JOURNAL OF BASIC SCIENCE AND ENGINEERING

A COMPREHENSIVE METHOD FOR MICRO EXPRESSION RECOGNITION: GRADIENT-FEATURE DIFFERENCE BASED SPOTTING AND AMALGAMATED LBP BASED CLASSIFICATION

Sreenivasu Bhukya[1*], Prof.L.Nirmala Devi[2], Dr.A.Nageswar Rao[3]

[1*]Research Scholar, Department of ECE, University College of Engineering, Osmania University, Hyderabad, Telangana State, India
[2]Professor, Department of ECE, University College of Engineering, Osmania University, Hyderabad, Telangana State, India
[3]Senior Manager (Design), Strategic Electronics Research and Design center, HAL, Hyderabad, India

Abstract: Micro Expression (ME) recognition from video sequences is a challenging task that plays a significant role in various fields, including security, psychology, and human-computer interaction. In this paper, we propose a novel method for ME recognition that encompasses two distinct phases: ME spotting and ME classification. The ME spotting phase identifies key frames by analyzing feature differences based on gradient attributes, capturing significant changes that are indicative of micro-expressions. In the ME classification phase, we employ a combination of Local Binary Patterns (LBP) and Local Average Binary Patterns (LABP) to extract expression-related features from these key frames. These features are then classified using a Support Vector Machine (SVM), which is well-suited for handling complex decision boundaries and achieving high classification accuracy. Our experimental results demonstrate that this approach outperforms existing methods, achieving an accuracy of 72.36% on the CASME II dataset, surpassing previous techniques in both precision and reliability. The proposed method provides a robust and effective framework for ME recognition, offering significant advancements over prior methods and laying the groundwork for future research in this domain.

Keywords: Micro Expression (ME), ME Spotting, ME Classification, Local Binary Patterns (LBP), Local Average Binary Patterns (LABP), Support Vector Machine (SVM) and CASME II Dataset

1. INTRODUCTION

Recently, advancements in technologies like computer vision, machine learning, and artificial intelligence have made intelligent Human-Computer Interaction (HCI) a crucial aspect of modern life. As society moves towards a future driven by intelligence, Intelligent HCI will become an essential component of our daily activities. Such HCI systems will not only focus on completing tasks but also consider the emotional states of users to enhance interactions. To analyze these emotional states, HCI systems utilize various input sources, including text, speech, and facial expressions. According to psychologists, facial expressions are the most effective at conveying emotions, accounting for approximately 55% of emotional communication, while speech and text contribute 38% and 7%, respectively [1].

Although facial expressions can reveal a person's mental state, there are situations where individuals deliberately express or mask certain emotions, making it challenging to accurately analyze their true feelings. Such expressions are known as Facial Micro Expressions (MEs), which are brief and often imperceptible [2]. MEs are involuntary and can uncover genuine emotions that a person may be trying to conceal or suppress. Due to their fleeting nature, MEs are difficult to manipulate and provide a more accurate reflection of a person’s emotional state [3].

The concept of MEs was first identified by Haggard and Isaacs in 1966 during a psychotherapy study, where they observed that these short-lived facial expressions are hard to detect [4]. Later, Ekman et al. [5] highlighted MEs in a video of a conversation between a psychologist and a patient who was depressed but attempted to mask this emotion with a pained smile. Researchers recognize that MEs represent strong, spontaneous, and unconscious emotional responses, providing a more authentic glimpse into a person's true feelings compared to macro or typical facial expressions. MEs are especially valuable in high-stakes situations due to their ability to reveal underlying emotions with high accuracy [6-9].

MEs are brief, involuntary facial expressions that occur as a direct manifestation of a person's genuine emotions. These expressions typically last only a fraction of a second, usually between 1/25th to 1/5th of a second [10-11], and are often difficult to detect with the naked eye. Unlike regular facial expressions, which can be consciously controlled and manipulated, MEs are automatic responses that reveal true emotions, even when an individual attempts to hide or suppress them [12]. The Micro-Expression Training Tool (METT) is a specialized program designed to help individuals recognize and interpret micro expressions accurately. Developed by Dr. Paul Ekman [13], a renowned psychologist and expert in the field of emotions and facial expressions, METT provides training that improves the ability to detect these subtle emotional cues. However, Frank et al. [14] found that the detection performance of METT is quite limited, with an accuracy of approximately 40%. This underscores the need for developing a more effective micro-expression recognition model, which is the primary motivation for our work.

An automatic ME recognition model involves two stages: ME spotting and ME recognition. In the ME spotting stage, key frames (frames that contain emotion attributes) are extracted from the input video. In the ME recognition stage, the specific type of emotion is identified based on features trained to the classification system. Since MEs are present in only a few frames, processing an entire video with a large number of frames would introduce significant computational complexity to the recognition system. Therefore, ME spotting is essential to accurately and precisely identify the frames containing emotion attributes, thereby reducing computational load and enhancing the efficiency of the recognition system. On the other hand, the classification of MEs is also a big challenging ask because emotional attributes looks similar in nature due to the low intensities.

This paper introduce a comprehensive two-phase approach for ME recognition that integrates ME spotting and ME classification. This structured methodology first extracts significant key frames using gradient-based techniques and then classifies these frames using advanced feature extraction and classification methods. This framework offers a systematic and effective solution for detecting and analyzing micro-expressions. The major contributions of this work are outlined as follows;

1. The proposed method employs a gradient-based Feature Difference Analysis to extract key frames that capture significant changes indicative of MEs. This technique improves the accuracy of ME detection by focusing on subtle variations in facial expressions, which are crucial for recognizing micro-expressions
2. We propose a combination of Local Binary Patterns (LBP) and Local Average Binary Patterns (LABP) for feature extraction. LBP captures texture details, while LABP provides complementary information about local average patterns. This combined approach enhances

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

the representation of expression-related features from key frames, leading to more effective micro-expression classification.

