GMR
The Original

QUANTITATIVE ANALYSIS OF MICRO-EXPRESSION PHENOTYPES VIA CUMULATED OPTICAL FLOW AND LBP-TOP UNDER VARYING ENVIRONMENTAL CONSTRAINTS

P. Surekha[1*], P. Vidya Sagar[2], G. Ramesh[3]

[1]PhD Research Scholar, Department of CSE, Koneru Lakshmaiah Education Foundation Vaddeswaram, A.P, India,
Email: ramesh680@gmail.com
[2]Associate Professor, Department of CSE, KoneruLakshmaiah Education Foundation Vaddeswaram, A.P, India
[3] Associate Professor, Department of CSE, GRIET, Hyderabad, Telangana
*Corresponding Author: P. Surekha.

ABSTRACT
Micro-expression recognition is a challenging problem in affective computing because the facial muscle movement is subtle and transient. This paper proposes a hybrid framework that combines a Cumulated Optical Flow Vector (COFV) to capture very subtle inter-frame motion and Local Binary Patterns (LBP-TOP) to encode spatiotemporal texture representation. COFV is used to for motion estimation under illumination changes, while LBP-TOP preserves appearance dynamics.The COFV method captures the cumulative motion patterns over consecutive frames and enhances discriminability in micro-expression analysis. LBP-TOP encodes dynamic texture variations, complementing motion-based features. The extracted feature sets are fused to result in a hybrid representation and, consequently, superior recognition performance. Experimental results show that our approach outperforms traditional optical flow and LBP-based methods through improved robustness against illumination variations and noise. The method was evaluated on the CASME II dataset and achieved 85.2% accuracy, 84.1% precision, 83.3% recall, and 83.7% F1-score, outperforming only optical flow based as well as the LBP-TOP based baseline models.The proposed method provides a structured and reliable feature extraction pipeline for micro-expression recognition and is very suitable for psychology, security, and human-computer interaction applications.

KEYWORDS: Micro Expression (ME) Recognition, Cumulated Optical Flow Vector (COFV), Local Binary Pattern on Three Orthogonal Planes (LBP-TOP), Illumination Variance.

INTRODUCTION
Humans naturally recognize the motion of objects around them, whereas computer-based systems cannot. Various techniques have been developed throughout the years to give machines this ability to recognize movements. For a given input video, successive frames can be captured, which helps recognize motion as time passes. Each frame contains a large number of pixels. These pixels contain information about the intensity and color in that frame. The difference between successive frames is understood as motion.

For the problem of Micro Expression Recognition, it is crucial to understand various optical flow techniques. Micro Expressions (MEs) are extremely minor movements seen in a human face for a short period (less than 0.5 seconds). Optical flow estimation plays a vital role in determining a video's motion flow [8]. These techniques work on changes at the pixel level for successive frames. The first of these include the Horn-Schunck algorithm, which follows a global or dense approach [12], and the Lucas-Kanade method, which follows a local or sparse approach [13]. Peter O'Donovan [17] studied various techniques and applications of optical flow. Typical applications include the detection of real-time speed detection, traffic analysis, ME detection, trajectory analysis of rockets, motion aware cameras used in sports and many more. To implement this concept of optical flow, some assumptions need to be made, like the constancy of brightness and non-arbitrary motion in the pixels through successive frames.

While optical flow estimation is used for capturing motion in successive video frames, texture-based feature extraction is possible using the Local Binary Pattern (LBP) technique. This method is illumination invariant as it is not affected by lighting changes. This means the relative difference between pixel values is always the same irrespective of the illumination present [16]. LBP works based on calculating surrounding pixel values in binary concerning the central pixel in the respective 3x3 kernel [3]. If the surrounding pixel values are greater than or equal to the central pixel value, the result is 1, and otherwise, 0. This binary pattern is then converted into a decimal number. Hence, information about the edge or texture is found using these LBP values across the image. But, when it is required to deal with videos that can be seen as multiple static images stacked orthogonally, minor actions or movements are most interesting. We have the Local Binary Pattern on Three Orthogonal Planes (LBP-TOP) technique to work with LBP in this scenario. Three planes will be present: XY, XT, and YT (T represents time) [25].

Existing systems for ME recognition often struggle with changing illumination and micro facial movements. Handling noise is also challenging. Recent studies have also explored transformer or hybrid CNN-based models for micro-expression analysis, showing that richer spatiotemporal modelling improves recognition performance [22], [24]. Despite introducing these deep models, lightweight as well as handcrafted fusion remains desired for datasets such as CASME II. To address these challenges, this study proposes a cumulated optical flow vector for extracting successive

frames in a video while handling illumination variance. Additionally, LBP-TOP enhances feature extraction by capturing spatial and temporal variations, further improving ME recognition.
The objective of this research is to develop an illumination-robust recognition framework that fuses motion-sensitive COFV features with LBP-TOP texture descriptors to improve recognition accuracy on spontaneous ME datasets.

