{LEIC}-y_{REIC})^2+(x_{LEIC}-x_{REIC})^2}} \quad (10)$$

The new Landmark points of the aligned face are calculated through the following expression

$$(x'_n, y'_n) = (x_n, y_n) \times IR^T \quad (11)$$

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7171
***Article Received**: 22 April 2023 **Revised:** 10 June 2023 **Accepted:** 26 June 2023*

Where $n \in \{LEIC, REIC, NSPM\}$, (x'_n, y'_n) denotes the new landmark points of the aligned face and they are used to further process. The nonaligned face and aligned faces are shown in figure 3.

(a) (b)

Figure 3 Results of facial alignment
(a) Non-aligned faces and (b) Aligned face

D. Feature Extraction

After the completion of facial alignment in each frame, they are subjected to feature extraction. Here we employed LBP based feature descriptor to extract appearance based features from each facial frame. Here we proposed a new LBP variant called as composite LBP (C-LBP) which is a combination of LBP [33] and LMBP [34]. LBP was initially introduced in 90s and widely employed in different computer vision related applications like human action recognition, facial expression recognition, texture analysis and objective action etc. According to the standard procedure, the computation of LBP is done as follows; for a Centre pixel surrounded by 8 neighbour pixels in a block radius *r*, it is encoded based on the relation between their pixel intensities. If the pixel intensity of neighbour pixel is greater than the intensity of centre pixel, then the corresponding neighbour pixel is encoded as 1 otherwise it is encode as 0. In such way all pixels are encoded into binary and accumulated in an anti-clockwise direction to Form an eight bit string. Then the eight bit string is encoded decimally to derive LBP value of the centre pixel. For a Centre pixel q_c surrounded by p neighbour pixels on circle of radius r, the LBP is completed as

$$LBP_{r,p}(q_c) = \sum_{n=0}^{P-1} s(q_{r,p,n} - q_c)2^n \quad (12)$$

Where

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (13)$$

Next LMBP is a variant of LBP which considered the mean of the pixel intensities in the radius *r*. Consider is the mean of pixel intensities is Q_p, then LMBP of a center pixel is computed as

$$LMBP_{r,p}(q_c) = \sum_{n=0}^{P-1} s(q_{r,p,n} - Q_p)2^n \quad (14)$$

Where

$$Q_p = \frac{1}{P}\sum_{p=1}^{P} q_{r,p,n} \quad (15)$$

Finally each pixel is represented with two decimal codes; one is through LBP code and another is through LMBP code. To determine efficiency of two texture descriptors, we conduct a simulation study for both LBP and LMBP individually and observations are demonstrated in the result section. The process of LBP and LMBP computation is shown in figure.4 and figure.5 respectively.

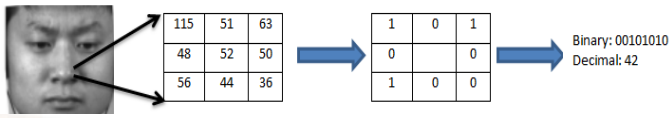


Figure.4 LBP Process

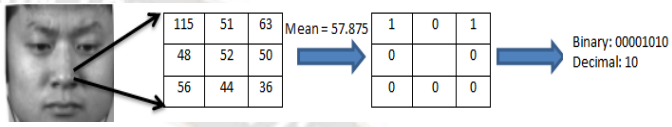


Figure 5. LMBP Process

E. FDA

Once each frame of ME video is represented with either LBP or LMBP then they are processed for Feature Difference Analysis (FDA) to identify the key frames. Here FDA determine totally three frames as ME frames such as onset, Apex and offset. For this purpose FDA considered each frame as a Current Frame (CF) and computed the deviation in motion appearance with its pre and post frames. For the pre and post analysis, we consider two frames such as pre-frame and post-frame as inputs and derived a new frame called as Mean Frame (MF). MF is obtained as an average of pre-frame and post-frame. Now FDA computes Chi-Squared Distance (CSD) between CF and MF which declare the levels of motion variations in the facial area. Moreover, the CSD can determine the Rapid facial moments from temporally lengthy videos. Except for the first and last frames, FDA computes CSD. Here the CSD is computed over the normalized histograms of CF and MF. For the FDA computation, initially the CF and MF are equally divided into several blocks and the histograms are computed for each block. Here the CSD is initially measured between the histogram bins in same block.