The paper is organized as follows: Section 2 provides a comprehensive literature survey, reviewing current methods for micro-expression recognition, highlighting their strengths and limitations, and establishing the need for advanced techniques. Section 3 details the proposed method, describing a two-phase approach for micro-expression recognition that includes ME spotting using gradient-based techniques and ME classification using a combination of Local Binary Patterns (LBP) and Local Average Binary Patterns (LABP) with a Support Vector Machine (SVM). Section 4 presents the results and analysis, outlining the experimental setup, discussing the performance of our method on the CASME II dataset, and comparing it with existing techniques to demonstrate its effectiveness. Finally, Section 5 offers the conclusion, summarizing the contributions of our work, its impact on the field, and suggesting directions for future research.

2. LITERATURE SURVEY

Several methods are proposed in the past towards the accurate recognition of Micro Expressions from videos. A. Moilanen et al. [15] proposed a simple method for automatically spotting rapid facial movements from videos. The method relies on analyzing differences in appearance-based features of sequential frames. In addition to finding the temporal locations, the system is able to provide spatial information about the movements in the face. Micro-expression spotting experiments are carried out on three datasets consisting only of spontaneous micro-expressions. Baseline micro-expression spotting results are provided for these three datasets including the publicly available CASME database.

Su-Jing Wang et al. [16] proposed the Main Directional Maximal Difference (MDMD) analysis for micro-expression spotting. MDMD uses the magnitude of maximal difference in the main direction of optical flow as a feature to spot facial movements, including micro-expressions. Based on block-structured facial regions, MDMD obtains more accurate features of the movement of expressions for automatically spotting micro-expressions and macro-expressions from videos. This method obtains both the temporal and spatial locations of facial movements. The evaluation was performed on two spontaneous databases (CAS(ME)^2) and CASME) containing micro-expressions and macro-expressions.

H. Ma et al. [17] proposed a novel Region Histogram of Oriented Optical Flow (RHOOF) feature to spot the apex frame automatically. First, a set of facial landmarks are detected and then 5 Regions of Interest (ROIs) are selected from facial region based on the frequency of occurrence of action units. Finally, they extract optical flow fields frame-by-frame and compute HOOF in these ROIs. Experiments are conducted on two ideal spontaneous micro-expression databases, i.e., CASME and CASME II.

Carlos Arango Duque [18] proposed a novel method for micro-expression spotting and recognition using the Riesz Pyramid. This method involves constructing a Riesz Pyramid from input video frames and calculating phase differences to detect subtle facial movements. The process includes face alignment, feature extraction using the Riesz Pyramid, and classification using a shallow CNN.

A. K. Davison et al. [19] focused on detecting micro-movements in videos using Histogram of Oriented Gradients (HOGs). They pre-processed frames by cropping, aligning, and noise removal, then divided each frame into blocks to calculate HOGs. The chi-squared distance between consecutive frames' spatial appearances was normalized to detect peaks, identifying key frames.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

D. Patel et al. [20] captured motion feature continuity by computing Optical Flow Vectors for small spatial regions and integrating them into temporal regions. They used heuristics to eliminate non-MEs, effectively determining ME frames.

Zhang et al. [21] employed the SMEConvnet deep learning model to extract Spatio-temporal features from lengthy videos, using a sliding window for apex frame spotting. Pre-processing included frame alignment and cropping. V. Burni and D. Vitulano [22] introduced "Frozen Frames," which appear just before or after a micro-expression (ME) and signal attempts to hide emotions. These frames were detected using a simplified Adelson and Bergen Energy model for motion perception. The authors used groups of frozen frames to identify ME frames.

Y. Han et al. [23] proposed a method for ME spotting called Feature Difference Analysis (FDA). This method involves partitioning a face image into several uniform Regions of Interest (ROIs) and computing features from these regions. They used Fisher Linear Discriminant Analysis (LDA) for evaluation, which assigns a weight to each ROI. Initially, FDA utilized two independent features: Local Binary Patterns (LBP) and Histogram of Optical Flow (HOOF). Later, the authors introduced MDMO into FDA [24] and proposed a collaborative strategy known as Collaborative Feature Difference (CDA). CDA leverages two complementary features, LBP for texture information and MDMO for motion information, to enhance ME detection. Additionally, V. Esmaeili & S. O. Shahdi [25] introduced a new LBP-based texture descriptor called Cubic LBP. This method calculates LBP on 15 distinct planes and demonstrated that analyzing these 15 planes is effective for identifying the apex frame where maximum facial movements occur.

Y. Wang e al. [26] developed 2 expression descriptors namely “LBPSIP and LBP on Mean Orthogonal Planes (MOP)”. They are efficient in preserving the key patterns and reduce the redundancy. Moreover this method provided sufficient discrimination capability to the recognition system.

X. Huang et al. [27] developed a new mechanism named as “Spatio-Temporal Completed Local Quantization Pattern (STCLQP)” for the analysis of MEs. At first STCLQP extracts three important attributes from face image such as sign, magnitude and orientation. Next, they introduced an efficient vector quantization and code book selection mechanism for every attribute in the temporal domain to analyze the discrimination capability and compactness of the code book structure to generalize the traditional pattern representation methods.