LITERATURE SURVEY

Local Binary Patterns (LBP) and Texture Analysis

Local Binary Patterns (LBP) is one of the important techniques in Computer Vision for classification ofsubtle textures while having limitations like illumination variance. In [1] introduced a new method, namely completed LBP (CLBP), stating significant improvement in texture classification accuracy. Ke-Chen et al. [2] thoroughly reviewed various applications and advantages of LBP. The benefits mentioned include the grayscale and rotational invariance while having low computational complexity. Chen et al. [3] proposed a new variation, specified as the Robust LBP (RLBP), to upgrade the effectiveness of LBP in various aspects like noise. In [4] discussed the relevance of LBP-based texture analysis techniques. They also emphasized that the important advantage of LBP is the illumination invariance. Liu et al. [5] provided a scheme of enhanced version of LBP for the purpose of texture classification. As a part, two intensity-based descriptors and two difference-based descriptors are introduced. Ahonen et al. [6] utilized the LBP technique to extract feature distributions and concatenated them into a feature vector, which could be used as a face descriptor. Face recognition is a field where LBP finds the most relevance. Rahim et al. [7] further validated the effectiveness of LBP in facial feature extraction. As an extension of existing LBP methods, Goyani and Patel [27] proposed a Local Mean Binary Pattern (LMBP) to address noise and illumination changes while generating highly discriminative code. While LBP is generally utilized to extract texture-based features from static images, LBP-TOP (LBP on Three Orthogonal Planes) has another dimension other than X and Y dimensions. The third dimension represents time making three planes, namely XY, XT, and YT, placed orthogonally (at a 90-degree angle). While dealing with videos, a continuous array of frames, the LBP-TOP can also give a 3D view of how the pixels change over time axis. Mattivi and Shao [25] utilized LBP-TOP for human action recognition. De Freitas Pereira et al. [26] proposed countermeasures based on LBP-TOP while mentioning the importance of the prevailing issue of face spoofing attacks.

Optical Flow Estimation and Motion Analysis

Motion detection across multiple video frames is done by identifying changes in pixels as time passes. Optical flow estimation techniques are standard for determining small motion in a minute area by analysing pixel to pixel changes. Along with classical differential techniques like the Horn-Schunck, others are being researched recently. Fleet and Weiss [8] provided fundamental knowledge on gradient-based optical flow estimation. Sun et al. [9] researched various optical flow estimation algorithms while revealing valuable hidden insights. Otte and Nagel [10] reviewed advances in optical flow estimation, comparing multiple approaches. Fortun et al. [11] surveyed the 35-year progress of introducing new methodological concepts in optical flow estimation. This provided a detailed history recap of this field.
Horn and Schunck [12] first introduced a global optical flow formulation based on brightness constancy and smoothness, which remains a foundational reference. Later, Bruhn et al. [13] combined this global formulation with the local Lucas–Kanade strategy to obtain more accurate and stable flow fields. Sharmin and Brad [14] optimized the Lucas-Kanade technique for estimation of optical flow, indicating that the Gaussian filter outperformed other image filters. Yedjour [15] also presented an improved Lucas-Kanade-based optical flow technique. Ross et al. [16] worked on quantifying descriptor illumination variance in the scenario of optical flow estimation.
O'Donovan [17] explored various optical flow techniques like differential methods, including the pioneering Horn and Schunk and Lucas-Kanade approaches and other correlation-based methods. In [18] reviewed the effectiveness of optical flow methods in detecting moving objects. Verri et al. [19] analyzed differential techniques for optical flow, while Barron et al. [20] evaluated the performance of various optical flow methods. Recently, Ma et al. [21] studied yet another technique for optical flow called the Farneback technique, which is a dense approach. Several studies have focused on application-driven motion analysis. Perš et al. [29] used optical flow histograms for efficient body motion representation. In [30] leveraged optical flow for moving object detection and tracking in traffic surveillance applications.

Deep Learning Approaches

Deep learning methods are popular as they find applications in every scenario, including the Micro Expression (ME) recognition. To study this topic, Li et al. [22] provided a report on deep learning methods used for ME recognition. This study includes concepts like convolutional neural networks (CNN), region-based convolutional neural networks (RCN), and long short-term memory (LSTM). They utilized widely used and available datasets like the CASME II [28]. Liong et al. [23] expressed interest in the advancements of deep neural networks in ME recognition. This motivated them to design a Shallow Triple Stream Three-dimensional CNN (STSTNet) to extract details of MEs while keeping the system computationally light. Wang et al. [24] explored 2D-3D CNN approaches for micro-expression detection.
The research on various optical flow estimation techniques resulted in progressive insights to build improved systems for ME recognition. Yet, limitations like handling illumination variance and noise showcase the need for new

optimized methods. While providing better accuracy, algorithms like Farneback and Horn-Schunck are computationally expensive, which becomes a limitation for real-time applications like ME recognition. Traditional optical flow methods struggle with varying lighting conditions, negatively impacting motion estimation accuracy. A lack of temporal information limits standard LBP. This makes working with videos much harder. LBP also struggles with non-monotonic lighting conditions, affecting feature extraction. Deep learning methods showed promising results. Nonetheless, this has some shortcomings, too. Issues like dependency on large datasets for training, computational complexity, and operation as black boxes make them challenging to implement in many applications.
Although deep learning approaches have shown strong performance, they typically require larger annotated ME corpora and heavier training pipelines. Since CASME II is relatively small and our aim is to design a reproducible, lightweight solution, we prioritize a handcrafted fusion strategy in this work.
This research addresses these limitations by proposing a Cumulated Optical Flow Vector (COFV) that enhances optical flow estimation by extracting successive frame variations, making it more robust to illumination changes and noise. To make it computationally efficient while maintaining accuracy, the Lucas-Kanade method is used. This research also contributes by integrating the LBP-TOP (Local Binary Pattern on Three Orthogonal Planes) with the COFV, making it well-suitable for video-based analysis. By utilizing this hybrid approach, this research also reduces reliance on large, making it an efficient and effective solution for ME recognition.

