Composite Binary Pattern Assisted Micro-Expressions Spotting Through Feature Difference Analysis

Sammaiah Seelothu[1], Dr. K. Venugopal Rao[2]

[1]*Research Scholar, Department of Computer Science Engineering, Jawaharlal Nehru Technological University, Hyderabad, Telangana, India*

[2]*Professor, Department of Computer Science Engineering, G Narayanamma Institute of Technology and Science, Hyderabad, Telangana, India.*

Corresponding Author Email: sammaiahmanuu@gmail.com

Abstract: *Micro-Expressions (MEs) are one kind of facial movement which is very spontaneous and involuntary in nature. MEs are observed when a person attempts to hide or conceal the experiencing emotion in a high stakes environments. The duration of ME is very short and approximately less than 500 milliseconds. Recognition of such kind of expressions from a lengthy video consequences to a limited Micro Expression Recognition Performance and also creates he computational burden. Hence, in this paper, we propose a new ME spotting (detection of ME frames) method based on a new texture descriptor called as Composite Binary Pattern (CBP). As a pre-processing, we employ viola jones algorithm for landmark regions detection followed by landmark points detection for facial alignment. Next, every aligned face is described through CBP and subjected to feature difference analysis followed by threshold for ME spotting. For simulation, MEVIEW dataset is used and the performance is measured through Recall, Precision and F-Score.*

Keywords: **Micro Expressions, Face alignment, Composite Binary Pattern, Feature difference, MEVIEW dataset, Recall.**

1. INTRODUCTION

Facial Micro-Expression (MEs)are involuntary, subtle and very brief facial movements those usually happen when somebody either unconsciously or deliberately obscuresher or his honest feelings or emotions. Generally, the MEs occur when a person experiences an emotion but intentionally attempts to conceal his or her feelings [1]. MEs occur in highly risk environments because there is more danger to show the true feelings [2]. In recent years, the study and analysis of MEs has gained an increased attention of psychologists and affective computing researchers due to its widespread applicability in several fields like security, criminal interrogations, business negotiations and clinical diagnosis [3] etc.The first most effort to enhance the capability of human beings at the recognition of MEs is developed by Ekman [4] in which a tool called as “Micro-Expression Training Tool (METT)” was developed. METT provides ability to human to identify totally 7 different types of MEs. But, Frank et al. [5] found that the METT has attained a detection performance of only 40% when it was implemented with the help of undergraduate students. In one more study with U. S. coast guards, the METT has gained more than 50%. Hence there is a necessity to develop an automatic ME Recognition (MER) system for the detection of MEs those are expressed in typical and lays conditions, particularly with the recent progressions in parallel functionalities with multi-core and computational power applications.

Full Facial expression recognition has been widely studied in the field of computer vision. However, compared to the full facial expression recognition, MER constitutes so many difficulties. Compared to the normal expression, the Micro-expressions have the following characteristics that make MER quite difficult from full facial expression recognition. 1. Micro-expressions are rapid facial movements, typically exists for a very little time span and it is approximately ranges from 0.4 second to 0.04 to 0.2 seconds. 2. The intensity of short-lived micro-expressions is very low in terms of facial muscles’ movement [6, 7] and 3. Micro-expressions typically involve a fragment of the facial region. In the

presence of such kinds of events, the recognition of MEs is very tough for human beings with naked eye.

Automatic MER system is accomplished in two stages; (1) Spotting of ME and (2) Recognition of ME. The former one denotes to issue of identifying the time span of micro movements in the given emotional video with several frames. The second task refers to the classification of the MEs existed in the input emotional video sequence. Processing entire fames of a video sequence to ME recognition system constitutes a huge processing overhead. Hence an accurate and precise identification of frames containing the micro movements is required for further analysis. Hence automatic MEs spotting methods are developed automatically to identify the temporal dynamics of MEs in the video sequence.

In this paper, we propose a novel MEs spotting framework which can identify the temporal interval of MEs with more precision and accuracy. Our method is a simple and effective method which concentrates on the coordinates of landmarks regions of facial images for MEs spotting. The complete methodology is employed in four phases, (1) Landmark region detection through Viola Jones algorithm, (2) Face alignment through landmark points, (3) Texture invariant feature extraction through a Composite Binary Pattern (CBP) and (4) Feature difference analysis and thresholding. The proposed approach is an adaptive and it can identify the MEs even in unknown ME videos.

Rest of the paper is organized as follows; the details of literature survey are explored in section II. The overall details of proposed ME spotting mechanism is discussed in section III. Section IV explores the details of simulation experiments and the section V provides concluding remarks.

2. LITERATURE SURVEY

Automatic Micro expression spotting methods tries to find out the temporal dynamics of Micro Movements of facial expressions in the video sequence means, the discovery of time span of micro movements of facial expressions that include four frames such as Start (Onset), Peak (Apex), End (Offset) and Neutral. According to [8], the onset is defined as the phase at which the contraction of facial muscles is observed and the variations in facial appearance become stronger. Next theApex fames of Peak frame is defined as the phase at which the peak emotion lies. Further the offset is defined as the phase at which the relaxation of facial muscle is observed and the returning of face to neutral mode is observed. At offset phase, we can observe the facial muscles with only a small or even no activations. In general, the facial emotion is a video lies in the fashion of neutral-onset-apex-offset-neutral. Hence the major task of ME spotting is the identification of onset, apex and offset frames.In this regard, different authors proposed different spotting methods. The entire ME spotting methods are categorized into two categories. Hey are Movement Spotting and Apex spotting. In the former category, the authors tried to find the order of “neutral-onset-apex-offset-neutral” frames while in the second category, the authors tried to get only apex frame.