Optical Flow vectors (OFV) is a standard method which can explore the information about motion features and hence several earlier authors employed for the recognition of MEs. Y. J. Liu et al. [28] developed a simple approach and named it as “Main Directional Main Optical Flow (MDMO)”. This is a “Region of Interest (RoI)” oriented normalized statistical attribute which considers constant motion information and spatial location. The interesting fact about this method is its smallest size. At last they used the SVM algorithm to recognize the MEs.

S. L. Happy and A. Routray [29] distinguished the effectiveness of temporal features and associated them for MEs. Through these methods, they developed a “Fuzzy Histogram of Optical Flow Orientation (FHOFO)” method. FHOFO created through suitable histograms with the help of optical flow vector orientations based on the Fuzzification of histograms that encodes the temporal pattern. They also explored and investigated the inclusion and exclusion of magnitudes of the motion at the instance of feature extraction.

Further, Lu et al. [31] developed “Fusion of Motion Boundary Histograms (FMBH)” method which combines the horizontal and vertical displacements of Discrepancy OFVs. S. T. Liong et al. [30] contemplate only 2 frames from every ME video and proposed a new proposition

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

method for MER. The two frames are namely an apex frame and onset frame. They are also proposed a simple and effective expression descriptor and called it as "Bi-Weighted Oriented Optical Flow (Bi-WOOF)" that encodes the important features in the apex frame.

3. PROPOSED APPROACH

The proposed method for recognizing Micro Expressions (MEs) from videos consists of two phases: ME spotting and ME classification. ME spotting extracts key frames by analyzing feature differences based on gradient attributes, identifying significant changes indicative of MEs. ME classification then uses a combination of Local Binary Patterns (LBP) and Local Average Binary Patterns (LABP) to represent expression-related features from these key frames. Finally, a Support Vector Machine (SVM) algorithm classifies the extracted features, leveraging its effectiveness in handling complex decision boundaries to accurately distinguish between different micro-expressions. This approach ensures a robust and reliable system for ME recognition from video sequences. Figre.1 shows the overall block diagram of proposed approach.

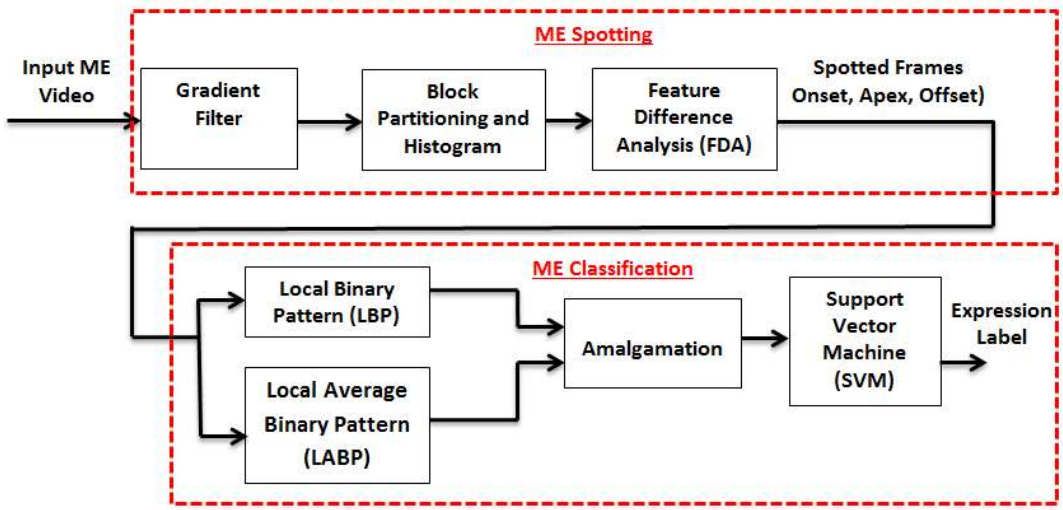


Figure.1 Block diagram of proposed approach

3.1 ME spotting

ME spotting is a crucial preliminary step in the process of ME classification. It serves to identify and extract the key frames from a video where the micro-expressions occur. The proposed ME spotting technique is novel in that it considers the gradients of the image rather than the traditional pixel intensities. This approach leverages the fact that gradients capture changes in intensity more effectively, highlighting the subtle and brief movement's characteristic of micro-expressions better than raw pixel values. The novelty of using gradients lies in their ability to emphasize edges and transitions within the facial features, which are critical for detecting the minor muscle movements involved in micro-expressions. After representing the frames based on gradient attributes, Feature Difference Analysis (FDA) is used to extract the key frames. These key frames consist of three crucial frames: onset (the beginning of the expression), apex (the peak of the expression), and offset (the end of the expression). This gradient-based ME spotting technique not only enhance the detection accuracy but also ensures that the most relevant frames are selected for further analysis. By focusing on the gradient changes, the method effectively

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

isolates the moments of significant facial movement, laying a robust foundation for subsequent ME classification.