METHODOLOGY

Overview

Traditional optical flow methods often have some limitations regarding illumination variance and minor movement detection. On the other hand, texture-based methods like Local Binary Pattern on Three Orthogonal Planes (LBP-TOP) lack precise motion information evident from the literature survey. To overcome these challenges, the proposed method introduces a Cumulated Optical Flow Vector (COFV) to improve motion estimation across successive frames while preserving very minute movement details.
The methodology proposed in this paper follows a structured flow of using various techniques, as shown in Figure 1. Firstly, the Micro Expression (ME) input video is given from which the standard preprocessing steps, including frame extraction and RGB to grayscale conversion, are performed. This results in the pre-processed input frames. Then, we get to the Motion Estimation part, where the Lucas-Kanade approach of optical flow is employed to extract motion vectors from consecutive frames, which are then used in the calculation of the COFV. This stated descriptor captures the subtle and prominent motions, making it suitable for varying lighting conditions. Alongside this process, the LBP-TOP is applied to the frames to extract the spatiotemporal texture features. Finally, a feature fusion mechanism combines COFV and LBP-TOP features, leveraging motion and texture information for improved micro-expression recognition. The final fused feature vector is classified using a Support Vector Machine (SVM) with an RBF kernel, chosen for its effectiveness on medium-dimensional handcrafted features.
Thus, this proposed methodology enhances the robustness and accuracy of ME recognition and analysis by considering the limitations of optical flow and texture-based methods. The detailed breakdown of these is discussed below.

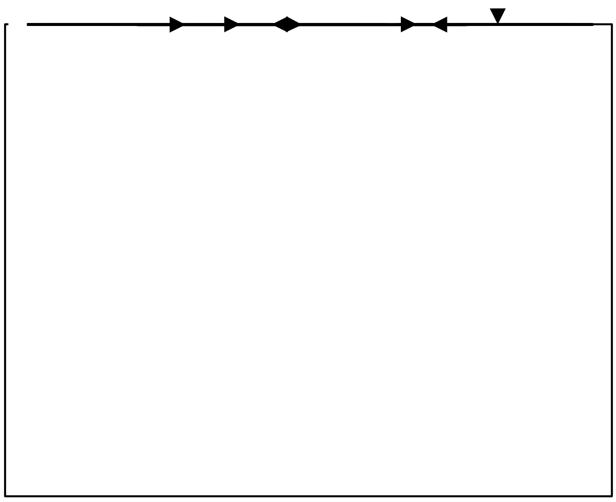

Fig.1.Proposed COFV–LBP-TOP framework for micro-expression recognition.

Preprocessing steps

The proposed methodology includes some key preprocessing techniques applied to an input ME video, which are frame extraction and changing of frames in RGB format into grayscale format.
Let V be the input ME video consisting of N frames as the following

$$V = \{F_1, F_2, \ldots, F_N\} \quad (1)$$

These frames as mentioned in above eq. (1) are then converted into grayscale. This is done using the standard luminance transformation formula as follows

$$I(x,y) = 0.2989 \cdot R(x,y) + 0.5870 \cdot G(x,y) + 0.1140 \cdot B(x,y) \quad (2)$$

Here, the symbols R, G and B in the eq. (2) represent the red, green, and blue pixel intensities at coordinates (x, y). This step reduces computational complexity and ensures homogeneity, making the process fast and efficient. These pixel intensity values are further normalized using min-max normalization.

Optical Flow Estimation

The motion vectors represent the direction of any movement in the pixels. We employ the Lucas-Kanade approach for this. This method assumes that the displacement of the image contents is minute in a specified small amount of time. The motion will also be believed to be constant (approximately) in range of a small neighbourhood. For a given pixel (x, y), let I (x, y, t) be the intensity of that pixel at some time t. Then, the optical flow constraint equation will be

$$I_x u + I_y v + I_t = 0 \quad (3)$$

Here, I_x and I_y denote the spatial intensity gradients, I_t is the temporal gradient, and (u, v) represent the horizontal and vertical components of the optical flow vector at pixel (x, y).We utilize a weighted least-squares approach over a small window w (x, y) to solve for the variables.

