



A Novel Approach on Motion Estimation for Micro-Expression Recognition Using Phase Only Correlation with All Block Search (POC-ABS)

Ulla Delfana Rosiani^{1,2,4*} **Priska Choirina⁴** **Niyalatul Muna⁵** **Eko Mulyanto^{1,2}**
Surya Sumpeno^{1,2,3} **Mauridhi Hery Purnomo^{1,2,3*}**

¹*Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Indonesia*

²*Department of Computer Engineering, Institut Teknologi Sepuluh Nopember, Indonesia*

³*The Science and Technology Center of Artificial Intelligence for Healthcare and Society (PUI AIHeS), Indonesia*

⁴*Department of Technology Information, Politeknik Negeri Malang, Indonesia*

⁵*Department of Health, Politeknik Negeri Jember, Indonesia*

* Corresponding author's Email: hery@ee.its.ac.id, ulla.rosiani14@mhs.ee.its.ac.id

Abstract: Micro-expression is an expression when a person tries to hold or hidden, but the leak of this emotion still occurs in one or two areas of the face or maybe a short expression that across in the whole-face. Not more than 500ms, micro-expressions can be difficult to recognize and detect where the leakage area is located. This study presents a new method to recognize and detect the subtle motion on the facial components area using Phase Only Correlation algorithm with All Block Search (POC-ABS) to estimate the motion of all block areas. This block matching method is proposed by comparing each block in the two frames to determine whether there is movement or not. If the two blocks are identical, then the motion vector value is not displayed, whereas if the blocks are non-identical, the motion vector value of the POC is displayed. The motion vector, which is as a motion feature, estimates whether or not there are movements in the same block. In order to further confirm the reliability of the proposed method, two different classifiers were used for the micro-expression recognition of the CASME II dataset. The highest performance results are for SVM at 94.3 percent and for KNN at 95.6 percent. Finally, this algorithm detects leaks of motion based on the ratio of the motion vectors. The left and right eyebrows are dominant when expressing disgust, sadness, and surprise. Meanwhile, the movements of the right eye and left eye were the most dominant when the happiness expression.

Keywords: Micro-expression, Phase-only correlation, Facial component, Motion estimation, Leakage.

1. Introduction

Micro-expression occurs when someone wants to hide feelings from oneself (repression) or others [1], because may not be appropriate or may be due to cultural display rules [2]. But there is an emotional leak with low muscle activity in certain facial areas with a duration of less than 500 ms [3]. It is not easy to detect these emotional leaks, high-speed cameras and high-quality automatic systems are needed [4, 5].

Observation of subtle and fast motion is needed in the automatic detection system of micro-expression [6]. Various video analysis applications such as detection of moving objects [7], human motion detection [8], hand gesture analysis [9],

cinematography [10], etc. use a motion estimation approach to detect and analyze motion [11].

Optical flow-based is the method most widely used in the recognition of micro-expressions with a motion-based approach [12, 13]. Optical flow and optical strain were first proposed to detect fine movements in faces [14]. This method measures the optical flow-motion and strain derived and cannot be used when the translation motion of head is large and fast [14, 15]. Other studies based on optical flow include: Liong et al. using features of optical strain, optical strain weight, and a combination of the two also reported spontaneous emotions [16]. Xu et al. [17] have implemented a dynamic face map (FDM) that estimates the optical flow on the motion of the facial components. This technique can be used to

avoid the aggravation of head pose variations. A maximum directional difference (MDMD) analysis was performed by Wang et al. [18], achieving higher spotting precision in long images. Happy and Routray [19] proposed a temporal feature descriptor named Fuzzy Histogram of Optical Flow Orientation (FHOFO), which is a refinement of the Histogram of Oriented Optical Flow (HOOF), with the Pair-Wise (PWFP) function to improve previous research results. Meanwhile, Li et al. [19] conducted a deep learning method based approach using HOOF to detect the direction of movement of facial muscles, a multi-task learning to find facial features and region-based normalized HOOF features. This is the only technique to date that uses advanced ME spotting approaches. Although the optical flow strategy can manage an over and fast head translation movement situation, the classification accuracy is not high [12]. In addition, computer-intensive and high computational time requirements are needed, so this approach is not sufficient for real-time and practical implementation [20].