Consider C_j^i and M_j^i be the histograms of j^{th} bin in i^{th} block of CF and MF respectively, then the CSD is calculated as

$$\mathcal{X}^2(C_j^i, M_j^i) = \frac{(C_j^i - M_j^i)^2}{C_j^i + M_j^i} \quad (16)$$

Where $\mathcal{X}^2(C_j^i, M_j^i)$ denotes the CSD. Here CSD consider two blocks in CF and MF located at the same position as inputs. Then the obtained CSDs are used to compute an initial difference vector notated as F_i as

$$V_i = \frac{1}{M}\sum_{j=1}^{M} \mathcal{X}^2(C_j^i, M_j^i) \quad (17)$$

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7171
***Article Received**: 22 April 2023 **Revised:** 10 June 2023 **Accepted:** 26 June 2023*

Here V_i explores the difference between i^{th} blocks in CF and MF. Here, we have totally L number of blocks and hence the size of initial difference vector is L. Based on these values, we calculate a local difference vector L_i as

$$L_i = V_i - \frac{1}{2}(V_{i+k} - V_{i-k}) \quad (18)$$

Based on the obtained L_i, we compute a threshold (T) which determines the motion threshold. Mathematically, the threshold is calculated as

$$T = L_{mean} \mp (L_{max} - L_{mean}) \quad (19)$$

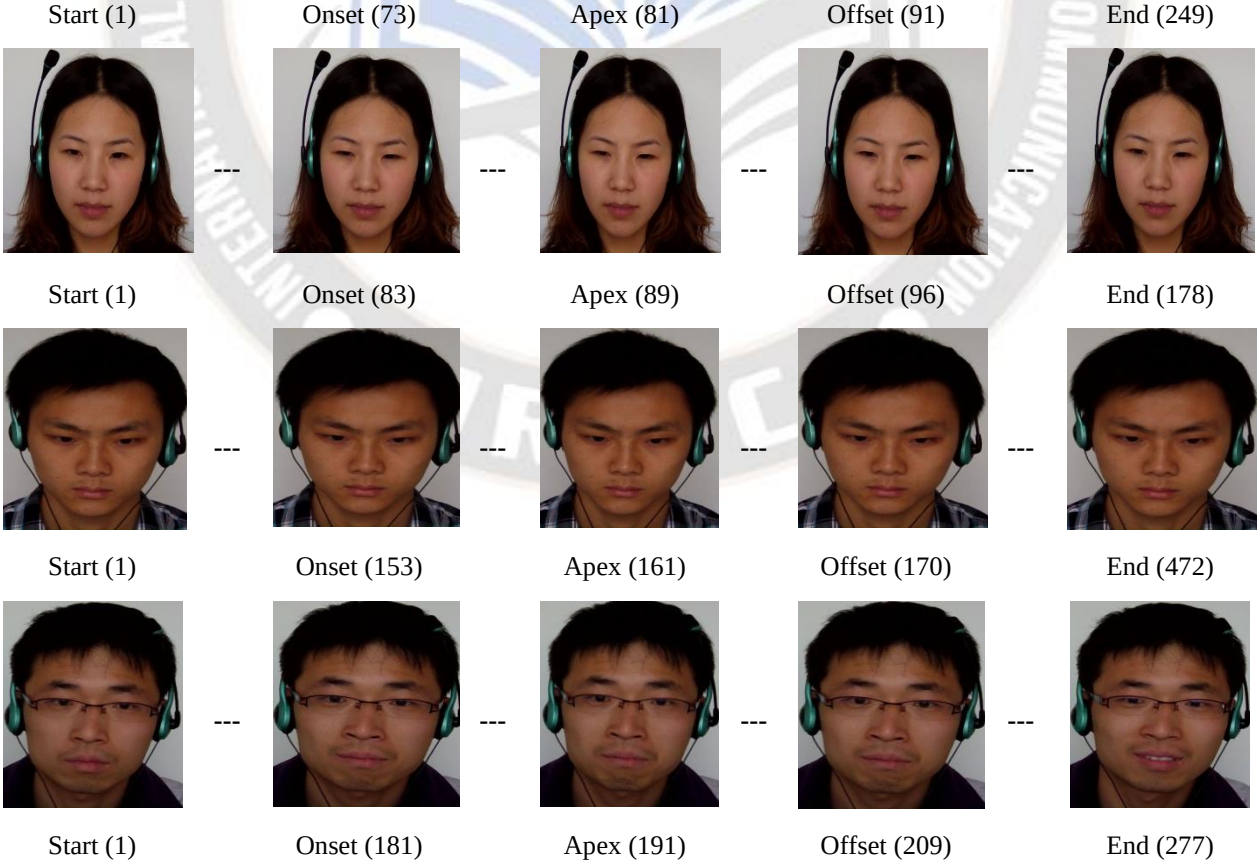
Based on the Threshold, the key frames are identified. The frames those L_i value more than the threshold are considered as spotted frames. Among the spotted frames, the firs frames is considered as onset frames, last frame is considered as offset frame and the center frame is considered as Apex frame.