Polikovsky et al. [9, 10] adopted 3D Histogram oriented Gradients (HOG)and K-means clustering for feature extraction and clustering the extracted features to specific Action Units (AUs). In this approach, the ME spotting is treated as a classification task and each frames is classified into four categories such as Offset, Apex, Onset and Neutral. Further to evaluate the performance, the results are compared with the labels of ground truth frames. Even though it attained a spotting performance of 68%, it was tested on the posed facial ME videos which don’t have exact nature of ME. Further, the running of experiment as a classification task is not all suitable for ME spotting.

Wu et al. [11] also treated the spotting of MEs as a classification task and employed Gabor filer for feature extraction and gentle Support Vector Machine (SVM) for classification. Based on the label of resultant frame, the MEs were computed according to the transition points and frame rate. For METT trained database, this approach has gained a superior spotting performance but they have simulated by inserting a synthesized frame of ME in the middle of video sequence. In real world, the transition of frames of MEs is much more abrupt.

Unlike the above method, Shreve et al. [12, 13] employed to assess the temporal relation between the frames and applied thresholding to find the MEs with instantaneous movements. This is first contribution towards the spotting the MEs without using machine learning and also worked on both macro and micro expressions. In their work, they employed a new image deformation model called as optical strain is calculated from the facial regions those were chosen. Afterwards spotting of MEs is done by tracing the magnitudes of optical strains through the frames through (1) the magnitude of optical strain is larger than the threshold and (2) the duration of leak frame lasts at most $1/5^{th}$ of a second. In their further work [14], they employed a peak detector to locate the MEs sequences with the help of strain maps. But, there is no much information about the thresholding and peak frames detection.

Moilanen et al. [15] modeled the changes in facial frames through feature difference analysis of appearance based features extracted by Local Binary Pattern (LBP). In this approach, for feature difference analysis, they have used Chi-Squared distance and computed between the current frame and average feature frame that was obtained by averaging the head and tail frames. A contrasting difference vector is measured to determine the relevant peaks from the sequence. On the simulation over CASME-A and CASME-B, the true positive rate is observed as 52% and 66% respectively. The similar ME spotting mechanism was implemented by Li et al. [16] and it was tested over several ME datasets like CASME-II, SMIC-E-HS, SMIC-E-VIS, and SMIC-E-NIR. Based on the obtained results, they stated that the LBP had shown an outstanding performance compared to the optical flow vectors. Next, for spotting the MEs in a newly created $CAS(ME)_2$ dataset, Wang et al. [17] employed the same approach that was employed by [15]. With the help of their proposed "Main Directional Optical Flow (MDMD)" approach, they gained an average detection rate, precision, F-score after the implementation over $CAS(ME)_2$ dataset is 32%, 35% and 33% respectively.

Davinson et al. [18] applied HOG as a feature descriptor for ME spotting in a high speed in-house video dataset. This approach initially aligns and crops the face regions from all frames of video sequence through automatic facial point detection and affine transformation. Next, the facial region of every frame is subjected to block division and then to HOG. The chi-Squared dissimilarity check is employed to determine the spatial appearance between frames. In their later works [19, 20], they introduced a 'individualized baselines' those were measured by considering a neutral video sequence for the participant through Chi-Squared distance to obtain a preliminary features for the sequence of baseline. The threshold is determined as the maximum value of the baseline feature.

Xia et al. [21] proposed a probabilistic method for the detection of ME spots in the video sequence. In this probabilistic model, a random walk model is employed for the computation of probability of individual frames to be ME frames. The Adaboost model is used for the prediction of initial probability for each frames and the correlation between frames. Further, to describe the geometric shape of human face, an active shape model along with procrustes analyses are employed which makes the system robust for hear postures and light variations.

Duque et al. [22] proposed to design a workflow for the automatic ME spotting based on the Riesz pyramid and video magnification method. In addition, they proposed a filtering and masking schemes to segment the motion of interest without introducing additional artifacts of delays. This approach is able to provide a sufficient differentiation between eye movements and MEs.

Z. Zhang et al. [23] proposed a new Convolutional Neural Network model for spotting MEs from long videos. The complete methodology is accomplished in three stages. They are alignment and cropping of video, extraction of high level features from processed long videos through SMEConvNet and processing the feature matrixes further though a sliding window.

3. PROPOSED APPROACH

3.1 Overview

In this paper, we propose a simple and effective method for spotting of Micro Expressions. This method finds the frames at which the maximum ME occurred as well as the spatial locations of spotted movements. This method doesnot require pre-labeling or training and it is adaptive for even unknown videos or real time videos. This approach depends on analyzing the differences between appearance based features of a continuous video sequence. Here the complete methodology of ME spotting is done in four phases; (1) landmark points detection through Viola Jones algorithm, (2) Face alignment based on landmark points, (3) Features extraction through Composite LBP and (4) Feature difference analysis followed by thresholding. The overall architecture of develop ME spotting mechanism is depicted in Figure.1.