After this, to capture the cumulative motion features across the frames, this methodology employs the calculation of the Cumulated Optical Flow Vector (COFV) as

$$\mathrm{COFV} = \sum_{t=1}^{N-1} \sqrt{u_t^2 + v_t^2} \quad (4)$$

The eq. (4) gives us the COFV where 〖(u〗 _t,v_t) are flow vectors between frames F_t and F_(t+1)COFV presented here enhances micro-expression motion representation by aggregating motion magnitude. When integrated with LBP-TOP discussed below, this stands as a refreshed and efficient approach to ME recognition.

Local Binary Patterns on Three Orthogonal Planes (LBP-TOP)

This method extends the Local Binary Patterns (LBP) into three slices of orthogonal planes XY, YT, and XT, ensuring it captures spatial and temporal patterns. LBP-TOP is also more computationally more effective than standard LBP, as there is no need to consider the difference between the every other neighboring pixel with respect to the center pixel through the frames to compute the LBP binary code. Instead, we only take a specific number of neighboring pixels. For instance, when only three orthogonal planes (XY, XT, and YT) are considered, the feature space is reduced to 3×2^8 possible patterns instead of 2^{26}. Thus, using LBP-TOP recognizes texture and edge descriptions quickly and effectively. The binary code for successive frames is calculated as follows.

$$\mathrm{LBP}(x,y,t) = \sum_{p=0}^{P-1} s(I_p - I_c) * 2^p \quad (5)$$

As in eq. (5), I_s are the surrounding pixel intensities, I_c is the center pixel intensity while s(x)=1 if x≥0, else 0. The resulting histograms from XT, XY, as well as the YT plane are then concatenated to create the feature vector representing LBP-TOP.

Feature Fusion and Output

To develop a hybrid feature set, we concatenate the unique features derived through earlier steps of the proposed methodology: Motion estimation and Texture feature extraction. This key representation, done by integrating the feature set of COFV and LBP-TOP, ensures more accurate and improved ME recognition than traditional and existing methods. This also achieves robustness towards illumination variance and better handling minor movements in ME video input. The two feature sets are first z-score normalized and then fused using weighted concatenation.

$$F = \alpha \cdot \mathrm{Norm}(\mathrm{COFV}) \oplus \beta \cdot \mathrm{Norm}(\mathrm{LBP-TOP}), \alpha + \beta = 1 \quad (6)$$

where $\alpha = 0.6$ and $\beta = 0.4$ represent the weighting coefficients assigned to the motion (COFV) and texture (LBP–TOP) feature sets respectively, and $\oplus$ denotes the concatenation operator used to combine the normalized feature vectors.

The output of this methodological process is a high-dimensional feature vector representing ME expression dynamics. This can be further utilized as input to a classifier like a Support Vector Machine (SVM). It is also very helpful for performing clustering or visualization.

Algorithmic Flow of Proposed Solution

Algorithm 1. COFV and LBP-TOP Based Micro-Expression Recognition

Input ME video V Perform frame extraction and grayscale + normalization Compute optical flow (Lucas–Kanade) between consecutive frames Accumulate flow to form COFV descriptor Apply LBP-TOP on preprocessed video frames Normalize both feature sets and perform weighted concatenation Classify fused feature vector using SVM (RBF)

Output predicted ME label.

The above algorithm helps understand the solution proposed through the fusion strategy of COFV and LBP-TOP methods.

Experimental Analysis

Dataset Details

The CASME II dataset [28] is a well-known benchmark dataset for micro-expression recognition. It includes 255 video sequences of subtle micro-expressions of 26 Chinese people aged 22–32. The dataset collects five categories of emotion: happiness, disgust, surprise, repression, etc. Captured using high-resolution footage at 200 frames per second (fps), it has been created to time-stamp fast and fine micro-expressions between 0.05 and 0.5 seconds. For each video, also onset, apex, and offset frames are annotated.

CASME II has all the facial regions pre-aligned and lighting consistency that also reduces noise beside the human face, which is more suitable for feature extraction methods, such as the Cumulated Optical Flow Vector (COFV) or the LBP-TOP. The high temporal resolution of the data helps increase accuracy in analysing even the most subtle expressions. While CASME II is a standard benchmark for spontaneous MEs, it contains a relatively small number of subjects and imbalanced emotion categories, which may limit generalization and can lead to overfitting if cross-subject protocols are not followed. However, it serves as a valuable dataset for micro-expression recognition research.

Experimental Setup

To provide an experimental setup for the evaluation of the proposed methodology, the CASME II dataset is utilized and offers high-resolution video sequences annotated for micro-expression recognition. The implementation is done in Python, utilizing libraries like OpenCV for optical flow computation and texture analysis and Scikit-learn for classification tasks. The experiments are done utilizing cross-validation with 10 folds for robustness as well as generalizability.

Hyperparameters are highly tuned and documented for replicability. Hence, the Lucas-Kanade optical flow uses a window size of 15×15 a maximum pyramid level of 3 and a threshold of 10 iterations for convergence. The LBP-TOP feature extraction applied a radius of R=3 and uniform patterns to the three orthogonal planes to extract spatiotemporal features. Concatenating the COFV and LBP-TOP feature vectors enables feature fusion to yield a unified feature representation. A Support Vector Machine (SVM) along with a radial basis function (RBF) kernel are utilized for classification at hyperparameter settings of $C=1.0$ and $\gamma=0.01$, optimized with grid search.

