Jurnal Teknik Informatika (JUTIF) Vol. 6, No. 2, April 2025, Page. 593-608
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4235

Improved Micro-expression Recognition: An Apex Frame-Based Approach Feature Tracking and KLT

Priska Choirina*[1], Indah Martha Fitriani[2], Ulla Delfana Rosiani[3], Muhammad Nabil Mufti[4], Firmanda Ahmadani Arsistawa[5], Pangestuti Prima Darajat[6]

[1,5,6] Information Technology, Faculty of Science and Technology , Universitas Islam Raden Rahmat, Indonesia
[2,4] Teknik Elektro, Faculty of Science and Technology, Universitas Islam Raden Rahmat, Indonesia
[3]Information Technology, Department of Information Technology, Politeknik Negeri Malang, Indonesia

Email: [1]priska_choirina@uniramalang.ac.id

Received : Dec 18, 2024; Revised : Jan 28, 2025; Accepted : Feb 10, 2025; Published : Apr 26, 2025

Abstract

This research develops a real-time facial micro-expression recognition system, focusing on analyzing the onset and apex phases of micro-expression on the Spontaneous Activity and Micro-Movements (SAMM) dataset. Micro-expressions are very brief (0.04 - 0.2 seconds) facial muscle movements that often occur when a person is trying to hide emotions. The developed system aims to improve computation time efficiency and micro-expression recognition accuracy by optimizing feature extraction techniques and selecting more specific facial areas, including facial components such as eyebrows, eyes, and mouth. This research successfully improved the computation time efficiency by 51.96%, almost half the time required by the previous method. In addition, this study shows an increase in efficiency compared to previous studies, with an increase of 34.3% for SVM with Manual Sampling technique and 32.6% for MLP-Backpropagation. In the Random Sampling technique, SVM efficiency increased by 6.1%, but MLP-Backpropagation accuracy decreased by 4.8%. This method achieved 77.9% accuracy for MLP-Backpropagation, which is higher than the previous method. This research contributes to accelerating micro-expression recognition systems and improving accuracy, which opens opportunities for real-time emotion analysis applications such as lie detection or human behavior monitoring in a broader context.

Keywords: *DRMF, emotion classification, feature tracking, KLT, micro-expression.*

1. INTRODUCTION

Micro-expressions refer to involuntary facial muscle movements that last for a short period, usually occurring when individuals try to hide feelings in stressful situations [1]. This phenomenon is often known as emotional leakage [2] and can describe six main categories of emotions: happiness, fear, sadness, surprise, anger and disgust [3]. In contrast to macro expressions with more obvious movements, micro expressions last between 0.04 to 0.2 seconds [4] with subtle movements and are limited to certain areas of the face [5]. Previous research has shown that micro-expressions occur spontaneously and are challenging to manipulate [3]. In the context of emotion recognition, the movements' short duration and subtle nature make micro-expressions often undetectable to the naked eye, which can add complexity to the identification process [6]. Therefore, micro-expression recognition is a challenge in psychology and computer science, especially in applications such as police lie detection and social behavior analysis [7].

The ability to recognize micro-expressions considerably assists experts in identifying such expressions accurately [8], [9]. This study aims to deliver quick and accurate insights into detecting

concealed emotional shifts by comparing different phases of facial expressions. Micro-expressions are facial movements that unfold dynamically and sequentially across several stages: neutral, onset, apex, offset, and back to neutral [6], [10]. The neutral phase is the initial emotionless state, followed by the onset phase, where emotions emerge with increased muscle activity [11], [12]. The apex phase is the peak where expression is most evident; the offset phase occurs when expression begins to fade, and muscles return to a neutral state [13]. This process describes rapid and often unconscious emotional changes and requires careful analysis to recognize the type of emotion [14]. Previous research by a facial expression expert stated that frames extracted when the expression reaches its apex can be easily analyzed as emotional messages [15].

A previous study by Liong et al. [16] introduced the Bi-Weighted Oriented Optical Flow (Bi-WOOF) method, which utilizes two key frames, namely the apex (highest intensity of expression) and onset (neutral reference), for micro-expression recognition. The classification stage employed the Support Vector Machine (SVM) method with the Leave-One-Subject-Out Cross-Validation (LOSOCV) technique. This method was tested on five datasets, including CASME II and SMIC-HS, achieving the best F1-scores of 0.61 and 0.62, respectively. In terms of efficiency, on the SMIC-HS dataset, processing two frames required an average of 3.95 seconds per video, significantly faster compared to 128.71 seconds for the entire video sequence (33 times faster) using MATLAB on an Intel Core i7-4770 3.4 GHz CPU. This method successfully reduced computational time without compromising recognition accuracy.

Precise identification of facial feature points is essential for micro-expression recognition. Friesen and Ekman [17] in Vitale et al. [18] highlighted six important regions: eyes, eyebrows, nose, cheeks, and mouth, whereas Lekdioui et al. [19] categorized facial landmarks into three groups: mouth, left eye + eyebrow, and right eye + eyebrow. Extracting features from specific facial areas reduces irrelevant data and improves processing speed. This study combines facial components (eyebrows, eyes, and mouth) using Regions of Interest (ROI) but focuses only on a small subset of marker points representing facial muscle movements, determined using the Discriminative Response Map Fitting (DRMF) method.

The proposed research builds on Choirina et al.'s work [13], which used entire frames from the SAMM dataset and achieved unsatisfactory accuracy (56.3% for SVM and 58.6% for MLP-Backpropagation) due to potential data redundancies in handcrafted data. Choirina et al. applied two data settings: handcrafted data (60% training, 40% testing) and Random Sampling. This study introduces a new approach to micro-expression recognition by comparing the apex and neutral phases, inspired by Liong et al. [16], utilizing apex frame data from the SAMM dataset and expression classes like disgust, fear, happiness, and sadness. The approach uses Kanade Lucas Tomasi (KLT) for tracking facial components such as eyebrows, eyes, and mouth corners, with points labeled using the DRMF method to improve accuracy [20]

This research introduces a novel approach by comparing the apex and onset phases, utilizing apex frame data from the SAMM dataset to minimize data redundancy and enhance recognition accuracy. The KLT feature point tracking method targets specific facial components (eyebrows, eyes, and mouth corners), with points formed by DRMF, aiming to improve feature extraction efficiency and accuracy. The study focuses on recognizing disgust, fear, happiness, and sadness expressions in the SAMM dataset, evaluating whether this method reduces data redundancy and boosts computational efficiency compared to previous approaches. The findings could contribute to the development of more precise and efficient micro-expression recognition techniques, including applications in lie detection.