3.1.1 Gradient Attributes

Gradient features are the most prominent features which gained huge popularity in several image processing applications. The gradient of an image gives information about the directional change of the color or intensity in an image. For instance, the canny edge operator uses image gradient for the detection of edges in images. The gradients help in providing sufficient information about the presence of different regions in images. In the case of images with facial expressions, the expression portion is totally different from other parts of images. In such cases, the accomplishment of gradients over facial images differentiates the regions effectively. Hence, this paper considers extracting the gradient features from facial image to do the segmentation. The gradient is simply defined as Directional derivative of scalar field. Gradient gives information about the image. The gradients are calculated in two directions, they are vertical and horizontal directions. Based on these two gradients, the final gradient magnitude and the corresponding direction are measured. Considering this fact, here we represent each pixel with four gradient features they are G_x, G_y, G_M and G_θ . They are calculated as follows

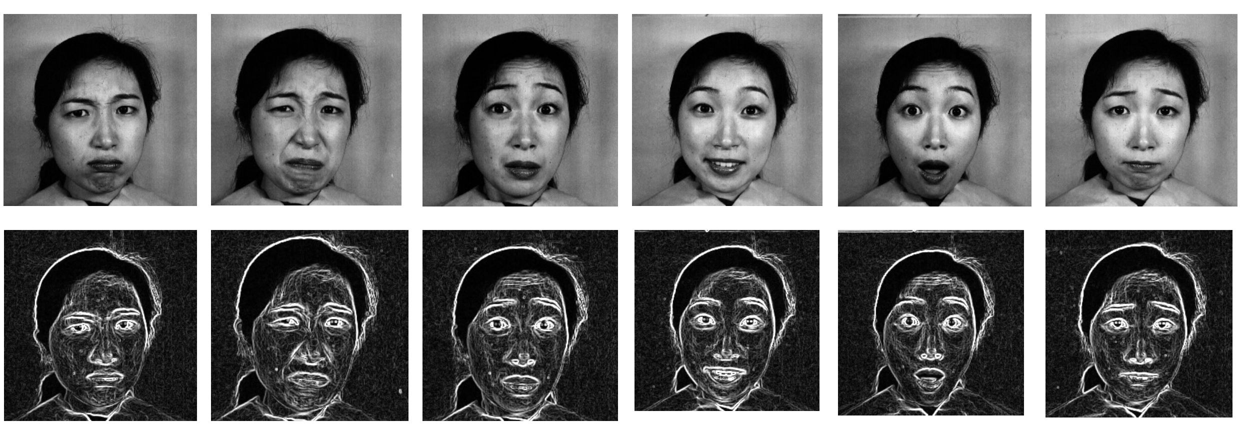
$$G_x(x,y) = |C(x+1,y) - C(x-1,y)| \quad (1)$$

$$G_y(x,y) = |C(x,y+1) - C(x,y-1)| \quad (2)$$

$$G_M(x,y) = \sqrt{\left(G_x(x,y)\right)^2 + \left(G_y(x,y)\right)^2} \quad (3)$$

$$G_\theta(x,y) = \tan^{-1}\left(\frac{G_y(x,y)}{G_x(x,y)}\right) \quad (4)$$

Where $G_x(x,y)$ is the gradient of pixel (x,y) along X-direction and $G_y(x,y)$ is the gradient of pixel (x,y) along Y-direction, $G_M(x,y)$ is the magnitude of a gradient and $G_\theta(x,y)$ is the direction of a gradient $G_M(x,y)$. We explore the difference between successive pixels and $G_\theta(x,y)$ and express the direction of movement of the corresponding pixel. Figure.2 shows some examples of original facial expression images and their corresponding gradient images.



(a) (b) (c) (d) (e) (f)

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

Figure.2 top row - facial expression images and bottom row - gradient representation (a) Angry (b) Disgust, (c) Fear, (d) Happy, (e) Surprise and (f) Sad

3.1.2 FDA

Once each frame of the ME video is represented with gradients, they are processed using Feature Difference Analysis (FDA) to identify the key frames. FDA determines three key frames: onset, apex, and offset of the micro-expression. To do this, FDA treats each frame as the Current Frame (CF) and calculates the motion deviation compared to its preceding and succeeding frames. For this analysis, two frames—one is preceding and one is succeeding—is averaged to create a Mean Frame (MF). FDA computes the Chi-Squared Distance (CSD) between the CF and the MF to quantify motion variations in the facial area. This distance helps identify rapid facial movements within lengthy videos. Except for the first and last frames, FDA calculates the CSD for all frames. The CSD is computed over the normalized histograms of CF and MF, providing a measure of similarity or dissimilarity between them. For FDA computation, both CF and MF are divided into several blocks, and histograms are computed for each block, ensuring detailed analysis of localized facial movements. 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

$$\mathcal{X}^2\left(C_j^i, M_j^i\right) = \frac{\left(C_j^i - M_j^i\right)^2}{C_j^i + M_j^i} \qquad (5)$$

Where $\mathcal{X}^2\left(C_j^i, M_j^i\right)$ 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

$$V_i = \frac{1}{M}\sum_{j=1}^{M} \mathcal{X}^2\left(C_j^i, M_j^i\right) \qquad (6)$$

Here V_i explores the difference between i^{th} block 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

$$L_i = V_i - \frac{1}{2}(V_{i+k} - V_{i-k}) \qquad (7)$$

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

$$T = L_{mean} \mp (L_{max} - L_{mean}) \qquad (8)$$

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.

3.2 ME Classification

ME Classification is the second phase of the proposed approach, utilizing the key frames identified during the ME spotting phase. These key frames serve as inputs for the classification

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

process. ME classification involves applying Local Binary Patterns (LBP) and Local Average Binary Patterns (LABP) to extract texture-related features from the images. These features capture the essential details of facial expressions, enabling a robust representation of micro-expressions. Once the features are extracted, they are used to train the system through SVM algorithm. SVM, known for its high accuracy and efficiency in classification tasks, learns to distinguish between different micro-expressions based on the training data. After training, the system undergoes testing to evaluate its performance in recognizing and classifying micro-expressions. This phase ensures that the system can accurately identify subtle and brief facial expressions from video sequences, completing the process of micro-expression recognition.