The modules in the prototype application consist of frame extract, grayscaling, optical flow, LBP-TOP, and feature fusion. All videos have been pre-processed to have the same number of frames, the exact resolution (128×128 pixels), and the same input shape. This modular pipeline allows researchers to swap out or tune down parts of it. Also, for debugging and verification reasons, the implementation provides all latent outputs, like the computed motion vectors, the compiled LBP histograms, and the computed fused features. To validate and achieve reproducibility, both the random number generation and the dataset splits across all experiments utilize the same seed. Also, we share detailed documentation and code to enable other researchers to replicate the study.

RESULTS OF FEATURE EXTRACTION

The feature extraction results demonstrate the proposed methodology's effectiveness in capturing motion and texture details for micro-expression recognition. COFV effectively identifies subtle motion between frames, while LBP-TOP captures spatiotemporal texture features across three orthogonal planes, ensuring robust performance under varying conditions. The following table summarizes key feature extraction metrics.

Table 1. Quantitative Metrics of Motion and Texture Features Extracted Using COFV and LBP-TOP for ME Recognition

Feature Type	**Metric**	**Value(Mean± Std)**
Motion (COFV)	Optical Flow Magnitude	0.856 ± 0.034
	Motion Vector Variance	0.421 ± 0.022
Texture (LBP-TOP)	XY Plane Entropy	4.872 ± 0.231
	XT Plane Entropy	5.130 ± 0.189
	YT Plane Entropy	4.986 ± 0.214

As shown above, in Table 1, the motion features extracted through COFV exhibit high sensitivity to subtle micro-expressions, as evident from the optical flow magnitude and motion vector variance. Here, entropy measures the information content of the extracted feature distribution where higher is more discriminative while motion vector variance reflects the spread of the optical flow magnitudes across the face, indicating how well subtle motions are captured.These metrics confirm the method’s ability to effectively capture minor variations in facial movements. This variance is maintained throughout the sequence and proves resistant to noise and illumination changes. LBP-TOP gradually captures these texture characteristics to present spatial and temporal relationships. The variation of entropy over XY, YT as well as XT plane indicates the temporal and spatial differences in micro-expression sequences, and the XT plane entropy is the highest owing to the most clear temporal features in the sequences of ME. The COFV and LBP-TOP features complement each other and provide additional information for the recognition framework.

Performance Comparison

We distinguish the method proposed through this methodology to baseline models like Horn-Schunck, Farneback as well as Lucas-Kanade optical flow methods. The integration of COFV and LBP-TOP improves the proposed method's precision, recall, accuracy along with the F1-score compared with existing traditional methods. The following section emphasizes the hybrid method's performance abilities in adjusting for slight motions and variations in illumination to successfully detect micro-expressions.

Table 2. Performance comparison of proposed and baseline methods (mean ± standard deviation)

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
Horn-Schunck + LBP-TOP	72.8 ± 2.1	70.5 ± 2.3	69.4 ± 2.7	69.9 ± 2.5
Farneback + LBP-TOP	75.3 ± 1.9	73.1 ± 2.0	71.8 ± 1.8	72.4 ± 2.2
Lucas-Kanade (COFV) Only	78.5 ± 2.4	77.2 ± 2.5	76.0 ± 2.3	76.6 ± 2.4
Proposed (COFV + LBP-TOP)	85.2 ± 1.8	84.1 ± 1.9	83.3 ± 1.7	83.7 ± 1.8

Table 2 presents a performance comparison and suggests, proposed method (COFV + LBP-TOP) significantly is outperforming the baseline models in micro-expression recognition. It achieves higher recall, precision, accuracy along with F1-score by fusing motion (COFV) and texture (LBP-TOP) features. Unlike conventional methods, such as the likes of Horn-Schunck and Farneback approaches, the proposed method can capture subtle facial movements while dealing with variation in illumination, thus providing robust recognition.

Our method, which integrates COFV and LBP-TOP, obtains superior performance showing accuracy of 85.2% as well as F1-Score of 83.7%. The notable improvement comes from COFV and LBP-TOP features' complementary nature. COFV provides a robust estimate of the optical flow in the presence of significant challenges, such as illumination variances, and LBP-TOP [166] captures spatiotemporal texture information across three orthogonal planes. This synergy empowers the proposed method to better model the intricate temporal and spatial interactions among the biological features crucial for finely recognizing micro-expressions.

The principal reason behind the effectiveness of the proposed approach is that it can enrich the feature representation from both motion and texture information. Furthermore, COFV, with its proficiency in addressing illumination variations and rendering fine-grained motion vectors, dovetails seamlessly with the prowess of LBP-TOP in capturing spatiotemporal traits, facilitating strong and consistent performance across various contexts. In addition, the three orthogonal planes appropriately describe the subtle micro-expression patterns using LBP-TOP from the perspective of the temporal dynamics. This demonstrably superior performance across all metrics indicates this method's strong robustness and reliability for manifestation of micro-expression recognition.