Jurnal Teknik Informatika (JUTIF) Vol. 6, No. 2, April 2025, Page. 593-608
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4235

2. METHOD

The research framework (see Figure 1) focuses on micro-expression recognition using apex frames by tracking feature points in facial regions like eyebrows, mouth corners, and eyes. A manual observation of the "surprise-015_5_3" video revealed no visible movement in the left eyebrow during the neutral to onset phase, while the onset to apex phase showed eyebrow movement. In the apex to offset phase, the eyebrow returned to its initial position, indicating a return to a neutral expression. This analysis suggests that comparing image frames with significant movement can simplify the identification of micro-expression phases.

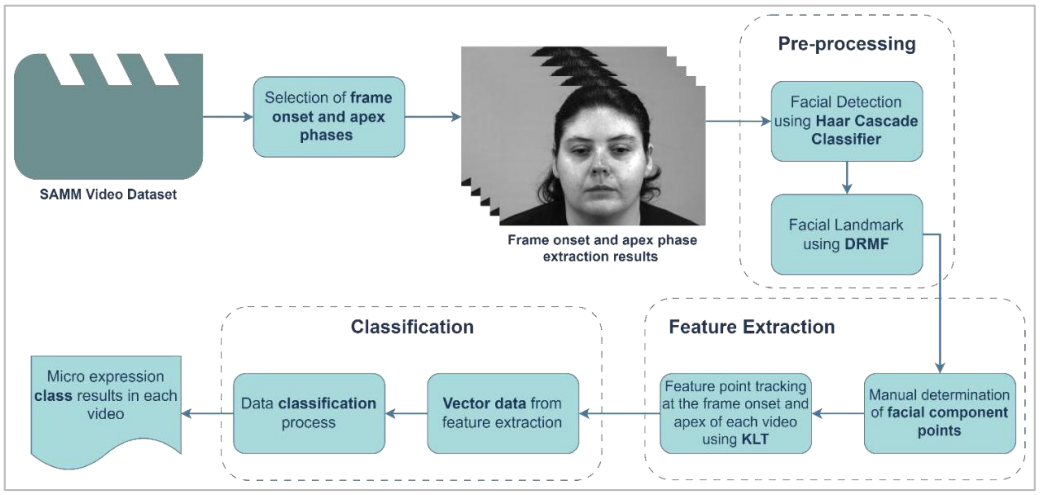


Figure 1. Framework of Apex Frame-Based Feature Point Tracking for Micro-expression Recognition using the SAMM Dataset

2.1. Selection of Frame Onset & Apex Phases

Frame extraction from SAMM videos is performed manually, selecting the onset and apex phases based on the labels provided in the dataset. Each video is extracted from the onset to offset frames, with the micro-expression phases encoded. The onset and apex frames are manually marked and stored in folders according to the expression class, as shown in Figure 2. For example, in the video "surprise-015_5_3," the onset frame is located at the 6050th frame, the apex at the 6073rd frame, and the offset at the 6127th frame. In this study, only the two main frames—onset and apex—are used to analyze the movement of facial components.

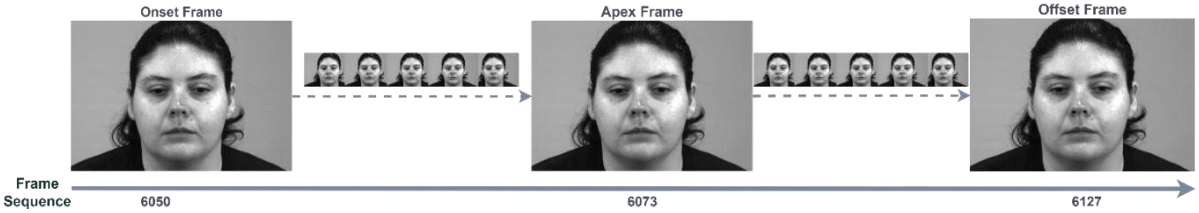


Figure 2. Onset to offset phase frames on SAMM video

2.2. Facial Detection using Haar Cascade Classifier

The pre-processing of the input data, i.e. the onset and apex frames, includes two stages: face location detection and face component point determination. Face detection is performed using the Haar Cascade Classifier, which is effective for real-time face detection with high accuracy from different angles and positions [20], [21]. Each frame, at the onset and offset phases, is extracted with

Jurnal Teknik Informatika (JUTIF) Vol. 6, No. 2, April 2025, Page. 593-608
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4235

Haar features to detect contrast differences between light and dark areas. The Adaboost algorithm selects the best features organized in a cascade structure to filter candidate areas. The detection process uses a sliding window at various scales until the entire frame area is processed, resulting in the Region of Interest (RoI) shown in Figure 3.

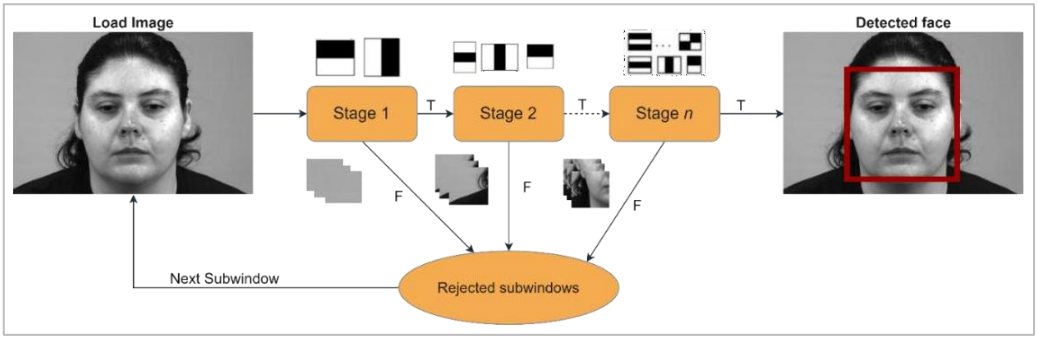


Figure 3. Illustrates the stages of face detection using the Haar Cascade Classifier

2.3. Facial Landmark using DRMF

In this study, landmark detection on the face uses the DRMF method, which identifies 66 points (see Figure 4) [20], [22], [23] across different facial regions. These points remain consistent even with movement, rotation, or scaling of the face, allowing for accurate facial component determination. The face can be aligned using three key points from the eyes and mouth. The DRMF method, based on a discriminative regression approach within the Constrained Local Model (CLM) framework, handles dynamic backgrounds, occlusion, and lighting variations effectively, ensuring high computational speed for real-time applications. With an average facial component box formation time of 0.08 seconds per frame, DRMF is ideal for real-time micro-expression recognition [20], [24].