3.2.1 Feature Extraction

After extracting key frames from the input ME video, these frames undergo feature extraction. We employ a Local Binary Patterns (LBP) based feature descriptor to extract appearance-based features from each key frame. In this process, we introduce a new LBP variant called Amalgamated LBP (A-LBP), which combines LBP [32] and Local Average Binary Patterns (LABP) [33]. LBP, initially introduced in the 1990s, has been widely used in various computer vision applications such as human action recognition, facial expression recognition, texture analysis, and object detection. The standard procedure for computing Local Binary Pattern (LBP) involves the following steps: For a center pixel surrounded by 8 neighboring pixels within a block radius *r*, each neighbor is compared to the center pixel based on their pixel intensities. If a neighbor's intensity is greater than the center pixel's intensity, it is encoded as 1; otherwise, it is encoded as 0. This binary encoding is applied to all neighboring pixels in an anti-clockwise direction, forming an eight-bit string. This string is then converted to a decimal value, which is the LBP value of the center pixel. For a Centre pixel q_c surrounded by *p* neighbour pixels on circle of radius *r*, the LBP is completed as

$$LBP_{r,p}(q_c) = \sum_{n=0}^{P-1} s\left(q_{r,p,n} - q_c\right)2^n \tag{9}$$

Where

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{10}$$

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

$$LMBP_{r,p}(q_c) = \sum_{n=0}^{P-1} s\left(q_{r,p,n} - Q_p\right)2^n \tag{11}$$

Where

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

Finally each pixel is represented with two decimal codes; one is through LBP code and another is through LABP 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.3 and figure.4 respectively.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

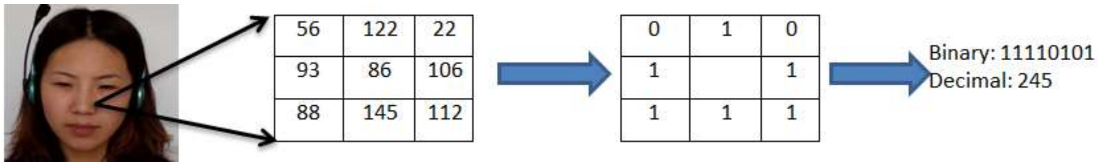


Figure.3 LBP Process

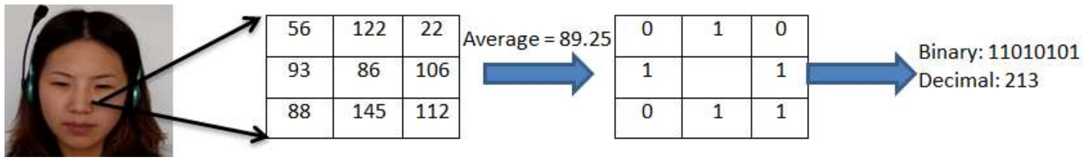


Figure.4 LABP Process

3.2.2 Classification

Support Vector Machine (SVM) algorithm plays a pivotal role in our work for the classification of Micro Expressions. SVM is a supervised machine learning algorithm that excels in binary and multi-class classification tasks by finding the optimal hyperplane that maximizes the margin between different classes in the feature space.

Significance of SVM in ME Classification:

1. **Enhanced Classification Accuracy:** SVM enhances the classification accuracy of MEs by effectively handling the high-dimensional feature space generated from the LBP and LABP descriptors. The algorithm's ability to maximize the margin between classes ensures that the decision boundary is as far as possible from any data point, reducing the risk of misclassification.
2. **Robust to Overfitting:** One of the key benefits of using SVM is its robustness to over-fitting, especially in high-dimensional spaces common in image processing tasks like ME classification. This is achieved through regularization parameters that control the trade-off between maximizing the margin and minimizing classification errors.
3. **Effective with Small Sample Sizes:** SVM is particularly effective when dealing with small sample sizes, which is often the case in ME datasets. Its reliance on support vectors (a subset of the training data) to define the decision boundary makes it less sensitive to the number of training samples, maintaining high accuracy even with limited data.
4. **Versatility with Kernels:** The use of kernel functions (linear, polynomial, radial basis function, etc.) allows SVM to model complex, non-linear relationships between features, making it adaptable to various patterns and distributions in the data. This versatility ensures that subtle and intricate variations in micro-expressions are captured and accurately classified.
5. **Scalability:** SVM can efficiently handle large feature sets, which is beneficial when working with the rich and detailed features extracted from facial images using LBP and LABP. Its scalability ensures that the classification process remains computationally feasible even as the feature dimensionality increases.

4. EXPERIMENTAL ANALYSIS

In this section, we validate the effectiveness of our proposed method using the CASME (Chinese Academy of Sciences Micro-Expression) dataset, a widely recognized standard dataset

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

for micro-expression recognition. The CASME dataset provides a comprehensive benchmark for evaluating the performance of various micro-expression recognition techniques.