Ablation Study

Table 3. Ablation study of COFV and LBP-TOP components (mean ± standard deviation)

Method	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
LBP-TOP Only	75.6 ± 1.9	74.3 ± 2.1	73.0 ± 2.2	73.6 ± 2.0
COFV Only	78.5 ± 2.4	77.2 ± 2.5	76.0 ± 2.3	76.6 ± 2.4
COFV + LBP (Without Fusion)	81.3 ± 2.0	80.1 ± 2.1	78.9 ± 2.0	79.5 ± 2.1
Proposed (COFV + LBP-TOP + Fusion)	**85.2 ± 1.8**	**84.1 ± 1.9**	**83.3 ± 1.7**	**83.7 ± 1.8**

Table 3 features the ablation study that measures the impact of each of the individual components present in the proposed method. LBP-TOP obtains moderate performance by capturing geometric texture features on spatio-temporal patterns while ignoring the obtained motion model information. Micro-expressions are states of minimal motion disturbance and are well captured by COFV(TOP) compared to existing LBP-TOP methods. The complementary information of motion and texture improves the performance when separate input is combined without fusion. The amalgamation with COFV, LBP-TOP, and feature fusion improves recall, precision, accuracy along with F1-score. This illustrates necessity of utilizing motion and texture features simultaneously and verifies the validity of the fusion approach for executing micro-expression recognition in harsh conditions.

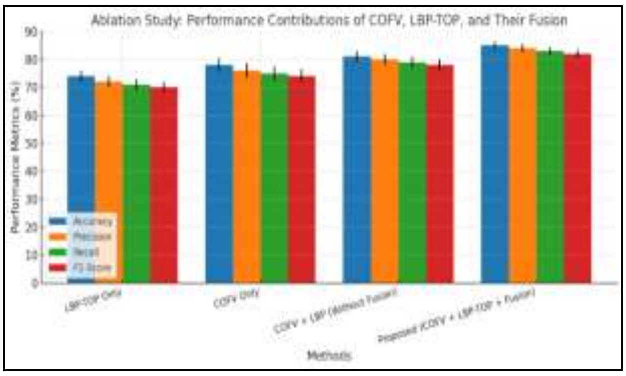

Fig. 2.Ablation Study Highlighting the Performance Contributions of COFV, LBP-TOP, and Their Fusion for ME Recognition.

As shown in Fig.2, an ablation study of the method is performed to analyze the contributability of individual components LBP-TOP and COFV, where the technique without fusion is denoted as LBP-TOP + COFV. The metrics taken into account include recall, precision, accuracy along with F1-score, with error margins included for reflecting variability in these results. This research has shown the incremental advantages of fusing COFV and LBP-TOP and the considerable gain from their combination.

The performance scores prove that the LBP-TOP method shows 75.6% accuracy and 73.6% F1-Score, indicating its efficacy for extracting spatiotemporal texture features. However, as it relies only on texture features, it cannot tackle the subtle motion deviations hidden in micro-expression. In contrast, the performance of COFV alone is 78.5% and 76.6% for accuracy and F1-Score, respectively. Because COFV is explicitly designed to robustly estimate motion vectors, it includes subtle motions indicative of micro-expressions, whereas LBP-TOP relies on static appearance changes to estimate the features. Without fusion applied to the combined single stream systems, COFV and LBP-TOP achieve an even higher accuracy of 81.3% and 79.5% , respectively, and F1-Score. By combining both components, which respectively exploit motion estimation and texture representation, we can improve recognition ability on both fronts. But without fusion, this synergy cannot be fully harnessed.

The fusion of COFV and LBP-TOP leads to the best performance in all metrics. In union with motion and texture information fusion, the approach achieves 85.2% of accuracy along with an F1-Score of 83.7%, helping low-level micro-expression complexities to be addressed. This model greatly enhanced performance by utilizing a fusion layer that captures these illumination variations and minor temporal dynamics from three orthogonal planes. The ablative study underlines the complementary role of motion and texture descriptors in micro-expression recognition. Though each part has its significant role to play in performance, when combined they take advantage of the complementary aspects of their strengths, yielding a powerful recognition system. These experimental results validate that presented method can improve micro-expression recognition and further stress that the current fusion COFV and LBP-TOP method is enhancing the performance of ME recognition.

Comparative Analysis

Table 4. Comparative Analysis of Methods Discussed in the Literature Review for ME Recognition and Motion Analysis

Method	Key Contribution/Focus	Strengths	Limitations	Reference
Completed Local Binary Pattern (CLBP)	Enhanced LBP for texture classification	Improved accuracy in texture analysis	Limited focus on spatiotemporal features	[1]
Robust Local Binary Pattern (RLBP)	Noise-resilient variation of LBP	Better performance in noisy environments	Computational complexity	[3]
LBP-TOP	Captures spatiotemporal features for videos	Effective for micro-expression recognition	Sensitive to parameter tuning	[25], [26]
Horn-Schunck Optical Flow	Global motion estimation with brightness constancy	Accurate for dense flow	Computationally expensive and illumination sensitive	[12]
Lucas-Kanade Optical Flow	Local motion estimation with the pyramidal approach	Efficient for subtle movements	May fail in large or abrupt motions	[13], [15]
Farneback Optical Flow	Dense optical flow estimation	High accuracy in dense motion	Computationally expensive	[21]