Figure 4. Distribution of 66 feature points with the DRMF method [25]

2.4. Manual Determination of Facial Component Points

This stage defines the formation of 20 facial feature points across components: the left eyebrow (points 2–5) and right eyebrow (points 7–10), each with 4 points to track shape and position; the left eye (points 20–23) and right eye (points 26–29), each with 4 points to identify shape and expression; and the corners of the mouth (points 32, 35, 38, and 41), represented by 4 points. Table 1 details the feature points and their distribution.

Table 1. Component Feature Points Face Area [13]

Facial Features	Points
Left Eyebrow	2-5
Right Eyebrow	7-10
Left Eye	20-23
Right Eye	26-29
Mouth Corner	32, 35, 38, and 41

2.5. Feature Point Tracking using KLT

The KLT method can detect subtle movements and track facial features accurately under various lighting conditions. The KLT method uses an optical flow approach to track feature points with sufficient texture to maintain high accuracy even when the face is moving [13]. Another advantage of KLT is that it can detect feature points with suitable characteristics for tracking attributes [26]. Good and strong feature points with striking textures can be selected using the DRMF method to determine feature points in a particular area. Point tracking with KLT involves tracking the movement of the tracked point between the current frame and reference frame [27], [28], shown in Figure 5. The KLT method can track motion in the face area through two steps, namely: a) find predefined feature points in the current frame (onset frame) and b) track those points based on the displacement in the reference frame (apex frame).

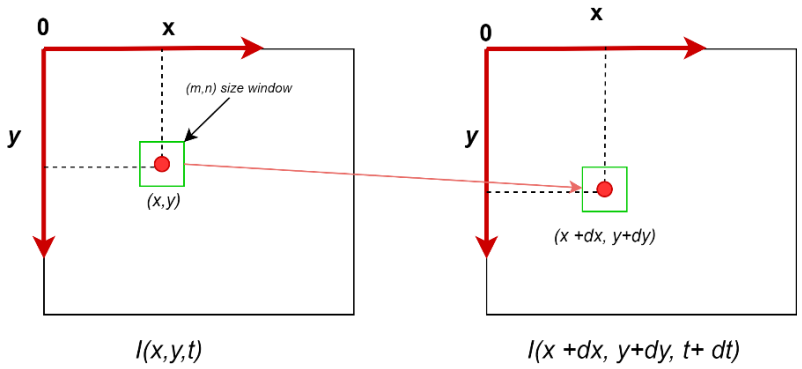


Figure 5. Illustration of point movement in the current frame (apex frame) compared to the reference frame (onset frame) [24]

The KLT process for tracking points on the face involves several key steps based on the principles of optical flow. Feature points initialized using the DRMF method are tracked from one frame to the next by calculating the positional shift of each point using the optical flow equation, as shown in Equation 1:

$$I(x, y, t) = I(x + \delta_{x,y} + \delta_y, t + 1) \quad (1)$$

where I represents the pixel intensity at position $(x,\ y)$ and time (t), and $(\delta_x,\ \delta_y)$ denote the positional change between two frames [24], [29]. This process is repeated for each frame in the video, enabling real-time feature point tracking. To minimize errors in motion estimation, the KLT algorithm minimizes the error function shown in Equation 2:

$$\epsilon(d) = \iint W[A(x) - B(x + d)]^2 dx \quad (2)$$

where $A(x)$ and $B(x\ +\ d)$ represent the pixel intensities in the initial and subsequent frames, respectively [30]. The solution for the displacement d is obtained by solving the system of linear equations derived from differentiating this error function.

P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4235

2.6. Vector Data from Feature Extraction

The feature extraction process in feature point tracking involves capturing the characteristics of the detected points' movements. After completing the feature point tracking stage, the tracking data produces the coordinates of the points in each video frame. From this data, vectors can be generated to represent the movement of the feature points [13]. First, for each feature point p at time t, its coordinates can be defined as $x_{p,t}$; $y_{p,t}$. To calculate the positional change or displacement of the feature point between the onset frame $(t-1)$ and the apex frame (t), Equations 3 and 4 are used:

$$C_X = (x_{p,t} - x_{p,t-1}) \quad (3)$$

$$C_Y = (y_{p,t} - y_{p,t-1}) \quad (4)$$

where C_X and C_Y are the components of positional change along the x and y-axes. These values provide information about the distance and direction of the feature point's movement from the onset frame to the apex frame. These values provide information about the distance and direction of the feature point's movement from the onset frame to the apex frame. Next, the magnitude of the displacement can be calculated using the Euclidean distance formula between two points, as expressed in Equation 5:

$$|\vec{PQ}| = \sqrt{(x_{p,t} - x_{p,t-1})^2 + (y_{p,t} - y_{p,t-1})^2} \quad (5)$$

where $|\ PQ\ |$ represents the magnitude of the displacement vector between the initial and final positions of the feature point. The magnitude provides information about the extent of the movement. Extracting the direction or orientation of the displacement vector is also important. The orientation, measured in degrees or radians, can be calculated using the inverse tangent (arctangent) function, as shown in Equation 6:

$$O(x,y) = tan^{-1}\left(\frac{C_Y}{C_X}\right) \quad (6)$$

In Equation 5, $O(x,\ y)$ is the orientation of the displacement vector formed by the C_X and C_Y components. The result of this calculation yields the angle that indicates the movement direction relative to the horizontal axis. The feature extraction process produces a vector comprising coordinate components (positional changes), magnitude (movement size), and orientation (movement direction).

2.7. Classification

The classification process in this study utilizes two main algorithms, namely SVM [31] and MLP-Backpropagation [32]. With a Radial Basis Function (RBF) kernel, SVM effectively handles high-dimensional data and reduces overfitting. SVM parameter settings include $Cost\ (C) = 1$ and epsilon $(\varepsilon) = 0.10$, which allows finding the optimal hyperplane to separate classes by maximizing the margin between nearby data points. On the other hand, MLP uses a backpropagation algorithm with 100 neurons in the hidden layer, a ReLU activation function to overcome vanishing gradient, and an Adam solver for adaptive learning rate adjustment. This setup allows MLP to capture complex patterns in the data efficiently. Both methods were chosen for their advantages in handling non-linear and complex data, with SVM excelling in stability and class separation, while MLP offers flexibility in handling more complex patterns.