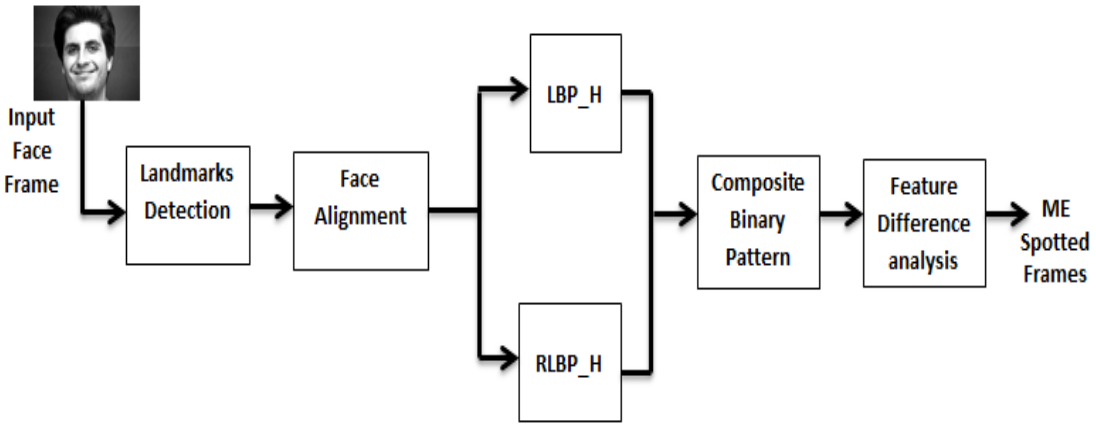

Figure.1 Overall architecture of Proposed ME spotting method

3.2 Landmark Regions Detection

Facial landmarks detection is very important for several applications in computer vision including face recognition, face expression recognition, or emotion recognition etc. Moreover, the facial landmarks are also used in the alignment of facial images to mean a shape of face such that after alignment the location of facial landmarks in all images is appropriately the same. Simply it denotes that a face recognition or expression recognition system trained with aligned faces would perform much better and it was already explored by several research papers. Mainly there are three landmark regions in the face at which the maximum possible expressions varies. For example, consider happy expression, the major variations can be observed at mouth region due to the rise of mouth muscle movements. Similarly for disgust expression, we can see the movements at nasal spine and eyebrows. Hence the detection of landmark regions makes the expression recognition much simple and effective.

For the detection of landmark regions from a face image, we employed the most popular Viola Jones algorithm. This algorithm is a fast technique for landmarks detection. The detection is employed based on the analysis of pixels in photo images of full front faces. The major advantages through the accomplishment of viola jones algorithm are of four fold: (1) High detection rate (almost it extracts with maximum exactness), (2) It can identify the faces from non-faces of arbitrary images, (3) this algorithm has high true positives and less false positives and (4) applicable in real time. Based on the methodology, the viols jones algorithm is employed in four stages, (1) Haar feature selection, (2) creation of an integral image and (3) training of integral images with Adaboost algorithm and (4) Cascading classifiers.

To obtain the facial landmark regions through viola jones algorithm, we used computer vision toolbox in MATLAB software. In the computer vision toolbox, we employed the command called “Vision.CascadeObjectDetector”. For a given input facial image, initially we fetch the required landmark with the help of command “Vision.CascadeObjectDetector”

by specifying the name of required landmark. For example, if we need Right Eye, then the command should be written as **RE = vision.CascadeObjectDetector('RightEye');.** After the extraction of required region, then we extract the x-coordinate, y-coordinate, length and width of the region by applying a step function over the face image through the obtained region. In this manner, we extract the remaining two landmarks region such as left eye and mouth. A simple output of VJ algorithm is shown in Figure.2.

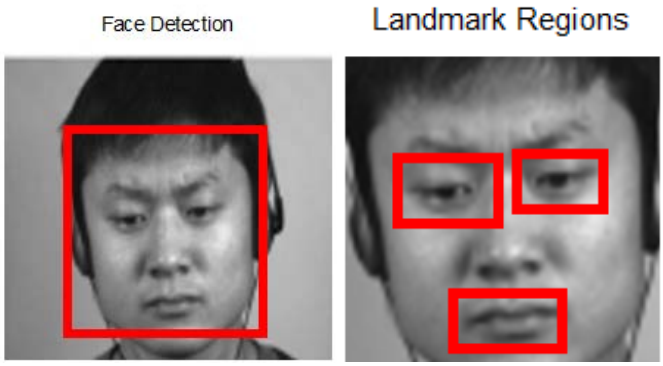

Figure.2 VJ output

As an output, for every landmark region in the facial image, we get totally four values; they are x- coordinate, y-coordinate, length and width of the region. Consider the four values is represented as (x_1, y_1, L, W), then the remaining coordinates of that region is calculated as, $(x_2 = x_1, y_2 = y_1 + L)$, $(x_3 = x_1 + W, y_3 = y_1)$, and $(x_4 = x_3 = x_1 + W, y_4 = y_2 = y_1 + L)$. In this work, even though there are several landmark region are available in facial image, we consider only three regions because our main intention is to perform face alignment through these landmark region. Based on the coordinates of landmark region, we compute the landmark points and applied for facial alignment.

3.3 Face Alignment

Face alignment has significant role in the micro expression recognition. Since the micro expressions are spontaneous and subtle in movements, a proper alignment of facial images is required to get a better performance in MER system. Moreover, the micro expressions are kept within the fraction of seconds and at such instant the face of respective person may be a non-frontal view. Along with these constraints there are some more issues like head postures, head movements etc. which may cause instability in the spotting of micro expressions.