IV. EXPERIMENTAL ANALYSIS

A. Dataset and Simulation set up

For experimental validation of our method, we applied it on the most widely used ME dataset named as "Chinese Academy of Sciences Micro-Expressions (CASME)" [35]. CASME consists of totally 1965 ME video clips and the frame rate of each video clip is 60 frames per second (fps). Approximately more than 1500 facial movements are referred to acquire these samples. Every sample is encoded with five attribute namely Action Units, Emotion labels, Onset, Apex and Offset frames. Totally 35 subjects are participated under the creation of this dataset and among them 22 are male and 13 are female. The mean age of subjects is identified as 22.03 years with a standard deviation of 1.60. The maximum duration of each video clip is maintained not more than 500ms. Additionally, some video clips even with more than 500s are also considered but they have onset duration less than 250ms because the expressions with fast onset duration can also be regarded as MEs.

Two different environments are used to acquire the CASME dataset and hence the entire dataset is classified into two classes namely Class A and Class B. The video clips under class A are acquired with the help of a camera called as BenQ31 and frame rate is 60 fps. The resolution of each frame is maintained as 1280×720. All the participants are kept under normal lightening conditions. On the other hand, class B video samples are acquired through GRAS-03K2C camera with the frame rate of 60 fps and resolution of as 610×480. All the participants are kept in a room with two LED lights. The video clips acquired through subjects numbering from 1 to 7 are kept in Class A and from 8 to 19 are kept in Class B. Some of ME video samples from b0th classes are shown in Figure 6.



International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7171
***Article Received**: 22 April 2023 **Revised:** 10 June 2023 **Accepted:** 26 June 2023*

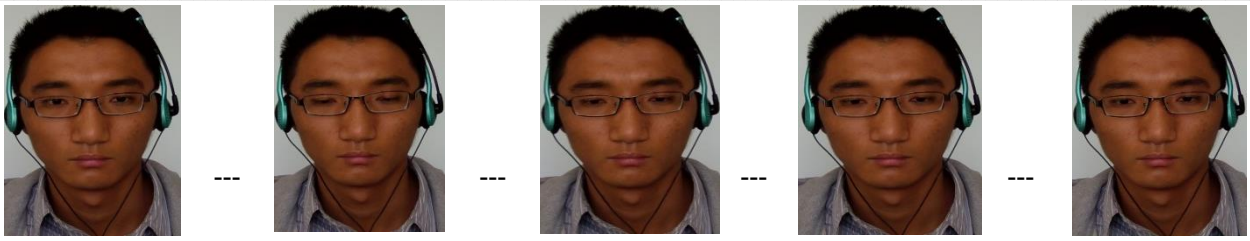


Figure.6 Some ME video samples of CASME dataset

B. Performance Metrics

Under the performance evaluation, we consider three performance metrics to assess the effectiveness of proposed spotting mechanism. The three metrics are namely Recall, Precision and F1-Score. These performance metrics are measured based on True Positives which are measured as follows;

$$TP = \frac{(I_{Spotted}) \cap (I_{Groudtruth})}{(I_{Spotted}) \cup (I_{Groudtruth})} \geq k \qquad (20)$$