The confusion matrix is a fundamental tool for evaluating the performance of a classification model, comparing actual and predicted labels. It consists of four components: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) [33]. Using these components, several key performance metrics can be calculated. Accuracy measures the proportion of correctly classified instances out of the total instances, calculated as Equation 7.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4235

It provides a general sense of model performance but may be less reliable for imbalanced datasets. Precision represents the proportion of correctly predicted positive instances out of all predicted positives, defined as Equation 8.

$$Precision = \frac{TP}{TP+FP} \tag{8}$$

This metric is crucial when minimizing false positives is important. Recall (or sensitivity) measures the proportion of correctly predicted positive instances out of all actual positives, calculated as Equation 9.

$$Recall = \frac{TP}{TP+FN} \tag{9}$$

High recall is essential in cases where missing positive instances has significant consequences. The F1-Score combines precision and recall into a single metric (see Equation 10).

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{10}$$

It is particularly useful for imbalanced datasets where accuracy alone may be misleading. Together, these metrics provide a comprehensive evaluation of a model's performance, highlighting its strengths and weaknesses in terms of accuracy, error rates, and the ability to correctly classify positive cases.

3. RESULT

3.1. Data set

The SAMM Video Dataset [34] is designed to analyze and understand micro-facial expressions, as well as spontaneous and involuntary facial movements. The dataset consists of high-speed videos (200 fps), which allow for the capture of rapid facial movements. These videos were recorded from a diverse group of participants, enhancing the representativeness of the data. Expressions in the dataset are coded using the Facial Action Coding System (FACS), a standardized method for describing all visible facial movements. Each facial muscle movement is identified as an Action Unit (AU), enabling objective analysis of the captured expressions. The SAMM dataset includes 159 videos, covering expression classes such as anger, contempt, disgust, fear, happiness, sadness, surprise, and others [34].

Table 2. SAMM Video Sample Data Analysis

Class Expression	Number of Data		Total
	Training	Testing	
Anger	34	23	57
Contempt	7	5	12
Disgust	5	4	9
Fear	4	4	8
Happiness	15	11	26
Sadness	3	3	6
Surprise	9	7	16

3.2. Preprocessing Result (Face Detection and Facial Feature Point Determination)

P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4235

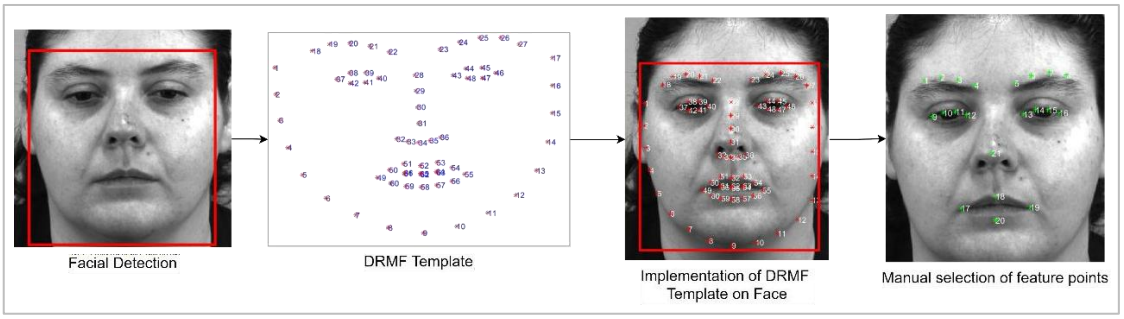


Figure 6. Implementation of 66 feature points from DRMF on SAMM Dataset

The DRMF method implementation begins with face detection using the Haar Cascade Classifier (Figure 3) as input data. The process starts with initialization, where the system determines initial landmark points on the face. In the response maps generation stage, discriminative response maps are created for each landmark point to indicate their possible locations. The fitting process uses the Gauss-Newton method to iteratively optimize landmark positions by minimizing errors. Finally, the refinement stage adjusts the results using a global face shape model to ensure anatomically accurate landmark configurations.

3.3. Feature Extraction Results

Figure 6 provides information on the displacement of point 7 after the tracking process. In the current frame, point 7 is located at coordinates $(x,\ y) = [575,221]$, while in the reference frame, a shift occurred to $(x,\ y) = [576,220]]$. This indicates that a displacement of the point occurred between the two frames. Subsequently, the distance of the displacement for all coordinate points will be calculated using Equation 1 and Equation 2.

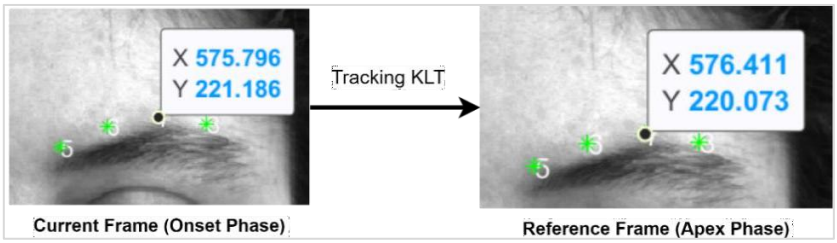


Figure 7. The displacement coordinates the left eyebrow region between neutral frame and apex frame

The motion vector feature analysis in Table 3. Motion Features Extraction Data KLT Tracking Results involves monitoring changes in the positions of facial feature points (such as eyebrows, eyes, and mouth corners) from the onset frame to the apex frame in the video sequence ("surprise-015_5_3"). Each feature point has two primary coordinates: Onset Frame Coordinates (the point's position in the initial frame) and Apex Frame Coordinates (the point's position in the peak frame). The positional change between these two frames produces vector components (R_X, R_Y), representing the displacement along the X and Y axes, respectively, as well as the magnitude $(|\overrightarrow{PQ}|)$, which measures the distance of displacement between the two points. The magnitude is calculated using the Euclidean distance formula provided in Equation 5.