In this paper, the face alignment is done with the help of landmark points those were derived from the coordinates of landmark regions extracted through Viola Jones algorithm. The coordinates of landmark regions are used for face alignment followed by the division of facial area into blocks. Here, based on the present coordinates we apply an in-plane rotation to correct the facial alignment. Here the main reason behind the selection of only three regions is their stability and less effect on the face expression. From these three landmark regions, we derive three landmark points. They are (1) Left Eye (LE) inner corner, (2) Right Eye (RE) Inner corner and (3) Nasal Spine Point above Mouth (M). These three landmark points have less effect on the facial expression and almost they are stable throughout the video.

Consider (x_1^i, y_1^i), (x_2^i, y_2^i), (x_3^i, y_3^i) and $(x_4^i, y_4^i)$are the coordinates of four corner points of a region i where $i \in \{RE, LE, M\}$. Based on these four coordinates, the coordinates of landmark points are calculated as

$$(x_{l_1}, y_{l_1}) = \left(\frac{x_1^{RE} + x_3^{RE}}{2}, \frac{y_1^{RE} + y_3^{RE}}{2}\right) (1)$$

$$(x_{l_2}, y_{l_2}) = \left(\frac{x_2^{LE} + x_4^{LE}}{2}, \frac{y_2^{LE} + y_4^{LE}}{2}\right) (2)$$

$$(x_{l_1}, y_{l_1}) = \left(\frac{x_1^M + x_2^M}{2}, \frac{y_1^M + y_2^M}{2}\right) (3)$$

Where (x_1^{RE}, y_1^{RE}) and (x_3^{RE}, y_3^{RE}) are the coordinates of first and third corners of right eye landmark region, (x_2^{LE}, y_2^{LE}) and (x_4^{LE}, y_4^{LE}) are the coordinates of second and fourth corners of left eye landmark region, and (x_1^M, y_1^M) and (x_2^M, y_2^M) are the coordinates of first and second corners of mouth landmark region. Based on these three landmark points, we perform face alignment. The face alignment removes the influence of head postures over the micro expression recognition. Based on the coordinates of right eye corner and left eye inner corner, we compute a rotation matrix R as follows;

$$R = \begin{bmatrix} R_1 & R_2 \\ R_3 & R_4 \end{bmatrix} (4)$$

Where

$$R_1 = \frac{x_{l_2} - x_{l_1}}{\sqrt{(y_{l_2} - y_{l_1})^2 + (x_{l_2} - x_{l_1})^2}} (5)$$

$$R_2 = \frac{y_{l_1} - y_{l_2}}{\sqrt{(y_{l_2} - y_{l_1})^2 + (x_{l_2} - x_{l_1})^2}} \quad (6)$$

$$R_3 = \frac{y_{l_2} - y_{l_1}}{\sqrt{(y_{l_2} - y_{l_1})^2 + (x_{l_2} - x_{l_1})^2}} (7)$$

$$R_4 = \frac{x_{l_2} - x_{l_1}}{\sqrt{(y_{l_2} - y_{l_1})^2 + (x_{l_2} - x_{l_1})^2}} (8)$$

After the computation of rotation matrix R, the in-plane rotation and facial size variations within the facial region are varied as

$$(x'_{l_i}, y'_{l_i}) = (x_{l_i}, y_{l_i}) * R^T (9)$$

After facial alignment, the new coordinates of three landmark points are obtained as (x'_{l_1}, y'_{l_1}), (x'_{l_2}, y'_{l_2}) and (x'_{l_3}, y'_{l_3}) of Right eye inner corner, left eye inner corner and nasal spine point respectively. Based on these new coordinates we measure the size of each block into which the facial region has to divide. The size (i.e., width and height)of each block is computed as follows;

$$w = \frac{(x'_{l_2} - x'_{l_1})}{2} (10)$$

And

$$h = \frac{(y'_{l_3} - y'_{l_1})}{2} (11)$$

Further, the starting point of each block is computed as

$$SP = (2x'_{l_1} - x'_{l_2}, 2y'_{l_1} - y'_{l_3}) (12)$$

Based on the width, height and starting point, the entire facial region is divided into equal sized blocks. Figure.3 shows a simple demonstration about the process of facial alignment.

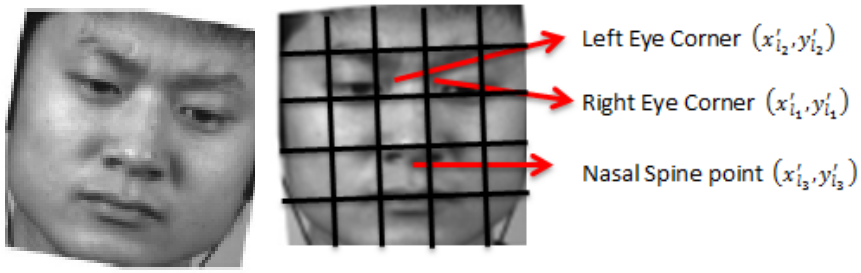

Figure.3 Facial Alignment & block division

3.4 Feature Extraction

Once the facial alignment and partitioning is completed, then we apply feature extraction over every block through Local Binary Pattern (LBP) based method. LBP was first introduced in the mid of 1990's [ ] and has been successfully applied in different fields of computer vision like human action recognition, object detection and texture analysis. Further, it was introduced into the face recognition and classification of micro expressions [ ]. According to the standard methodology, the LBP computation is as follows; for each pixel in the frame, it is compared with its P neighbor pixels along a circle. During this comparison, if the gray value of neighbor pixel is greater than or equal to the gray value of center pixel, then the pout put is 1 otherwise the output is 0. This process derives a *P*-point number that can be represented as a decimal number by weighing with the power of two. Based on this we can state that the LBP characterizes the structures of P pixels, those were distributed evenly in angle over a circle of radius r and centered at pixel q_c. For a center pixel q_c and its P neighbor pixels $\{q_{r,p,n}\}_{n=0}^{P-1}$ on the circle of radius r, the LBP pattern is computed as