Where $I_{Spotted}$ is the posted interval and $I_{Groudtruth}$ is ground truth interval. These two intervals defines the frame sin the period of $onset - offset$. The numerator in Eq.(20) indicates the commonality of interval between spotted and ground truth interval. As much as high the commonality, the fraction will be high and it is compared with a threshold *k*. The *k* value is set as 0.5 and if the obtained fraction value is greater than 0.5, then the spotted interval is considered as True Positive. In some ME video clips, there exists more number of spotting intervals. Consider there are *m* ground truth intervals and *n* spotted intervals and let *TP=a*, and then the FP is calculated as $n - a$ and FN is calculated as $m - a$, hen recall, precision and F1-score are calculated as follows;

$$Recall = \frac{a}{m} \text{ and } Precision = \frac{a}{n} \qquad (21)$$

And

$$F1 - score = \frac{2a}{m+n} = \frac{2TP}{2TP+FP+FN} \qquad (22)$$

C. Results

For a given input ME video clip, the spotted frames are considered as true positives if they lies within the specified ground truth range. Here we extended the range of detection slightly more to balance the annotation's uncertainty. Based on the available ME frames annotations, a frame is considered as correctly identified if it lies within the range of $\left[Onset - \left(\frac{N}{4}\right), Offset + \left(\frac{N}{4}\right)\right]$ where N denotes the maximum length of ME video clip. Figure.7 and Figure.8 shows a sample demonstration about the Feature difference between successive frames for sample video clips from CASME dataset.

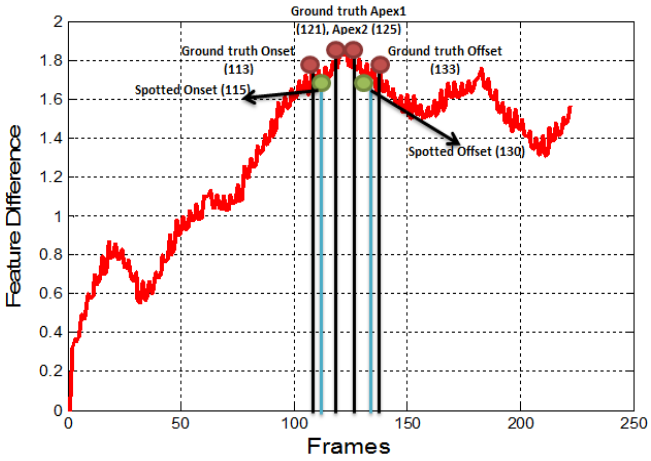


Figure 7. Feature Difference ME video clip (EP01_5) of CASME-A Dataset

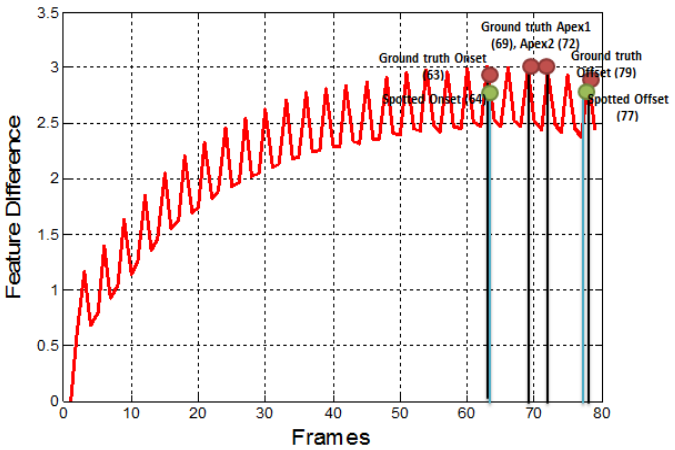


Figure 8. Feature Difference ME video clip (EP12_11_1) of CASME-B Dataset

As seen from Figure.7, the feature difference increasing up to certain level and there onwards it is decreasing. Based on the increment, it can be understood that the feature difference rises gradually due to the gradual rise in the facial muscle movements. It reaches to maximum values which indicate the presence of apex frame. Based on the FDA, we found that the spotted onset frame is 115 and offset frame is at 130. The spotted range is

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7171
***Article Received**: 22 April 2023 **Revised:** 10 June 2023 **Accepted:** 26 June 2023*

within the ground truth range (onset-113 and offset - 133), hence it is considered as True Positive. Similarly, the test result of one more sample ME video (EP12_11_1) from CASME-B is show in Figure.8. For this video clip, the Ground truth onset and offset frames are defined at frames 63 and 79 respectively. The proposed approach spotted the onset and offset frames at 64 and 77 respectively. The obtained range is within the ground truth range, hence it is counted under TP.