P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4235

Table 3. Motion Features Extraction Data KLT Tracking Results

Features	Points	Onset Frame Coordinates		Apex Frame Coordinates		Vector Component *(Motion Features)*			
		x	y	x	y	RX	RY	$\lvert\vec{PQ}\rvert$	O(x, y)
Left Eyebrow	*1*	*331*	*235*	*330*	*236*	*1*	*-1*	*1.41*	*135*
	2	*363*	*228*	*362*	*230*	*1*	*-2*	*2.24*	*153*
	3	*396*	*233*	*396*	*234*	*0*	*-1*	*1.00*	*180*
	4	*430*	*244*	*429*	*245*	*1*	*-1*	*1.41*	*135*
Right	*5*	*508*	*240*	*507*	*241*	*1*	*-1*	*1.41*	*135*
Eyebrow	*6*	*541*	*226*	*540*	*227*	*1*	*-1*	*1.41*	*135*
	7	*576*	*219*	*575*	*221*	*1*	*-2*	*2.24*	*153*
	8	*610*	*224*	*609*	*225*	*1*	*-1*	*1.41*	*135*
Left Eye	*9*	*348*	*305*	*347*	*306*	*1*	*-1*	*1.41*	*135*
	10	*369*	*298*	*368*	*299*	*1*	*-1*	*1.41*	*135*
	11	*392*	*296*	*391*	*297*	*1*	*-1*	*1.41*	*135*
	12	*414*	*302*	*414*	*303*	*0*	*-1*	*1.00*	*180*
Right Eye	*13*	*526*	*301*	*525*	*302*	*1*	*-1*	*1.41*	*135*
	14	*549*	*292*	*548*	*293*	*1*	*-1*	*1.41*	*135*
	15	*573*	*292*	*572*	*293*	*1*	*-1*	*1.41*	*135*
	16	*595*	*298*	*594*	*300*	*1*	*-2*	*2.24*	*153*
***M*outh**	*17*	*403*	*485*	*403*	*486*	*0*	*-1*	*1.00*	*180*
	18	*470*	*461*	*470*	*462*	*0*	*-1*	*1.00*	*180*
	19	*539*	*483*	*538*	*484*	*1*	*-1*	*1.41*	*135*
	20	*471*	*508*	*470*	*509*	*1*	*-1*	*1.41*	*135*

For example, for the feature point Left Eyebrow 1, the positional change from the onset frame (331, 235) to the apex frame (330, 236) results in vector components $RX = 1$ and $RY = -1$, with a magnitude of 1.41 and an orientation angle of 135°. This type of analysis is performed for all feature points across various facial regions, including the eyebrows, eyes, and mouth. The magnitude values provide insight into the extent of movement at each feature point between the two frames, while the orientation angle (θ) indicates the direction of movement relative to the initial position. Overall, this motion vector feature analysis enables the mapping of micro-muscular movements on the face, which can be used to detect changes in facial expressions or facial movements in videos. Larger magnitude values indicate more significant movement, while the vector components and orientation offer deeper information about the direction and type of movement occurring at each facial feature point.

3.4. Classification Results using SVM and MLP-Backpropagation

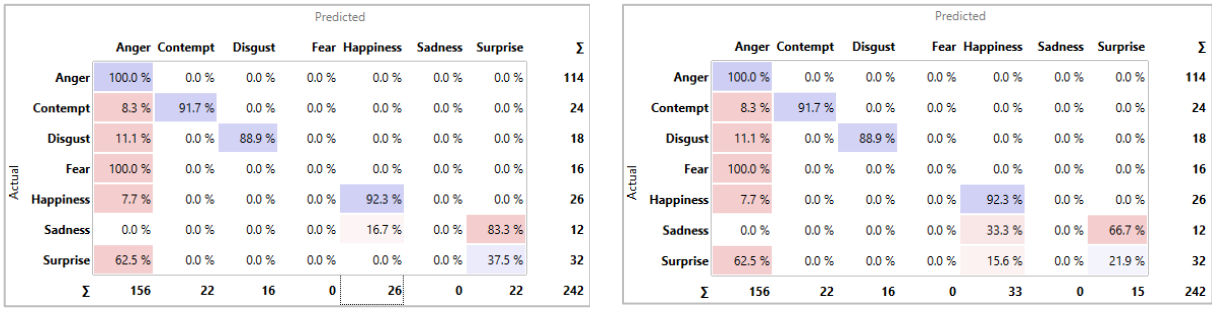
This study compares the classification results using two sampling techniques, namely manual sampling and random sampling, with 60% of the data used for training and 40% for testing, as explained in Table 3. Manual sampling allows the selection of samples based on specific criteria, while random sampling reduces bias by giving each population element an equal chance of being selected. The data is analyzed 10 times using random sampling for training and testing, comparing two classification methods: Support Vector Machine (SVM) and Multi-Layer Perceptron (MLP). SVM uses a Radial Basis Function (RBF) kernel to map the data to a higher-dimensional space, while MLP applies the backpropagation algorithm and a logistic sigmoid activation function. Emotion classification is based on motion features from the onset and apex frames, with fast data processing due to the small feature area, a learning rate of 0.0001, and 200 training epochs.

Table 4. Classification Results of Manual and Random Sampling for Extraction Features Data

Classification Method	Manual Sampling (handcrafted) (%)				Random Sampling (%)			
	Acc	F1	Recall	Precision	Acc	F1-Score	Recall	Precision
SVM	75,6	68,9	65,8	75,6	78,1	72,6	69,4	78,1
Backpropagation	77,7	72,1	68,9	77,7	76	69,6	66,4	76,6

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 2, April 2025, Page. 593-608
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4235

A. Manual Sampling Technique (Handcrafted)



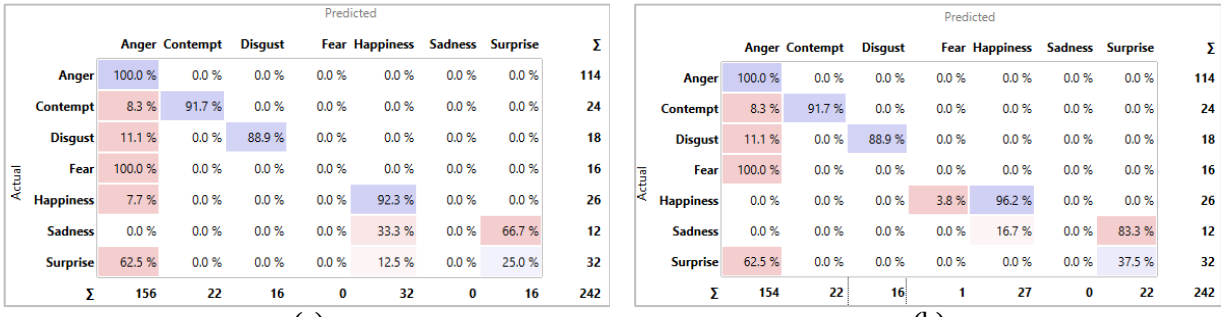
(a) (b)

Figure 8. Confusion Matrix Manual Sampling – MLP Backpropagation (a) dan SVM (b)

Figure 8 shows the classification results using MLP Backpropagation, achieving an accuracy of 77.7%, an F1-Score of 72.1%, precision of 68.9%, and recall of 77.7%. MLP demonstrates good performance, particularly in recall, which indicates that this method is effective at detecting existing micro-expressions. However, the slightly lower precision suggests that this method results in some false positives. On the other hand, SVM achieves an accuracy of 75.6%, an F1-Score of 68.9%, precision of 65.8%, and recall of 75.6%. Although SVM's accuracy and recall are slightly lower compared to MLP Backpropagation, SVM's precision is also lower, indicating a higher number of false positives compared to MLP Backpropagation, as shown in Table 4.