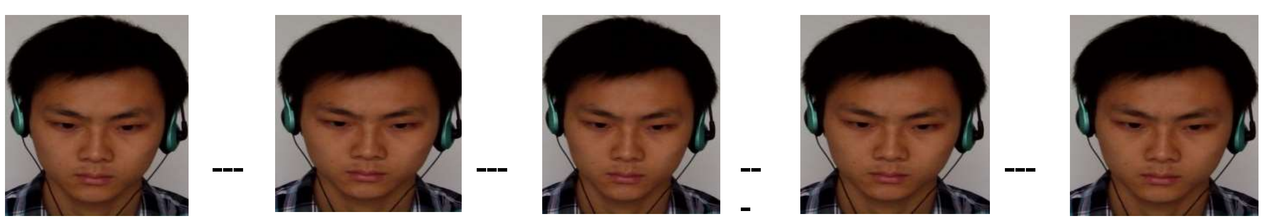
To assess the performance of our method, we employ several metrics: accuracy, recall, precision, and F-score. Accuracy measures the overall correctness of the method by comparing the number of correctly identified micro-expressions to the total number of micro-expressions. Recall, also known as sensitivity, indicates the ability of the method to identify true positive instances among all actual positive instances. Precision, on the other hand, measures the proportion of true positive instances among all instances identified as positive by the method. The F-score, the harmonic mean of recall and precision, provides a single metric that balances both precision and recall. Our results are then compared with state-of-the-art methods to highlight the advantages and potential improvements offered by our approach. By doing so, we demonstrate not only the effectiveness of our method in recognizing micro-expressions but also its competitiveness and superiority in the field.

4.1 Dataset and Simulation set up

CASME [34] dataset contains spontaneous micro-expressions collected in a controlled laboratory environment. These micro-expressions are elicited by showing participants emotion-evoking videos, ensuring a natural and authentic set of facial expressions. The dataset includes high-quality video recordings with detailed annotations, including the onset, apex, and offset of each micro-expression, along with the corresponding emotion labels. This makes CASME an ideal choice for evaluating the performance of micro-expression recognition methods.

CASME II is one of the most widely used databases in micro-expression research, consisting of a total of 247 micro-expression video clips acquired from 26 subjects. Each video clip is recorded at a frame rate of 200 frames per second. The database is categorized into five classes: Repression (27 samples), Disgust (64 samples), Surprise (25 samples), Happiness (32 samples), and other (99 samples). The original frame resolution is 640×480 pixels, which is reduced to 340×280 pixels after cropping. Additionally, CASME II includes Action Unit labels following the Facial Action Coding System (FACS). Some of the Video clips of CASEMII are shown in Figure.5.

Start (1) Onset (83) Apex (89) Offset (96) End (178)



Start (1) Onset (181) Apex (191) Offset (209) End (277)

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

Figure.5 Samples of CASME II video clips with start, onset, peak, offset and ending frames.

4.2 Results

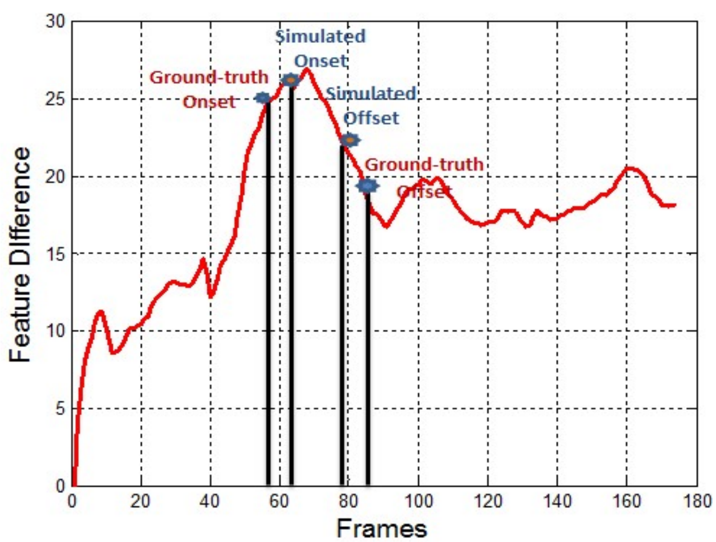
The evaluation of our proposed method is conducted in two distinct phases: spotting results and classification results. Each phase plays a crucial role in validating the effectiveness and accuracy of our approach.

Spotting Results: In the first phase, we focus on spotting results, which involve the identification of key frames that represent the onset, apex, and offset of micro-expressions. The performance of our proposed spotting mechanism is validated by comparing the ground truth key frames with the key frames obtained through our method. This comparison allows us to assess the recall of our spotting technique, ensuring that our method effectively captures the critical moments of micro-expressions. Under the performance evaluation, we consider F1-score to assess the effectiveness of proposed spotting mechanism. It is measured based on True Positives which are measured as follows;

$$TP = \frac{(I_{Spotted}) \cap (I_{Groudtruth})}{(I_{Spotted}) \cup (I_{Groudtruth})} \geq k \quad (12)$$

Where $I_{Spotted}$ is the posted interval and $I_{Groudtruth}$is ground truth interval. These two intervals defines the frame sin the period of $onset - offset$.

$$F1 - score = \frac{2TP}{2TP+FP+FN} \quad (13)$$



Vol. 21, No. 1, (2024)
ISSN: 1005-0930

Figure.6 Feature difference analysis between ground truth key frames and obtained key frames through proposed spotting mechanism

Figure.6 depicts the feature difference over a series of frames in a video clip, focusing on the key moments of a micro-expression. It explores the Feature difference analysis between ground truth key frames and obtained key frames through proposed spotting mechanism. The x-axis represents the frame number (0 to 180), while the y-axis represents the feature difference. Key frames are marked to show the ground-truth onset and offset, as well as the simulated onset and offset identified by the proposed spotting mechanism. The vertical lines highlight the duration of the micro-expression as detected by the proposed method. The graph shows an increase in feature difference peaking around the onset of the micro-expression and then decreasing, illustrating the intensity pattern of a micro-expression over time. This figure visually assesses the alignment between the proposed method and ground-truth annotations in detecting the onset and offset of micro-expressions.