Another faster and accurate approach is the block-based motion estimation or Block Matching Algorithm (BMA) [20]. Combining BMA method that uses Sum Absolute Difference (SAD) technique and simplified Optical Flow approach, Taylor Series Approximation can better classify emotions on micro-expression with the best accuracy was 85.07% [6]. There are some popular methods of matching criteria or distortion functions for BMA: SAD, Mean of Absolute Difference (MAD), Mean Squared Error (MSE) and Phase Only Correlation (POC). POC is an image matching method with high accuracy [21–24] and better than conventional full search using SAD [25]. The fact that the displacement between the two images that are in the phase of the cross-spectrum power is the basis of the algorithm POC [26] and it can be used to detect translation movements that occur in images [25]. The full-search POC algorithm (POC-FS) examines all positions in the search area in detail. This algorithm is optimal in the sense that the best matching position is guaranteed if the search range is correctly defined [20]. However, in order to find a motion vector for each block, a large number of computations are required [27].

To reduce complexity, a new block matching algorithm named POC-ABS is proposed. In this study, a block matching method is proposed by comparing each block in the two matching frames. If the two blocks are identical, then the motion vector value is not displayed, whereas if the blocks are non-identical, the motion vector value of the POC is displayed. This method does not look for the most accurate matching

results, but estimates whether or not there are movements in the same block. The observations are particularly prioritized on the eyes area and additionally the mouth area. Region-based research better results than global-based research (full-face observation) [28]. The motion features obtained from all areas of the facial components are used for the micro-expression classification.

The following is the outline of this paper: Section 2, Phase Only Correlation algorithm with All Block Search (POC-ABS) in brief. In section 3, the proposed micro-expression recognition based on POC-ABS are discussed. Section 4 explains the experiment setup and results, and Section 5 contains the discussion for the research results.

2. Phase only correlation algorithm with all block search (POC-ABS)

As the most common motion estimation method, BMA has proved highly efficient in terms of quality and bit rate. Each frame is divided into the same size macroblock and not overlap. A single vector of motion is applied to the entire object, it assuming undergoes translational motion. The direction and magnitude of the motion vector are the main objectives to be achieved from the block matching of the current framework to the candidate blocks that are most suitable in the reference frame. There are several BMA parameters that affect performance and accuracy in the estimation of motion: distortion block function, block size, and maximum movement displacement allowed, also known as search range [20].

The distortion block function or matching criterion is used to measure similarities between two-block, the current and reference block. There are several popular matching functions and compared to others, Phase Only Correlation (POC) is better [25]. The POC algorithm is also one of the most accurate image registration techniques [21] which is widely used in the field of computer vision and remote sensing, especially for applications such as super-resolution imaging [10], 2D and 3D image registration [29], and displacement estimation images with subpixel accuracy [23].

The POC algorithm is used to measure of similarity between two frame: current frame and reference frame. Given 2 frames of size $N_1 \times N_2$, being $f(n_1, n_2)$ and $g(n_1, n_2)$ as shown in Fig. 1 (a) and Fig. 1 (b), for simplification, it is assumed that $n_1 = -M_1 \dots M_1$ and $n_2 = -M_2 \dots M_2$ because $N_1 = 2M_1 + 1$ and $N_2 = 2M_2 + 1$. The Discrete Fourier Transformation form of the two frames are

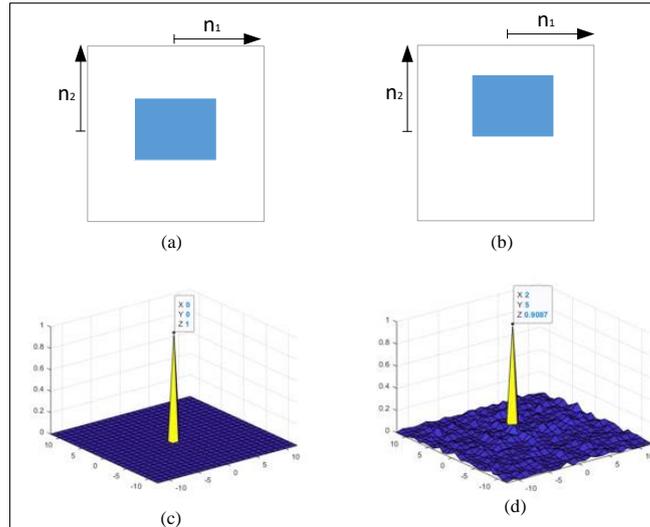


Figure. 1 POC matching example: (a) frame $f(n_1, n_2)$, (b) translation frame $g(n_1, n_2)$, (c) POC function between two identical frames between $f(n_1, n_2)$ and $f(n_1, n_2)$, and (d) POC function between two not identical frame $f(n_1, n_2)$ and $g(n_1, n_2)$