B. Random Sampling Technique

Figure 9 shows the classification results using MLP-Backpropagation with the Random Sampling technique, achieving an accuracy of 76%, an F1-Score of 69.6%, precision of 66.4%, and recall of 76.6%. MLP provides good accuracy with a high recall, indicating the model's ability to detect a wide range of micro-expressions. However, the lower precision suggests the presence of some false positives. On the other hand, SVM achieves an accuracy of 78.1%, an F1-Score of 72.6%, precision of 69.4%, and recall of 78.1%. SVM shows better overall performance, especially in precision and F1-Score, indicating that this model is more effective in identifying micro-expressions with fewer errors. Although both models perform well, SVM (see Table 4) is superior in terms of stability and overall classification accuracy.



(a) (b)

Figure 8. Confusion Matrix Random Sampling – MLP Backpropagation (a) dan SVM (b)

3.5. Computational Time Analysis

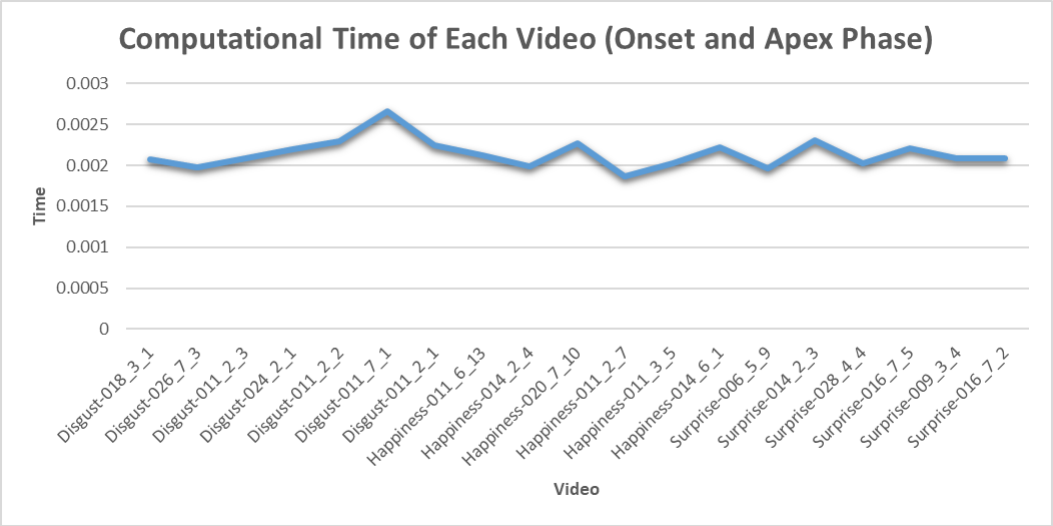


Figure 9. Computational Time of Each Video (Onset and Apex Phase)

This research identifies micro-expressions with real-time processing time using MATLAB software on a computer with specifications of 11th Generation Intel Core i3, 20GHz RAM, and 64-bit Windows Operating System. The average computation time for each video is about 0.0021 seconds, with slight variations between videos. In Figure 9, the computation time of the first video (Disgust-018_3_1) is 0.002069881 seconds, while the second video (Disgust-026_7_3) is lower, at 0.001973608 seconds. Other videos, such as Disgust-011_7_1, took 0.002658098 seconds, and Happiness-011_2_7 took 0.001867853 seconds. Although there are small fluctuations, the difference in computation time between videos is insignificant, indicating a consistent time efficiency of about 0.0021 seconds.

4. DISCUSSIONS

4.1. Comparison of Processing Time with Previous Research

This study demonstrates an improvement in the efficiency of micro-expression processing compared to previous research by Liong et al. [16], which used the Bi-WOOF method for feature extraction with an average processing time of 3.9499 seconds per video. The Bi-WOOF approach involves the onset and apex frames, as well as feature extraction from large blocks of the face area, resulting in relatively high computational time. In contrast, this study records an average processing time of 1.897 seconds per video, marking a 51.96% improvement in processing efficiency, nearly halving the time required by the previous method. This efficiency is achieved through the optimization of feature extraction techniques and the selection of more specific facial feature areas, avoiding the drawbacks of using large blocks that slow down processing.

The significant difference in processing time has a major impact on facial expression recognition applications, particularly in real-time scenarios. With faster processing times, the method in this study is better suited for practical applications such as automated surveillance systems, human-machine interaction, and behavior analysis tools. This efficiency not only improves system response speed but also enables the processing of more data in a shorter amount of time, making it more reliable for implementation in real-world scenarios that demand real-time performance.

4.2. Comparison of Micro-Expression Recognition Classification with Previous Research

The comparison of classification accuracy between this study and Choirina et al. [13] shows a notable difference, with this study achieving higher accuracy. Choirina et al. [13] used feature

extraction from the entire frame, resulting in accuracies of 56.3% for SVM and 58.6% for MLP-Backpropagation, which improved to 73.6% for SVM and 79.8% for MLP with Random Sampling. In contrast, this study, which focuses on feature extraction from the onset and apex phases, achieved 75.6% for SVM and 77.7% for MLP-Backpropagation using Manual Sampling, and 78.1% for SVM and 76% for MLP-Backpropagation with Random Sampling.

This difference indicates that a more focused feature extraction approach on the onset and apex phases is more effective. The distinction captures micro-expression changes, leading to a more significant accuracy improvement. This method shows an efficiency improvement of about 34.3% for SVM and 32.6% for MLP-Backpropagation in Manual Sampling compared to the research by Choirina et al. [9]. For Random Sampling, the efficiency improvement for SVM is 6.1%, while for MLP-Backpropagation, the accuracy efficiency decreases by 4.8%.