$$LBP_{r,p}(q_c) = \sum_{n=0}^{P-1} s(q_{r,p,n} - q_c)2^n \quad (13)$$

Where

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (14)$$

Where $s(.)$ Is a sign function. If the pixels are not observed to be at the center, then they are predicted through bi-linear interpolation. As shown in Figure.4, the decimal value of LBP pattern is given by the binary sequence of circular neighborhood, such as 241 = $(11110001)_2$. LBP is gray scale invariant and it can encode the significant local patterns like blobs, edges and lines, because it measures the differences between center pixel and its neighbor pixels. For a given $M * N$ texture image, the LBP pattern $LBP_{r,p}(q_c)$ is computed for every pixel thereby a textured image can be represented by the distribution of LBP values, and make the image represented by a LBP histogram vector. By altering the value of r and the P, we can encode more information about the center pixel q_c.

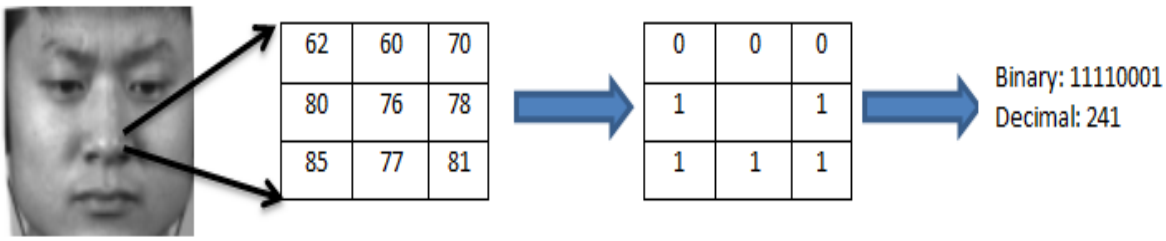

Figure.4 LBP computation

Even though the LBP has gained a better representation in ME, the main drawback is information loss due to the concept of thresholding with center pixel. Further, the proportion of uniform patterns may be too small to acquire the variations. Hence we propose an extended version of LBP called as Radial LBP (RLBP) which encodes the radian information between pixels at different radii. Unlike the conventional LBP which encodes the information between

center pixel and its neighbor pixels on a same circle (single scale), the RLBP considers the pixels on different circles (different scales) for encoding. Due to the consideration of pixels on the single scale, the LBP fails to encode the second order information of neighbor pixels between different circles. For every pixel in the image, we consider two circles or radius r and $r-\delta$ centered on the pixel q_c and P pixels are distributed on each ring evenly. For the computation of RLBP codes, we first measure the radial differences between pixels on two circles and then threshold against 0. Based on the formal definition, the computation of RLBP is done as follows;

$$RLBP_{r,p,\delta}(q_c) = \sum_{n=0}^{P-1} s(q_{r,p,n} - q_{r-\delta,p,n}) 2^n \quad (15)$$

Where r and $r-\delta$ are the radii of outer and inner circles respectively. Figure.5 shows the simple process of RLBP computation.

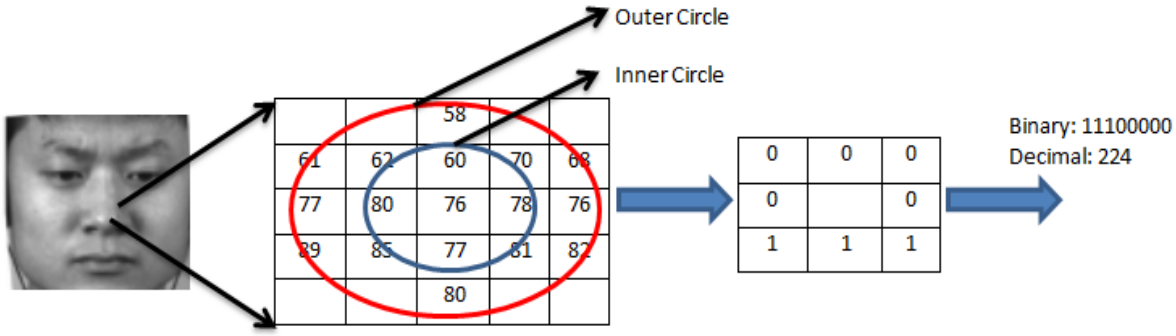

Figure.5 RLBP computation

After the computation of LBP and RLBP for every frame, then we apply histogram computation to extract the histogram features. Finally the obtained histogram of both LBP and RLBP are integrated and formulated into a single feature vector called as Composite Binary Pattern (CBP). Let's consider LBP_H as the histogram of LBP featured image and $RLBP_H$ as the histogram of RLBP featured image, the CBP is obtained by the horizontal concatenation of these two features. In this manner, every frame is represented with CBP vectors and then fed to feature difference analysis to find the exact ME spotting frames.