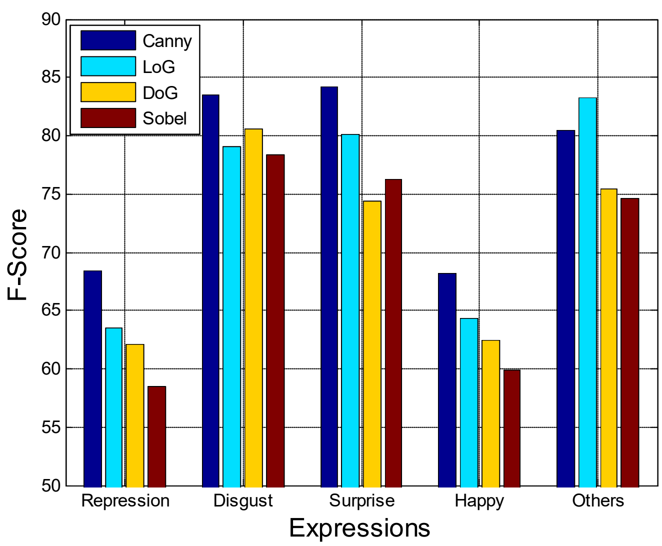


Figure.7 F-score Analysis with different gradient operators for different expressions

Figure.7 illustrates the performance of different gradient operators—Canny, LoG (Laplacian of Gaussian), DoG (Difference of Gaussian), and Sobel—evaluated based on the F-score metric across several facial expressions: repression, happy, surprise, disgust, and others. The study considers various facial expressions to see how well each gradient operator performs across different emotional states. The results highlight which gradient operators are more effective for specific expressions. For example, Canny might perform best for detecting "happy" expressions, while Sobel might be more effective for disgust. In evaluating the performance of gradient operators for facial expression recognition, the canny operator achieved the highest average F-score of 76.96, demonstrating the best effectiveness in edge detection compared to the others. The LoG operator followed with a solid average F-score of 74.07, showing good performance but slightly less effective than Canny. The DoG operator and Sobel operator had lower average F-scores of 71.00 and 69.51, respectively, indicating moderate and the least effectiveness in detecting facial expressions. Overall, Canny stands out as the most effective gradient operator for analyzing facial expressions, while Sobel shows the least effectiveness among the tested methods.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

Classification Results: In the second phase, we evaluate the classification results, which involve recognizing and categorizing the identified micro-expressions into their respective classes. This phase is based on comparing the ground truth expressions with the expressions obtained through our classification algorithm. Performance metrics such as accuracy, recall, precision, and F-score are utilized to measure the effectiveness of our method in correctly identifying and classifying the micro-expressions.

Table.1 Results of LBP over CASME II dataset

Emotion/Metric	Recall (%)	Precision (%)	F1-Score (%)	FNR (%)
Happy	71.4541	57.0541	63.5106	28.5459
Disgust	71.7696	85.4311	78.0453	28.2304
Surprise	80.8959	76.3492	78.5611	19.1041
Repression	48.2852	80.2453	60.6033	51.7148
Others	80.1958	75.9264	78.0065	19.8042

Table.2 Results of LABP over CASME II dataset

Emotion/Metric	Recall (%)	Precision (%)	F1-Score (%)	FNR (%)
Happy	72.3352	59.6312	65.2341	27.6648
Disgust	70.2145	85.2231	75.2431	29.7855
Surprise	79.4512	76.1345	73.2214	20.5488
Repression	44.2312	75.4215	58.6341	55.7688
Others	80.1478	78.3235	78.2020	19.8522

Table.3 Results of A_LBP over CASME II dataset

Emotion/Metric	Recall (%)	Precision (%)	F1-Score (%)	FNR (%)
Happy	73.3691	58.9691	65.4256	26.6309
Disgust	73.6846	87.3461	79.9603	26.3154
Surprise	82.8109	78.2642	80.4761	17.1891
Repression	50.2002	82.1603	62.5183	49.7998
Others	82.1108	77.8414	79.9215	17.8892

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

The results from the three methods—LBP, LABP, and A_LBP—show distinct variations in their performance metrics for facial emotion recognition on the CASME II dataset. A_LBP demonstrates the highest F1-Score for "Surprise" (80.48%) and "Others" (79.92%), indicating superior overall performance in capturing these emotions. In contrast, LBP shows better precision for "Disgust" (85.43%) compared to the other methods, while LABP has the lowest performance in terms of the F1-Score for "Surprise" (73.22%). A_LBP consistently provides higher F1-Scores for "Happy" (65.43%) and "Disgust" (79.96%), suggesting improved balance between recall and precision across most emotions compared to LBP and LABP. LABP also has higher recall for "Happy" (72.34%) but lower performance in detecting "Repression" (58.63%). Overall, A_LBP generally outperforms the other methods in F1-Score and recall, particularly for "Surprise" and "Others," while LBP and LABP show variable strengths and weaknesses across different emotions. Across the CASME II dataset, A_LBP generally outperforms both LBP and LABP in F1-Score and recall for most emotions, achieving the highest scores for "Surprise" and "Others." LBP excels in precision for "Disgust," while LABP shows the lowest F1-Score for "Surprise" and weaker overall performance in detecting "Repression." Overall, A_LBP provides the best balance of performance metrics across different facial expressions.