Table 5. Comparison of Classification Accuracy Results on the SAMM Dataset Between This Study and Choirina et al. [9]

Author	Dataset	Frame	Features Extraction	Accuracy	
				Manual Sampling	Random Sampling
Choirina, dkk[9]	SAMM	Seluruh frame	KLT	- SVM = 56,3% - MLP-Backpro = 58.6	- SVM = 73,6% - MLP-Backpro = 79.8%
This Study	**SAMM**	**Fase onset dan Apex**	**KLT**	- **SVM = 75,6%** - **MLP-Backpro = 77.7%**	- **SVM = 78,1%** - **MLP-Backpro = 76%**

This study makes a significant contribution to the field of computer science, particularly in the development of real-time micro-expression recognition systems. By optimizing feature extraction techniques and selecting more specific facial feature areas, this research successfully reduces processing time by nearly half compared to the previous Bi-WOOF method. This efficiency enables the application of facial expression recognition technology in real-time response applications, such as automated surveillance systems, human-machine interaction, and behavior analysis. The higher processing speed also opens opportunities for processing large volumes of data more efficiently, enhancing system performance and reliability in real-world scenarios that require instant and accurate performance. Therefore, this study not only refines the technical aspects of micro-expression recognition but also expands its applications across various sectors, accelerating the adoption of this technology in both industry and further research.

5. CONCLUSION

This study makes a significant contribution to the development of real-time facial micro-expression recognition systems, a relevant topic in the field of computer science, particularly for applications requiring quick responses. By comparing the methods used in this study with previous approaches, such as Bi-WOOF, the research successfully improves processing efficiency by 51.96% and accuracy by 34.3% using the SVM method with Manual Sampling. Optimizing feature extraction techniques and selecting more specific facial areas results in faster processing times, making the system more efficient and practical for various real-world applications, such as behavior analysis (e.g., lie detection).

The impact of the accuracy and computational time differences achieved in this study is significant. Faster processing time opens the potential for using this system in a wider range of real-time scenarios, where rapid and accurate responses are crucial for analyzing facial micro-expressions. These improvements also enhance the system's ability to handle large data volumes more efficiently in a shorter period. However, the accuracy difference between the tested methods and previous research needs more attention, especially in systems that are sensitive to detection errors. Although this study

shows better efficiency, the increased processing speed might sacrifice accuracy, requiring further optimization.

A limitation of this study lies in the influence of rapid facial movements and head motions beyond the scope of micro-expression, which can interfere with feature extraction analysis. As a suggestion for improvement, future research could consider using more robust and adaptable feature extraction techniques to address these variations. Additionally, enhancing the model to tackle the trade-off between speed and accuracy, along with further testing on more diverse datasets, will strengthen the real-world application of this system. Overall, this research provides a strong foundation for the development of real-time facial micro-expression recognition systems and opens many opportunities for practical applications across various industries.

CONFLICT OF INTEREST

The authors declared that there is no conflict of interest between the authors or with research object in this paper.

ACKNOWLEDGEMENT

The research team expresses their gratitude to the Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi for providing support through information resources and research funding via the Penelitian Dosen Pemula (PDP) scheme for the 2024 fiscal year.

REFERENCES

[1] X. Zeng *et al.*, "Affection of facial artifacts caused by micro-expressions on electroencephalography signals," *Front Neurosci*, vol. 16, p. 1048199, Nov. 2022, doi: https://doi.org/10.3389/fnins.2022.1048199.

[2] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *Journal of Personality and Social Psychology*, vol. 17, no. 2, pp. 124–129, 1971, doi: 10.1037/h0030377.

[3] P. Ekman, "Lie catching and microexpressions," *The philosophy of deception*, vol. 1, no. 2, p. 5, 2009.

[4] X. Ben *et al.*, "Video-based Facial Micro-Expression Analysis: A Survey of Datasets, Features and Algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021, doi: 10.1109/TPAMI.2021.3067464.

[5] B. Allaert, I. M. Bilasco, and C. Djeraba, "Micro and Macro Facial Expression Recognition Using Advanced Local Motion Patterns," *IEEE Transactions on Affective Computing*, vol. 13, no. 1, pp. 147–158, Jan. 2022, doi: 10.1109/TAFFC.2019.2949559.

[6] N. L. Yee, M. A. Zulkifley, A. H. Saputro, and S. R. Abdani, "Apex Frame Spotting Using Attention Networks for Micro-Expression Recognition System," *Computers, Materials and Continua*, vol. 73, no. 3, pp. 5331–5348, 2022, doi: 10.32604/cmc.2022.028801.

[7] K. S. Widjaja, C. C. Alamo, Anderies, and A. Chowanda, "Exploring the Accuracy of Artificial Intelligence in Detecting Lies Through Micro-expression Analysis," in *2023 6th International Conference on Information and Communications Technology (ICOIACT)*, Nov. 2023, pp. 218–223. doi: 10.1109/ICOIACT59844.2023.10455941.

[8] H.-X. Xie, L. Lo, H.-H. Shuai, and W.-H. Cheng, "An Overview of Facial Micro-Expression Analysis: Data, Methodology and Challenge," *IEEE Transactions on Affective Computing*, vol. 14, no. 3, pp. 1857–1875, Jul. 2023, doi: 10.1109/TAFFC.2022.3143100.

[9] F. Zhang and L. Chai, "A review of research on micro-expression recognition algorithms based on deep learning," *Neural Comput & Applic*, vol. 36, no. 29, pp. 17787–17828, Oct. 2024, doi: 10.1007/s00521-024-10262-7.

[10] J. He, L. Peng, B. Sun, L. Yu, and M. Guo, "Dual Multi-Task Network with Bridge-Temporal-Attention for Student Emotion Recognition via Classroom Video," in *2021 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2021, pp. 1–8. doi: 10.1109/IJCNN52387.2021.9533471.