given by

$$F(k_1, k_2) = \sum_{n_1, n_2} f(n_1, n_2) W_{N_1}^{k_1 n_1} W_{N_2}^{k_2 n_2} = A_F(k_1, k_2) e^{j\theta_{F(k_1, k_2)}} \quad (1)$$

$$G(k_1, k_2) = \sum_{n_1, n_2} g(n_1, n_2) W_{N_1}^{k_1 n_1} W_{N_2}^{k_2 n_2} = A_G(k_1, k_2) e^{j\theta_{G(k_1, k_2)}} \quad (2)$$

The cross-spectrum of $\hat{R}(k_1, k_2)$ between $F(k_1, k_2)$ dan $G(k_1, k_2)$ is indicated with the formula:

$$\hat{R}(k_1, k_2) = \frac{F(k_1, k_2) \overline{G(k_1, k_2)}}{|F(k_1, k_2) \overline{G(k_1, k_2)}|} = e^{j\theta(k_1, k_2)} \quad (3)$$

where $\overline{G(k_1, k_2)}$ states the complex conjugation of $G(k_1, k_2)$ and $\theta(k_1, k_2) = \theta_F(k_1, k_2) - \theta_G(k_1, k_2)$.

The POC function $\hat{r}(n_1, n_2)$ is the inverse 2D discrete Fourier transformation of $\hat{R}(k_1, k_2)$ and is formulated by Eq. (4):

$$\hat{r}(n_1, n_2) = \frac{1}{N_1 N_2} \sum_{k_1, k_2} \hat{R}(k_1, k_2) W_{N_1}^{-k_1 n_1} W_{N_2}^{-k_2 n_2} \quad (4)$$

with \sum_{k_1, k_2} defining $\sum_{k_1=-M_1}^{M_1} \sum_{k_2=-M_2}^{M_2}$.

When the two frames are identical, i.e. $f(n_1, n_2) = g(n_1, n_2)$, the POC function will be given by:

$$\hat{r}(n_1, n_2) = \frac{1}{N_1 N_2} \sum_{k_1, k_2} W_{N_1}^{-k_1 n_1} W_{N_2}^{-k_2 n_2} = \delta(n_1, n_2) = \begin{cases} 1 & \text{if } n_1 = n_1 = 0 \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

If the two image blocks are identical, the POC value will equal to 1 and produce a distinct sharp peak, as shown in Fig. 1 (c). Conversely, if it is non-identical, then the peak will drop significantly. The peak height is used as a measure of similarity for matching images, and the peak position indicates the translation movement between two images [25]. Therefore the translational motion and the fit value between the two frames can be obtained based on the position and height of the correlation peaks [22, 30] as shown in Fig. 1 (d).

POC Full Search (POC-FS) is a fully comprehensive search algorithm, and finding the optimal motion vectors (MV) for each macroblock is the simplest process. It determines the candidate block best suited by measuring the distortion function for all candidate blocks inside the search window. The search process is carried out by matching blocks in the search area. The matching value that has the highest POC value will be considered as the same block. MV consists of a pair (x, y) of horizontal and vertical displacement values pointing from the center of the current frame to the center of the most appropriate block [25]. Although POC-FS can produce optimal results globally, it has very intensive calculations and therefore very time-consuming [31]. To reduce the complexity, we propose a simple search strategy for a phase-based motion estimation method with modified the search range of blocks matching process, called POC-All Block Search (POC-ABS).