Table 1. Spotting Performance through F1-Score at Different Configurations with and without facial alignment

Method	Configuration (R, P)	F1-Score (%)
LBP	(R = 1, P = 8)	60.2210
	(R = 2, P = 8)	68.4530
	(R = 2, P = 12)	**70.3520**
	(R = 2, P = 16)	65.4420
	(R = 3, P = 12)	60.4520
	(R = 3, P = 16)	62.3380
LMBP	(R = 1, P = 8)	62.2210
	(R = 2, P = 8)	64.2530
	(R = 2, P = 12)	66.2000
	(R = 2, P = 16)	**68.3340**
	(R = 3, P = 12)	62.5230
	(R = 3, P = 16)	62.8610
Proposed	(R = 1, P = 8)	69.4810
	(R = 2, P = 8)	72.1420
	(R = 2, P = 12)	**75.3520**
	(R = 2, P = 16)	72.5420
	(R = 3, P = 12)	69.3360
	(R = 3, P = 16)	70.1280

Table.1 explores the spotting performance comparison of proposed method and existing LBP and LMBP. AT this simulation study, we simulated the ME video clips of CASEM dataset through three methods; they are LBP without prepressing, LMBP without preprocessing and proposed (LBP and LMBP with preprocessing). In the first two cases, we processed the frames for LBP and LBP without aligning the faces while in the third case study, we applied facial alignment over each frame and then processed for LBP and LMBP. Hence, the proposed method (Third cases study) has gained better F1-Score than the remaining case studies. Furthermore, a one more case study is also conducted by varying the values of P and R which are called as LBP variables. Here P denotes the number of pixels considered and R denotes the radius. For example R = 1 and P = 8 means, the eight neighbour pixels are considered which are present at a distance of R=1 from center pixel. Similarly, for R = 2 and P = 16 denotes totally sixteen neighbour pixels which are located at a radius of R = 2. From the results, it was observed that the LBP without pre-processing performed well for the configuration of (R = 2, P = 12) while LMBP performed well for the configuration of (R = 2, P = 16). Similar to LBP, the proposed method shows better performance for the configuration of (R = 2, P = 12) only. The average F1-Score of LBP, LMBP and Proposed methods is observed as 70.3520%, 68.3340% and 75.3520% respectively.

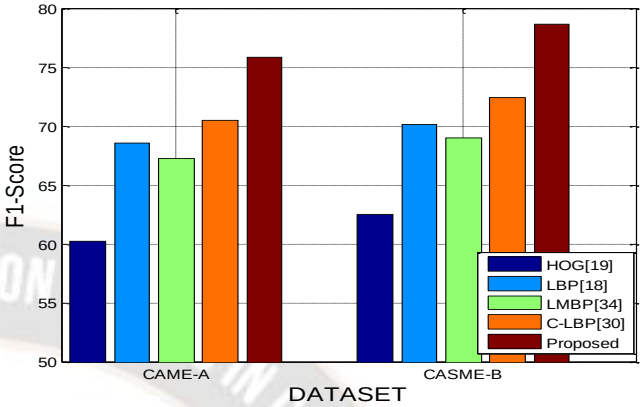


Figure 9. F1-score comparison between proposed and existing methods over CASME dataset

Spotting performance is evaluated through F1-Score and it is shown in Figure.9 where the proposed method gained better value for both CASME A and CASME B. The F1-score for CASME-A and CASME-B is approximated as 75.80% and78.68% respectively. Due to the high resolution of video frame sin CASME-B dataset, it contributed towards more accurate spotting than CASME-A. On the other hand, among the existing methods, HOG had shown poor performance because it cannot expose the minor changes in pixel intensities which are major attributes of MEs. Next, Compared to HOG, LBP and its variants have gained better performance and among them the recently proposed Cubic-LBP had shown better spotting performance. The Cubic-LBP considered totally fifteen planes at the computation of a LBP value for each pixel in the frame. These planes show their effectiveness at the determination of apex frame where the maximum facial muscle movements persists. However, they had experienced a limited performance at some ME video clips, for example the ME video clip with name 'EP02_2' belongs to Subject 12 in CASME-B. In such kind of MEs vide clips, initially the faces needs to be aligned properly because they may put a non-frontal view at the Apex frame. This advantage is present with the proposed method and hence, it had gained better spotting performance than all the existing methods.