[11] H. Zhang, L. Yin, and H. Zhang, "A review of micro-expression spotting: methods and challenges," *Multimedia Systems*, vol. 29, no. 4, pp. 1897–1915, Aug. 2023, doi: 10.1007/s00530-023-01076-z.
[12] G. Zhao, X. Li, Y. Li, and M. Pietikäinen, "Facial Micro-Expressions: An Overview," *Proceedings of the IEEE*, vol. 111, no. 10, pp. 1215–1235, Oct. 2023, doi: 10.1109/JPROC.2023.3275192.
[13] P. Choirina, U. D. Rosiani, and I. M. Fitriani, "Pelacakan Titik Fitur Dengan Kanade-Lucas-Tomasi (KLT) Untuk Pengenalan Ekspresi Mikro Wajah Pada Dataset SAMM," *G-Tech: Jurnal Teknologi Terapan*, vol. 8, no. 1, pp. 330–339, 2024, doi: 10.33379/gtech.v8i1.3755.
[14] R. Li *et al.*, "Micro-Expression Recognition by Motion Feature Extraction based on Pre-training," Jul. 10, 2024, *arXiv*: arXiv:2407.07345. doi: 10.48550/arXiv.2407.07345.
[15] P. Ekman and W. V. Friesen, "Nonverbal Leakage and Clues to Deception†," *Psychiatry*, vol. 32, no. 1, Art. no. 1, Feb. 1969, doi: 10.1080/00332747.1969.11023575.
[16] S.-T. Liong, J. See, K. Wong, and R. C.-W. Phan, "Less is more: Micro-expression recognition from video using apex frame," *Signal Processing: Image Communication*, vol. 62, pp. 82–92, Mar. 2018, doi: 10.1016/j.image.2017.11.006.
[17] E. Friesen and P. Ekman, "Facial action coding system: a technique for the measurement of facial movement," *Palo Alto*, vol. 3, no. 2, p. 5, 1978, doi: https://doi.org/10.1037/t27734-000.
[18] I. V. Vitale, "Facial Action Coding System Applied to Criminal Investigations: The Analysis of a Homicide Case in Southern Italy," *International Annals of Criminology*, vol. 59, no. 1, pp. 11–22, May 2021, doi: 10.1017/cri.2021.7.
[19] K. Lekdioui, Y. Ruichek, R. Messoussi, Y. Chaabi, and R. Touahni, "Facial expression recognition using face-regions," in *2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, May 2017, pp. 1–6. doi: 10.1109/ATSIP.2017.8075517.
[20] U. D. Rosiani, P. Choirina, S. Sumpeno, and M. H. Purnomo, "Menuju Pengenalan Ekspresi Mikro: Pendeteksian Komponen Wajah Menggunakan Discriminative Response Map Fitting," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, vol. 7, no. 2, Art. no. 2, 2018, doi: 10.22146/jnteti.v7i2.424.
[21] R. A. Pahlevi and B. Setiaji, "Analysis of Application Haar Cascade Classifier and Local Binary Pattern Histogram Algorithm in Recognizing Faces With Real-Time Grayscale Images Using Opencv," *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 1, pp. 179–186, 2023, doi: https://doi.org/10.52436/1.jutif.2023.4.1.491.
[22] H. Rahman, M. U. Ahmed, and S. Begum, "Non-Contact Physiological Parameters Extraction Using Facial Video Considering Illumination, Motion, Movement and Vibration," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 1, pp. 88–98, Jan. 2020, doi: 10.1109/TBME.2019.2908349.
[23] K. Lee, D. K. Han, and H. Ko, "Video Analytic Based Health Monitoring for Driver in Moving Vehicle by Extracting Effective Heart Rate Inducing Features," *Journal of Advanced Transportation*, vol. 2018, no. 1, p. 8513487, 2018, doi: 10.1155/2018/8513487.
[24] P. Choirina, U. D. Rosiani, and I. M. Fitriani, "Pengenalan Ekspresi Mikro Wajah Berdasarkan Point Feature Tracking Menggunakan Fase Apex Pada Database Ekspresi Mikro," *Edu Komputika Journal*, vol. 9, no. 1, Art. no. 1, Jun. 2022, doi: 10.15294/edukomputika.v9i1.56600.
[25] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Robust Discriminative Response Map Fitting with Constrained Local Models," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2013, pp. 3444–3451. doi: 10.1109/CVPR.2013.442.
[26] R. Miao, J. Qian, Y. Song, R. Ying, and P. Liu, "UniVIO: Unified Direct and Feature-Based Underwater Stereo Visual-Inertial Odometry," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–14, 2022, doi: 10.1109/TIM.2021.3136259.
[27] R. A. Asmara *et al.*, "Study of DRMF and ASM facial landmark point for micro expression recognition using KLT tracking point feature," *J. Phys.: Conf. Ser.*, vol. 1402, p. 077039, Dec. 2019, doi: 10.1088/1742-6596/1402/7/077039.

[28] C. He, "Face tracking using Kanade-Lucas-Tomasi algorithm," in *Fifth International Conference on Computer Information Science and Artificial Intelligence (CISAI 2022)*, Y. Zhong, Ed., SPIE, 2023, p. 125663P. doi: 10.1117/12.2667789.
[29] J.-C. Piao and S.-D. Kim, "Adaptive Monocular Visual–Inertial SLAM for Real-Time Augmented Reality Applications in Mobile Devices," *Sensors*, vol. 17, no. 11, Art. no. 11, Nov. 2017, doi: 10.3390/s17112567.
[30] J. Devasagayam, R. Bosma, and C. M. Collier, "A velocity program using the Kanade–Lucas–Tomasi feature-tracking algorithm with demonstration for pressure and electroosmosis conditions," *ELECTROPHORESIS*, vol. 43, no. 7–8, pp. 865–878, 2022, doi: 10.1002/elps.202100177.
[31] Z. Jun, "The Development and Application of Support Vector Machine," *J. Phys.: Conf. Ser.*, vol. 1748, no. 5, p. 052006, Jan. 2021, doi: 10.1088/1742-6596/1748/5/052006.
[32] J. Zhang, C. Li, Y. Yin, J. Zhang, and M. Grzegorzek, "Applications of artificial neural networks in microorganism image analysis: a comprehensive review from conventional multilayer perceptron to popular convolutional neural network and potential visual transformer," *Artif Intell Rev*, vol. 56, no. 2, pp. 1013–1070, Feb. 2023, doi: 10.1007/s10462-022-10192-7.
[33] G. Garg, D. Kumar, ArvinderPal, Y. Sonker, and R. Garg, "A Hybrid MLP-SVM Model for Classification using Spatial-Spectral Features on Hyper-Spectral Images," Jan. 01, 2021, *arXiv*: arXiv:2101.00214. doi: 10.48550/arXiv.2101.00214.
[34] A. K. Davison, C. Lansley, N. Costen, K. Tan, and M. H. Yap, "SAMM: A Spontaneous Micro-Facial Movement Dataset," *IEEE Transactions on Affective Computing*, vol. 9, no. 1, pp. 116–129, Jan. 2018, doi: 10.1109/TAFFC.2016.2573832.

Jurnal Teknik Informatika (JUTIF) Vol. 6, No. 2, April 2025, Page. 593-608
P-ISSN: 2723-3863 https://jutif.if.unsoed.ac.id
E-ISSN: 2723-3871 DOI: https://doi.org/10.52436/1.jutif.2025.6.2.4235