3.5 Feature Difference Analysis

The main intention to perform the feature difference analysis is to find the final ME spotting frames (onset, apex and offset frames) from an entire ME video sequence. Here we consider three notations such as Back Frame (BF), Front Frame (FF) and Current Frame (CF) for simplified representation. Here if we consider the k as the index of current frame, k-1 is the index of back frame and k+1 is the index of front frame. Expect eh first and last frames, all the remaining frames have BF and FF. For feature difference analysis, we seek the help of Chi-Squared $(\mathcal{X}^2)$ distance and it is computed between CF and Average Feature Frame (AFF). Here the AFF is measured by averaging the BF ad FF. the difference between CF and AFF indicates the level of changes in the facial area. Furthermore, due to rapid and spontaneous movements of ME, this can help to distinguish the rapid facial movements from temporally longer events. The entire process is repeated for entire frames except first and last frames of ME video sequence. For normalized histograms A and B with same number of bins, the Chi-Squared $(\mathcal{X}^2)$ distance is computed as

$$\mathcal{X}^2(A,B) = \sum_i \frac{(A_i - B_i)^2}{A_i + B_i} \quad (16)$$

Where the index i refers to the i-th bon of histogram. Here the calculation of dissimilarity check is done through blocks, means the $\mathcal{X}^2(A,B)$ denotes a two blocks in the CF and AFF at same position. Once the feature difference is computed for every block of each frame, then we compute an initial difference vector by taking an average of M greatest

difference values for each frame. For every frame, M value is calculated as 1/3rd of the entire blocks. For example, if the total number of blocks is 36, then the value of M is 12. Hence 12 blocks with maximum difference is extracted from every frame and the initial difference vector F is computed as

$$F = \frac{1}{M}\sum_{i=1}^{M} d_i (17)$$

Where d_i is the block difference value of ith block. From the obtained F value, we can find the local peaks which can provide the variations in local magnitude and background noise. The newly obtained local peak vector is called as local difference vector L_i and obtained as

$$L_i = F_i - \frac{1}{2}(F_{i+k} - F_{i-k}) (18)$$

The computation of L_i for each and every frame consequences to L_i of entire video sequence. Based on the obtained L_i values we compute the L_{mean} and L_{max} . Based on these values, we compute a threshold T and the local peak vectors of each frame are compared with threshold. The threshold computation is done as

$$T = L_{mean} + (L_{max} - L_{mean}) (19)$$

The set of frames those have higher L_i than the threshold are only selected as ME spotting frames. The first frame in the obtained resultant set of frames is called as onset frames and the end one is called as offset frame. The frame at the mid position is called as Apex frame.

4. EXPERIMENTAL ANALYSIS

In this section, we explore the details of simulation experiments conducted over the proposed model with the help of MEVIEW micro Expression dataset. For simulation purpose, we used the MATLAB 2015 software with image processing toolbox. Initially we explain the details of dataset and then we explore the details of simulation results obtained.

4.1 Dataset

For simulation purpose, we consider a standard ME dataset called as MEVIEW which was acquired by Peter Husak et al. [24]. This dataset consists of videos collected from TV interviewsand poker games. Since the poker game is a stressful game and the candidates try to hide the internal emotions. Players try to fake or conceal their real emotions in which the ME likes to be appearing. The MEs are still rarely available because the post production department removes the most valuable movements such as the details of players face when the cards are being uncovered or somebody raises, folds or calls his/her cards. In this dataset, the average length of each video is observed as 3 seconds. The video clip is captured by switching the camera often on and captures a complete shot with single face. This dataset have totally 31 video clips and they are acquired from 16 subjects. The frame rate of each video is 25 frames per second. Totally five emotions classes (contempt, surprise, fear, anger, happy) are observed in this dataset and every video is FACS encoded. The onset and offset frames of ME are labeled in long videos, FACS coded and the type of emotion is also annotated. Some examples of MEVIEW dataset are shown in the following Figure.6-8.

Figure.6 Video sample of Subject 11 for second time

Figure.7 Video sample of Subject 3

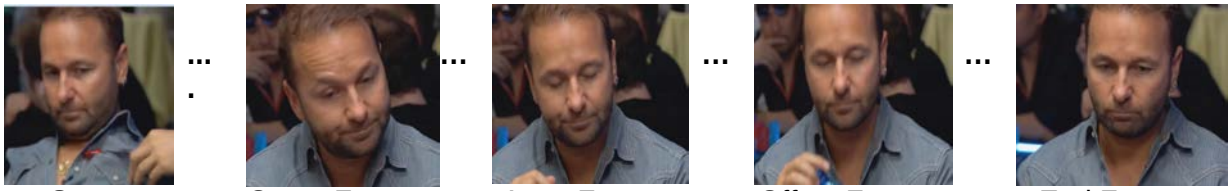

Figure.8 Video sample of Subject 5 for second time

Figure.9 Video sample of Subject 13 for second time

As shown in theFigure.6, the video sample is belongs to subject 11 and acquired from a poker game. According o the annotations available in the dataset, the onset frame is numbered at 57 and offset frame is at 67. Even though, there is no information about the apex frame, for every test video, we defined the apex frame manually by closely observing the entire video samples. The total number of frames present in this video is 69 and the time span is about 2 seconds. Similarly we add more two videos samples those are belong to subject 3 and subject 5. The total number of frames present for the video sample shown in Figure.7 is 138 and for video sample shown in Figure.8 are 174. Finally for the video sample shown in Figure.9, the onset frames is at 6 and offset frame is at 20. The annotations represented at respective samples are taken from the information provided by the authors at the database website.