Shallow Triple Stream 3D CNN (STSTNet)	Lightweight 3D CNN for micro-expression detection	Computationally efficient	Requires training on large datasets	[23]
2D-3D CNN Approach	Combines 2D and 3D CNNs for expression recognition	Effective for temporal and spatial analysis	High computational cost	[24]
Histograms of Optical Flow	Motion representation for human body actions	Simple and effective for body motion	Less suitable for facial micro-expressions	[29]
CASME II Dataset Baseline	Benchmark micro-expression dataset	Provides a standard for evaluation	Limited participant diversity	[28], [31]

The techniques and methods presented in the literature review are explored in Table 4, demonstrating the contributions of the techniques/methods and their strengths and weaknesses. Local Binary Patterns (LBP), including its various forms like Completed LBP (CLBP) and Robust LBP (RLBP), contributed to texture recognition by assisting against noise and increasing classification accuracy. However, these approaches are mainly focused on still images and lack spatiotemporal feature extraction, which is critical in video-based applications like facial microexpression recognition. Horn-Schunck, Lucas-Kanade, and Farneback and other optical flow methods, are widely used in motion estimation. Although Horn-Schunck is well-suited for dense motion analysis, its computational expense and susceptibility to illumination changes can hinder real-time applications. While Lucas–Kanade is more optimal for easily observable differences, especially micro-expressions, similar to more subtle facial expressions, it has global motion or sudden acceleration limitations. This dense approach of Farneback enables high precision, but its computational expenses can prevent scalability.

Long history-based approaches are one of the advanced methods used in three orthogonal planes, which are called LBP-TOP approaches, which are shown to be effective for combining spatial and temporal analysis of videos; thus, LBP-TOP approaches are suited for video-based texture systems [29]. Yet, its performance is susceptible to parameter tuning. As a similar approach, methods which are based on deep learning such as STSTNet and 2D-3D CNNs have shown promising results by leveraging spatiotemporal features. However, they typically necessitate higher computational expenses and huge datasets for training. The present scenario for ME recognition has enabled the CASME II dataset to be considered the de facto standard for benchmarking. Nonetheless, its narrow sample diversity limits generalization. These approaches give insight into the limitations of motion and texture analysis, providing a rationale for the hybrid approach (COFV and LBP-TOP integration) used in motion and texture-based analysis.

CONCLUSION

This research presents a novel approach using spatial domain motion-based features and spatial domain texture feature extraction techniques. COFV is used for motion modelling, and LBP-TOP is used for spatial-temporal texture classification. Each method produces a set of features concatenated to increase its discrimination power per ME. Lucas-Kanade Optical Flow method is employed with little movement of facial muscles between frames, and LBP-TOP descriptor is also helpful for encoding local spatiotemporal texture variations. The hybrid feature vector is then extracted and applied to classification or clustering so it can be used for ME recognition in different tasks. This combination has been responsible for better recognition and less dependence on illumination and noise, two significant issues with earlier methods. This paper forwards micro-expression analysis by introducing a cumulated flow vector for motion estimation and combining it with LBP-TOP texture-based features using its compositional nature as a feature extraction method.

Therefore, the proposed COFV–LBP-TOP fusion achieved up to 7–18% improvement in accuracy over traditional optical-flow-only and LBP-TOP-only baselines on CASME II, confirming the benefit of combining motion accumulation with spatiotemporal texture encoding.

Despite the gains, the approach is still constrained by the limited size and class imbalance of current ME datasets, and performance may vary on narrow scoped recordings. Future work will explore cross-dataset training and integration with transformer-based temporal encoders.

REFERENCES

1. Guo, Z., Zhang, L. and Zhang, D., 2010. Completed modeling of local binary pattern operator for texture classification.IEEE transactions on image processing, 19(6), pp.1657-1663.
2. Ke-Chen, S., Yun-Hui, Y.A.N., Wen-Hui, C.H.E.N. and Zhang, X., 2013. Research and perspective on local binary pattern.ActaAutomaticaSinica, 39(6), pp.730-744.
3. Chen, J., Kellokumpu, V., Zhao, G. and Pietikäinen, M., 2013. RLBP: Robust Local Binary Pattern. In BMVC.
4. Mäenpää, T. and Pietikäinen, M., 2005. Texture analysis with local binary patterns.In Handbook of pattern recognition and computer vision (pp. 197-216).