V. CONCLUSION

The major concentration of this paper to extract the key expressive frames from a ME video clip intern called as ME spotting. For this purpose, a new method is proposed which constitutes three stages; they are preprocessing, feature extraction and feature difference analysis. Under pre-processing, the non-aligned or non-frontal faces are aligned with the help of

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7171
***Article Received**: 22 April 2023 **Revised:** 10 June 2023 **Accepted:** 26 June 2023*

three standard landmark points derived from three landmark regions. For feature extraction, we employed two texture descriptors namely LBP and LMBP. Finally for spotting the frames, this work computes Chi-Square distance between successive frames and compared with a threshold. The frames whose CSD is less than threshold are extracted as spotted frames. Experimental evaluation is done over CASME dataset under different scenarios and the performance is measured through F1-score. The obtained spotted results prove that the proposed method is much better than several existing methods.

REFERENCES

[1] Mehrabian A. Communication without words. Psychological Today, 1968, 2(6): 53–55.

[2] Ekman, P. (2009b). Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage (revised edition).WWNorton & Company.

[3] P. Ekman, ``Darwin, deception, and facial expression,'' Ann. New York Acad. Sci., vol. 1000, no. 1, pp. 205-221, 2003.

[4] Haggard E A, Isaacs K S. Micromomentary facial expressions as indicators of ego mechanisms in psychotherapy. In: Methods of Research in Psychotherapy. Boston, MA, Springer US, 1966, 154–165.

[5] Ekman P, Friesen W V. Nonverbal leakage and clues to deception. Psychiatry 1969, 32(1):88–106.

[6] Ekman P. Lie catching and micro-expressions. The Philosophy of Deception, 2009, 1(2): 5

[7] O'Sullivan M, Frank M G, Hurley C M, Tiwana J. Police lie detection accuracy: the effect of lie scenario. Law and Human Behavior, 2009, 33(6): 530

[8] Matsumoto D, Hwang H S. Evidence for training the ability to read micro-expressions of emotion. Motivation and Emotion, 2011, 35(2): 181–191.

[9] Frank M, Herbasz M, Sinuk K, Keller A, Nolan C. I see how you feel: training laypeople and professionals to recognize fleeting emotions. In: The Annual Meeting of the International Communication Association. Sheraton New York, New York City, 2009, 1–35.

[10] Ekman P, Friesen W V. Constants across cultures in the face and emotion. Journal of Personality and Social psychology, 1971, 17(2): 124–129.

[11] Yan W J, Wu Q, Liang J, Chen Y H, Fu X L. How fast are the leaked facial expressions: the duration of micro-expressions. Journal of Nonverbal Behavior, 2013, 37(4): 217–230.

[12] Pande, S. D. ., & Ahammad, D. S. H. . (2021). Improved Clustering-Based Energy Optimization with Routing Protocol in Wireless Sensor Networks. Research Journal of Computer Systems and Engineering, 2(1), 33:39. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/17

[13] S. J. Wang, W. J. Yan, X. Li, G. Zhao, and X. Fu, (2014), "Micro-expression recognition using dynamic textures on tensor independent color space", in ICPR, 2014.

[14] Ekman, P. (2002). Micro-expression Training Tool (METT). University of California, San Francisco, CA.

[15] Frank, M. G., Maccario, C. J., and Govindaraju, V. (2009b). "Behavior and security," in Protecting Airline Passengers in the Age of Terrorism, eds P. Seidenstat and X. Francis, and F. X. Splane (Santa Barbara, CA: Greenwood Pub Group), 86–106.

[16] Valstar, M. F., and Pantic, M. (2012). Fully automatic recognition of the temporal phases of facial actions. IEEE Trans. Syst. Man Cybern. Part B 42, 28–43.

[17] Polikovsky, S., Kameda, Y., and Ohta, Y. (2009). "Facial micro-expressions recognition using high speed camera and 3d-gradient descriptor," in 3rd International Conference on Crime Detection and Prevention (ICDP 2009) (London, UK), 1–6.

[18] Polikovsky, S., and Kameda, Y. (2013). Facial micro-expression detection in hi speed video based on facial action coding system (facs). IEICE Trans. Inform. Syst. 96, 81–92.

[19] Moilanen A, Zhao G, Pietikäinen M. Spotting rapid facial movements from videos using appearance-based feature difference analysis. In: 2014 22nd international conference on pattern recognition. IEEE, 2014, 1722–17.