Table.4 Comparison with existing methods

Author(s)	**Accuracy (%)**	**Dataset used**	**Classifier**	**Feature Extraction**
Y. J. Liu et al. [28]	67.37	CASME II	SVM	MDMO
Yandan Wang et al. [26]	44.13	CASME II	SVM with RBF kernel	LBL-MOP
S. T. Liong et al. [30]	62.20	CASME II	SVM	Bi-WOOF
Lu et al. [31]	69.11	CASME II	SVM	FMBH
S. L. Happy and A. Routray [29]	56.64	CASME II	LDA, SVM and KNN	FHOFO
X. Huang et al. [27]	58.39	CASME II	Linear kernel assisted SVM	STLQP
Proposed	72.36	CASME II	FDA and SVM	Gradients and A_LBP

Table 4 compares the proposed method's performance with several existing methods for emotion recognition on the CASME II dataset. The proposed method achieves an accuracy of 72.36% using FDA and SVM with Gradients and A_LBP (Amalgamated Local Binary Patterns) for feature extraction. For instance, it outperforms Y. J. Liu et al. [28] (67.37%) and Lu et al. [31] (69.11%), demonstrating its superior performance in recognizing Micro expressions on the CASME II dataset. The use of Gradients and A_LBP for feature extraction contributes to this high performance. A_LBP captures nuanced facial features and variations more effectively than methods like MDMO, LBL-MOP, or Bi-WOOF, which may not fully leverage advanced local pattern techniques. Combining FDA with SVM enhances the classification process by improving class seperability and maximizing the margin between different emotional states. This theoretical approach provides a stronger foundation compared to simpler classifiers or less effective feature extraction methods used in previous studies. Coupled with the advanced A_LBP feature extraction technique, which captures detailed texture and edge information, this approach results in more accurate emotion classification compared to earlier methods that relied on less advanced feature extraction or classification techniques.

5. COCLUSION

In this paper, we presented a novel approach for recognizing Micro Expressions (MEs) from video sequences, characterized by a two-phase methodology that includes ME spotting and ME classification. Our approach effectively identifies key frames through the analysis of gradient attributes to detect significant changes indicative of MEs. Subsequently, we utilized a combination of LBP and LABP for feature extraction, capturing detailed expression-related information from these key frames. The features are then classified using a SVM, chosen for its robustness in handling complex decision boundaries and delivering accurate classification of micro-expressions. Experimental results demonstrate that our method outperforms existing techniques in terms of accuracy, achieving a notable improvement over previous methods. This performance is attributed to the synergistic use of advanced feature extraction techniques and a sophisticated classification approach. Overall, our proposed system offers a reliable and effective solution for micro-expression recognition, with the potential for applications in fields such as security, psychology, and human-computer interaction.

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.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

[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] 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.
[13] Ekman, P. (2002). Micro-expression Training Tool (METT). University of California, San Francisco, CA.
[14] 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.
[15] A. Moilanen, G. Zhao and M. Pietikäinen, "Spotting Rapid Facial Movements from Videos Using Appearance-Based Feature Difference Analysis," 2014 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 2014, pp. 1722-1727,
[16] Su-Jing Wang, Shuhang Wu, Xingsheng Qian, Jingxiu Li, Xiaolan Fu, A main directional maximal difference analysis for spotting facial movements from long-term videos, Neurocomputing, Volume 230, 2017, Pages 382-389.
[17] H. Ma, G. An, S. Wu and F. Yang, "A Region Histogram of Oriented Optical Flow (RHOOF) feature for apex frame spotting in micro-expression," *2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, Xiamen, China, 2017, pp. 281-286.
[18] Carlos Arango Duque, “Analysis of Micro-Expressions based on the Riesz Pyramid: Application to Spotting and Recognition," Signal and Image Processing. ED SIS 488, 2018. English. ffNNT : ff. fftel-02305659.
[19] 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,
[20] 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.
[21] Zhihao Zhang, Tong Chen, Hongying Meng, 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.
[22] 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.
[23] 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.
[24] 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.
[25] V. Esmaeili & S. O. Shahdi, “Automatic micro-expression apex spotting using Cubic-LBP”, *Multimedia Tools and Applications* (2020) 79:20221–20239.
[26] Yandan Wang, John See, Raphael C. W. Phan, Yee-Hui Oh, “Efficient Spatio-Temporal Local Binary Patterns for Spontaneous Facial Micro-Expression Recognition”, *PLoS ONE* 10(5): e0124674.

[27] X. Huang, G. Zhao, X. Hong, W. Zheng, M. Pietikäinen, Spontaneous facial micro-expression analysis using spatiotemporal completed local quantized patterns, Neuro-computing 175 (2016) 564–578.
[28] Y.-J. Liu, J.-K. Zhang, W.-J. Yan, S.-J. Wang, G. Zhao, X. Fu, "A main directional mean optical flow feature for spontaneous micro-expression recognition", *IEEE Trans. Affect. Comput.* 7 (4) (2016) 299–310.
[29] S. Happy, A. Routray, "Fuzzy histogram of optical flow orientations for micro-expression recognition", *IEEE Trans. Affect. Comput*. (2017).
[30] S.-T. Liong, J. See, K. Wong, R.C.-W. Phan, "Less is more: Micro-expression recognition from video using apex frame", *Signal Process., Image Commun*. 62 (2018) 82–92.
[31] H. Lu, K. Kpalma, J. Ronsin, "Motion descriptors for micro-expression recognition", *Signal Process., Image Commun*. 67 (2018) 108–117.
[32] C., Kertész: Texture-Based Foreground Detection, International Journal of Signal Processing, Image Processing and Pattern Recognition (IJSIP), Vol. 4, No. 4, 2011
[33] 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.
[34] W. J. Yan, X. Li, S. J. Wang, G. Zhao, Y. J. Liu, Y. H. Chen, X. Fu, CASME II: An improved spontaneous micro-expression database and the baseline evaluation, *PLoS One,* 9 (1) (2014) e86041.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930