4.2 Results

For valuation, we consider the entire video samples of MEVIEW dataset. For a given video sample, our method tries to find the ME spotting (frames having Micro expressions). Here the output is the frames of a video sample those have effective micro expressions. With the availability of onset and offset annotations for frames of the ME, then a frame is considered to be correctly detected if it lies in the range of $\left[Onset - \left(\frac{N}{4}\right),\ Offset + \left(\frac{N}{4}\right)\right]$, where N is the maximum considered length of ME. Here we considered the length of N as 8 because the frame rate is 25 Hz. To balance the uncertainty of the annotations, here we have expanded the range to a small margin.

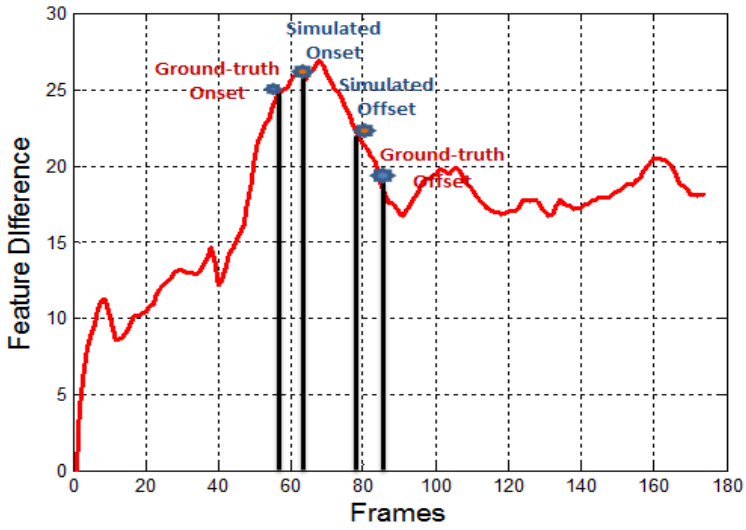

Figure.10 Feature difference analysis and annotations of ground truth and simulated ME spotting

In the above figure.10, we had shown a simple for feature difference analysis and annotated onset and offset frames. This figure is belongs to the video clip of subject 5, where the ground-truth ME spotting is declared as 75-84. From this figure, we can see that the obtained ME spotting interval also lies within the ground-truth range hence it is counted as TP. Moreover, we can also observe that with an increase of frame number along x-axis, the feature difference along y-axis also increasing. From this analysis, we can understand that the ME spotting frames interval has larger difference with first and last frames. Since the ME have peak magnitude sin the spotting interval, the frames shows larger deviations with starting and ending frames.

Next, we analyze the performance analysis of proposed approach through several performance metrics. Here the performance is measured with the True Positives and False Positives. The TP per interval in one video is defined based on the intersection between the spotted interval and ground truth interval. Consider the proposed spotted interval is P_{Spoted} and ground-truth spotted interval is G_{Spoted} , then the TP is measured as

$$TP = \frac{(P_{Spotted}\ )\cap(G_{Spotted}\ )}{(P_{Spotted}\ )\cup(G_{Spotted}\ )} \geq k (20)$$

Where the k value is set as 0.5, and $G_{Spotted}$ represents the ground-truth interval of micro-expression (onset to offset). For a given test video, if the obtained ME interval satisfies the above condition, ten it is considered as TP otherwise it is considered as False Positive (FP). Consider a video have x ground-truth intervals and y spotted intervals, then the TP is counted as z $(z \leq x \,\&\, z \leq y)$, hence $FP = z - x$and $FN = z - y$. Then the spotting performance for once video can be measured as

$$DR = \frac{z}{x} \text{and} PPV = \frac{z}{y} (21)$$

And

$$F - Score = \frac{2z}{x+y} (22)$$

The above expressions for performance metrics are to for only one video. For an entire dataset, the overall TP is obtained by the accumulation of all TPs. Consider the test dataset have L videos with X micro-expressions and Y ME intervals, the performance is measured as

$$DR = \frac{Z}{X} \text{and} PPV = \frac{Z}{Y} (23)$$

Where

$Z = \sum_{i=1}^{L} z_i,\ X = \sum_{i=1}^{L} x_i$ and $Y = \sum_{i=1}^{L} y_i$(24)

And

$F - Score = \frac{2*DR*PPV}{DR+PPV}$(25)

Table.1 Performance analysis for MEVIEW dataset

Method	Recall (%)	Precision (%)	F-Score (%)
LBP	45.2314	19.8520	27.6285
RLBP	40.9965	16.4785	23.7017
CBP	53.4412	22.1012	31.2682

The results shown in the Table.1 are obtained based on metrics specified in Eq.(23-24) after the simulation of entire videos available in MEVIEW dataset. In the MEVIEW dataset, the total number of available videos is 40. Irrespective of the notes specified in the dataset information, we consider the entire videos for simulation. In the notes at some videos, they specified that there is no micro expressions butt have macro expression or eye blinking etc. For such kind of videos also they specified the ground-truth spotting intervals and hence we consider such kind of videos also for simulation. Since out main objective is to spot the peak expressions, we simulated all the videos and validated with ground-truth interval in every videos. Here we employed both LBP and CBP feature extractors for feature extraction and simulated individually. Absolutely the proposed CBP is observed to have more recall, precision as well as F-Score compared to the conventional LBP. Due to the very limited accomplishment of MEVIEW dataset for ME spotting validation, we didn't provide a comparison with the existing method in literature survey. Compared to CBP and LBP, the RLBP have very less performance due to the consideration of only neighbor pixels information excluding t center pixel.