[20] A. K. Davison, M. H. Yap and C. Lansley, "Micro-Facial Movement Detection Using Individualised Baselines and Histogram-Based Descriptors," 2015 IEEE International Conference on Systems, Man, and Cybernetics, Hong Kong, China, 2015, pp. 1864-1869,

[21] Paul Garcia, Anthony Walker, Luis Gonzalez, Carlos Pérez, Luis Pérez. Improving Question Generation and Answering Systems with Machine Learning. Kuwait Journal of Machine Learning, 2(2). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/187

[22] Shaik, D. ., & Santhosh Gollapudi, S. K. . (2023). Analogy of Distinct Constructions of FinFET GDI Full Adder. International Journal of Intelligent Systems and Applications in Engineering, 11(1s), 120–135. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2484

[23] Patel D, Zhao G, Pietikäinen M. Spatiotemporal integration of optical flow vectors for micro-expression detection. In: International conference on advanced concepts for intelligent vision systems. Springer, 2015, 369–380.

[24] Xia Z, Feng X, Peng J, Peng X, Zhao G. Spontaneous micro-expression spotting via geometric deformation modeling. Computer Vision and Image Understanding. 2016, 147: 87–94.

[25] Wang S J, Wu S, Qian X, Li J, Fu X. A main directional maximal difference analysis for spotting facial movements from long-term videos. Neurocomputing, 2017, 230: 382–389.

[26] Davison A, Merghani W, Lansley C, Ng C C, Yap M H. Objective micro-facial movement detection using facs-based regions and baseline evaluation. In: IEEE International Conference on Automatic Face and Gesture Recognition (FG). IEEE, 2018, 642–649

[27] Li J, Soladie C, Seguier R, Wang S J, Yap M H. Spotting micro-expressions on long videos sequences. In: 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019). IEEE, 2019, pp.1–5.

[28] Duque, C., Alata, O., Emonet, R., Legrand, A.-C., and Konik, H. (2018). "Microexpression spotting using the Riesz pyramid," inWACV 2018 (Lake Tahoe).

[29] Zhihao Zhang, Tong Chen, Hongying Meng, Guangyuan Liu, and Xiaolan Fu, "SMEConvNet: A Convolutional Neural Network for

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7171
***Article Received**: 22 April 2023 **Revised:** 10 June 2023 **Accepted:** 26 June 2023*

Spotting Spontaneous Facial Micro-Expression from Long Videos", IEEE Access, Vol. 6, 2018, pp.71143-71151.

[30] V. Burni and D. Vitulano, "A Fast Preprocessing Method for Micro-Expression Spotting via Perceptual Detection of Frozen Frames", J. Imaging 2021, 7, 68, pp1-17.

[31] Adelson, E.H.; Bergen, J.R. Spatiotemporal energy models for the perception of motion. J. Opt. Soc. Am. A 1985, 2, 284–299.

[32] Y. Han, B. Li, Y. -K. Lai and Y. -J. Liu, "CFD: A Collaborative Feature Difference Method for Spontaneous Micro-Expression Spotting," 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 2018, pp. 1942-1946.

[33] Mei Chen, Machine Learning for Energy Optimization in Smart Grids , Machine Learning Applications Conference Proceedings, Vol 2 2022.

[34] V. Esmaeili & S. O. Shahdi, "Automatic micro-expression apex spotting using Cubic-LBP", Multimedia Tools and Applications (2020) 79:20221–20239.

[35] Zhihao Zhang, Fan Mo, Ke Zhao, Tong Chen, and Xiaolan Fu, "TSW-FD: A Novel Temporal and Spatial Domain Weight Analysis of Feature Difference for Micro-Expression Spotting", Journal of Physics: Conference Series, 1828, 2021.

[36] Viola, P.; Jones, M. (2001). "Rapid object detection using a boosted cascade of simple features". Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001.

[37] C., Kertész: Texture-Based Foreground Detection, International Journal of Signal Processing, Image Processing and Pattern Recognition (IJSIP), Vol. 4, No. 4, 2011

[38] Qing Xue et al., "Improved LMBP algorithm in the analysis and application of simulation data," 2010 International Conference on Computer Application and System Modeling (ICCASM 2010), Taiyuan, 2010, pp. V6-545-V6-547.

[39] 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, IEEE, pp 1–7.