5. Liu, L., Zhao, L., Long, Y., Kuang, G. and Fieguth, P., 2012. Extended local binary patterns for texture classification.Image and Vision Computing, 30(2), pp.86-99.
6. Ahonen, T., Hadid, A. and Pietikainen, M., 2006. Face description with local binary patterns: Application to face recognition. IEEE transactions on pattern analysis and machine intelligence, 28(12), pp.2037-2041.
7. Rahim, M.A., Hossain, M.N., Wahid, T. and Azam, M.S., 2013. Face recognition using local binary patterns (LBP). Global Journal of Computer Science and Technology, 13(4), pp.1-8.
8. Fleet, D. and Weiss, Y., 2006. Optical flow estimation.In Handbook of mathematical models in computer vision (pp. 237-257). Boston, MA: Springer US.
9. Sun, D., Roth, S. and Black, M.J., 2010, June. Secrets of optical flow estimation and their principles.In 2010 IEEE computer society conference on computer vision and pattern recognition (pp. 2432-2439).IEEE.
10. Otte, M. and Nagel, H.H., 1994. Optical flow estimation: advances and comparisons. In Computer Vision—ECCV'94: Third European Conference on Computer Vision Stockholm, Sweden, May 2–6, 1994 Proceedings, Volume I 3 (pp. 49-60). Springer Berlin Heidelberg.
11. Fortun, D., Bouthemy, P. and Kervrann, C., 2015. Optical flow modeling and computation: A survey. Computer Vision and Image Understanding, 134, pp.1-21.
12. Horn, B.K. and Schunck, B.G., 1981. Determining optical flow.Artificial intelligence, 17(1-3), pp.185-203.
13. Bruhn, A., Weickert, J. and Schnörr, C., 2005. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. International journal of computer vision, 61, pp.211-231.
14. Sharmin, N. and Brad, R., 2012. Optimal filter estimation for Lucas-Kanade optical flow.Sensors, 12(9), pp.12694-12709.
15. Yedjour, H., 2021. Optical flow based on Lucas-Kanade method for motion estimation. In Artificial Intelligence and Renewables Towards an Energy Transition 4 (pp. 937-945). Springer International Publishing.
16. Kumar, M. Sunil, and D. Harshitha. "Process innovation methods on business process reengineering." Int. J. Innov. Technol. Explor. Eng 8.11 (2019): 2766-2768.
17. Sangamithra, B., BEManjunathSwamy, and M. Sunil Kumar. "Evaluating the effectiveness of RNN and its variants for personalized web search." Optical and Quantum Electronics 55.13 (2023): 1202.
18. Burada, Sreedhar, B. E. Manjunathswamy, and M. Sunil Kumar. "Early detection of melanoma skin cancer: A hybrid approach using fuzzy C-means clustering and differential evolution-based convolutional neural network." Measurement: Sensors 33 (2024): 101168.
19. Natarajan, V. Anantha, and M. Sunil Kumar. "Improving qos in wireless sensor network routing using machine learning techniques." 2023 International Conference on Networking and Communications (ICNWC).IEEE, 2023.
20. Kumar, M. Sunil, et al. "Automated Extraction of Non-Functional Requirements From Text Files: A Supervised Learning Approach." Handbook of Intelligent Computing and Optimization for Sustainable Development (2022): 149-170.
21. Ma, Z., Wang, T., Xu, S., Mu, X., Wang, Q. and Guo, Q., 2023, July. Moving object Detection Based on Farneback Optical Flow.In 2023 42nd Chinese Control Conference (CCC) (pp. 7350-7355).IEEE.
22. Li, Y., Wei, J., Liu, Y., Kauttonen, J. and Zhao, G., 2022. Deep learning for micro-expression recognition: A survey. IEEE Transactions on Affective Computing, 13(4), pp.2028-2046.
23. Liong, S.T., Gan, Y.S., See, J., Khor, H.Q. and Huang, Y.C., 2019, May. Shallow triple stream three-dimensional cnn (ststnet) for micro-expression recognition.In 2019 14th IEEE international conference on automatic face & gesture recognition (FG 2019) (pp. 1-5).IEEE.
24. Wang, L., Jia, J. and Mao, N., 2020, July. Micro-expression recognition based on 2D-3D CNN. In 2020 39th Chinese control conference (CCC) (pp. 3152-3157). IEEE.
25. Mattivi, R. and Shao, L., 2009. Human action recognition using LBP-TOP as sparse spatio-temporal feature descriptor. In Computer Analysis of Images and Patterns: 13th International Conference, CAIP 2009, Münster, Germany, September 2-4, 2009. Proceedings 13 (pp. 740-747). Springer Berlin Heidelberg.
26. Kumar, M. Sunil, and K. JyothiPrakash. "Internet of things: IETF protocols, algorithms and applications." Int. J. Innov. Technol. Explor. Eng 8.11 (2019): 2853-2857.
27. Goyani, M.M. and Patel, N., 2017. Recognition of facial expressions using local mean binary pattern. ELCVIA: electronic letters on computer vision and image analysis, 16(1), pp.54-67.
28. Yan, W.J., Li, X., Wang, S.J., Zhao, G., Liu, Y.J., Chen, Y.H. and Fu, X., 2014. CASME II: An improved spontaneous micro-expression database and the baseline evaluation. PloS one, 9(1), p.e86041.
29. Perš, J., Sulić, V., Kristan, M., Perše, M., Polanec, K. and Kovačič, S., 2010. Histograms of optical flow for efficient representation of body motion.Pattern Recognition Letters, 31(11), pp.1369-1376.
30. Aslani, S. and Mahdavi-Nasab, H., 2013. Optical flow based moving object detection and tracking for traffic surveillance. International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering, 7(9), pp.1252-1256.