5. Conclusion

Micro-Expression spotting is much important for Micro Expression recognition. Without an appropriate spotting of ME in lengthy videos, the recognition accuracy is much affected. Moreover, the ME is very spontaneous and for such kind of objects, processing an entire video creates a huge computational burden. Hence, in this paper, we proposed a simple and effective method for ME spotting based on Local Binary Patterns. A new feature descriptor called as CBP is proposed here by combining the conventional LBP with the improved version of LBP called as RLBP. Before feature extraction, as a preprocessing method, we proposed a new landmarks detection followed by facial alignment. Finally after the CBP, we apply feature difference analysis for ME spotting. The recall rate obtained after the simulation of proposed method on MEVIEW dataset proves that performance effectiveness.

REFERENCES

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

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

[3] Q.Wu, X.-B. Shen, and X.-L. Fu, ``Micro-expression and its applications,"*Adv. Psychol. Sci.*, vol. 18, no. 9, pp. 1359_1368, 2010.

[4] Ekman, P. (2002).Microexpression Training Tool (METT). University of California,San Francisco, CA.

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

[6] 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.
[7] Yan, W.-J., Wu, Q., Liang, J., Chen, Y.-H., and Fu, X. (2013a). How fast are the leaked facial expressions: the duration of micro-expressions. J. Nonverb. Behav. 37, 217–230.
[8] 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.
[9] 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.
[10] 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.
[11] Wu, Q., Shen, X., and Fu, X. (2011). "The machine knows what you are hiding: anautomatic micro-expression recognition system," in Affective Computing andIntelligent Interaction ACII 2011 (Memphis), 152–162.
[12] Shreve, M., Godavarthy, S., Goldgof, D., and Sarkar, S. (2011). "Macro-and micro-expressionspotting in long videos using Spatio-temporal strain," in 2011IEEE International Conference on Automatic Face & Gesture Recognition andWorkshops (FG 2011) (Santa Barbara), 51–56.
[13] Shreve, M., Godavarthy, S., Manohar, V., Goldgof, D., and Sarkar, S. (2009)."Towards macro-and micro-expression spotting in video using strain patterns,"in 2009 Workshop on Applications of Computer Vision (WACV) (Snowbird),1–6.
[14] Shreve, M., Brizzi, J., Fefilatyev, S., Luguev, T., Goldgof, D., and Sarkar, S. (2014).Automatic expression spotting in videos. Image Vis. Comput. 32, 476–486.
[15] Moilanen, A., Zhao, G., and Pietikainen, M. (2014). "Spotting rapid facialmovements from videos using appearance-based feature difference analysis,"in 2014 22nd International Conference on Pattern Recognition (ICPR)(Stockholm), 1722–1727.
[16] Li, X., Xiaopeng, H., Moilanen, A., Huang, X., Pfister, T., Zhao, G., et al. (2017).Towards reading hidden emotions: a comparative study of spontaneous microexpressionspotting and recognition methods. IEEE Trans. Affect. Comput.
[17] Wang, S.-J., Wu, S., Qian, X., Li, J., and Fu, X. (2016). A main directionalmaximal difference analysis for spotting facial movements from long-termvideos.Neurocomputing 230, 382–389.
[18] Davison, A. K., Yap, M. H., and Lansley, C. (2015). "Micro-facial movement detection using individualized baselines and histogram-based descriptors," in 2015 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (Kowloon), 1864–1869.
[19] Davison, A., Lansley, C., Costen, N., Tan, K., and Yap, M. H. (2016). SAMM: a spontaneous micro-facial movement dataset. IEEE Trans. Affect. Comput. 9, 116–129. doi: 10.1109/TAFFC.2016.2573832
[20] Davison, A. K., Lansley, C., Ng, C. C., Tan, K., and Yap, M. H. (2016). Objective micro-facial movement detection using facs-based regions and baseline evaluation.arXiv preprint arXiv:1612.05038.
[21] Xia, Z., Feng, X., Peng, J., Peng, X., and Zhao, G. (2016). Spontaneous micro-expression spotting via geometric deformation modeling.Comput. Vis. Image Understand. 147, 87–94.
[22] Duque, C., Alata, O., Emonet, R., Legrand, A.-C., and Konik, H. (2018). "Microexpression spotting using the Riesz pyramid," inWACV 2018 (Lake Tahoe).
[23] Zhihao Zhang, Tong Chen, HongyingMeng, Guangyuan Liu, and Xiaolan Fu, "SMEConvNet: A Convolutional Neural Network for Spotting Spontaneous Facial Micro-Expression from Long Videos", IEEE Access, Vol. 6, 2018, pp.71143-71151.
[24] PetrHusa´k, Jan C˘ ech, Jir˘´ı Matas, "Spotting Facial Micro-Expressions "In the Wild", 22nd Computer Vision Winter Workshop Nicole M. Artner, Ines Janusch, Walter G. Kropatsch (eds.) Retz, Austria, February 6–8, 2017.