Algorithm 1 Pseudocode for POC-ABS	
Input	- Bs= Block Size; - img = image sequence
Output	- Highest Value POC(Hpoc) - Motion Vector (Mv)
<pre> 1. Begin 2. for f = 1: (sum of image sequence -1) do 3. CurrentFrame (CF) ← img(f) 4. ReferenceFrame (RF) ← img(f+1) or img(f-1) 5. width (w), height (h) ← Current frame size 6. blockX ← ((w/Bs) * Bs), division of the block index on X(row) 7. blockY ← ((h/Bs) * Bs), division of the block index on Y(column) 8. for i ← 1: Bs : blockY do 9. for j ← 1: Bs : blockX do, repeat every block of CF 10. block ← [j until (j+B_s-1); i until (i+B_s-1)] 11. imgCur ← CF(block) 12. imgRef ← RF(block) 13. xCenter ← (blockCur[1]/ 2), determines the midpoint of the X coordinat on imgCur 14. yCenter ← (blockCur[2]/2), specify the midpoint of the Y coordinates of imgCur 15. F ← Two-dimensional Fourier transform on imgCur 16. G ← Two-dimensional Fourier transform on imgRef 17. R ← calculates the cross-phase spectrum of F and G 18. highR ← max(R), find the largest value of the matrix R 19. xMax,yMax ← find(highR == R), find the coordinate position of the highR value in the R matrix 20. tempR(j) ← [highR, xMax, yMax], temporary of data storage for each RF block looping 21. MaxCoor ← find(tempR[:,1] == max(tempR[:,1])), find the largest value and index of tempR 22. Hpoc ← tempR[MaxCoor,1] 23. p1 ← [xCenter, yCenter], midpoint of blockCur 24. p2 ← [tempR[MaxCoor, 2], tempR[MaxCoor, 3]], maximum midpoint of blockRef 25. dp ← p2 - p1 26. quiver(p1[1],p1[2],dp[1],dp[2]), Motion Vector image with starting point p1 and end point on dp 27. end for 28. end for 29. end for 30. end begin </pre>	

The POC-FS search area consists of several blocks, in this proposal the search area is only on one block. Our proposed system's pseudocode is given in Algorithm 1. Search strategy by comparing each block in the current frame directly to each block in the reference frame, where both blocks are in the same block position. If the two blocks are equal, the highest POC value of 1 will be generated at the center coordinates (0,0). However, if the two blocks are different, the highest POC value will less to 1 and its position will not in the center coordinate. The position of the highest value is the translation

movement between the two blocks at positions as far as x and y. The MV will then be generated from the center position of the block to the x, y points. It is assumed, the movement of all pixels in one block is considered the same.

3. Proposed micro-expression recognition based on POC-ABS

The micro-expression recognition research scheme on facial components, as shown in Fig. 2, uses feature extraction based on the POC-ABS matching algorithm to obtain motion features. The proposed system consists of three main stages: facial component detection and tracking, feature extraction using the POC-ABS matching algorithm, and micro-expression recognition.

Firstly, facial images are detected and divided into various areas of facial components using a series of facial feature points obtained from the DRMF alignment point method [32]. Facial movements are then tracked using the Kanade-Lucas-Tomasi (KLT) tracking method[33] then the position of the marked area of the face component is always traceable (Section 3.1), as shown in Fig. 2 (a). Next, in the feature extraction step, the motion features are obtained from the POC-ABS matching results (Section 3.2), as shown in Fig. 2 (b). For the final stage, the classification of micro-expressions using the SVM and KNN will be presented in Sections 3.3 and Fig. 2 (c). The performance of these two classifications is compared for all match block sizes.

3.1 Facial component detection and tracking

This section is the pre-processing stage for recognition of micro-expressions. Micro-expression movements consist of low muscle activity occurring in certain areas of the face and in a short time, and therefore a pre-processing stage that is accurate and stable is required. Observations made explicitly in the region of the eye and mouth throughout this report. By separating it into four parts, the area of the eye can be more detailed: right eyebrow, left eyebrow, right eye and left eye.

This study uses a sequence of images from a micro-expression video dataset. Region of Interest (ROI) of faces and facial components detection using Viola-Jones method and DRMF [32] is performed on the first frame [34]. To ensure accuracy and precision, each frame movement must have precise markings. In this analysis, facial-point tracking used the feature tracking algorithm Kanade-Lucas-Tomasi (KLT) feature tracking algorithm[33]. In each frame motion, tracking is done at forty-nine points. Then the marker

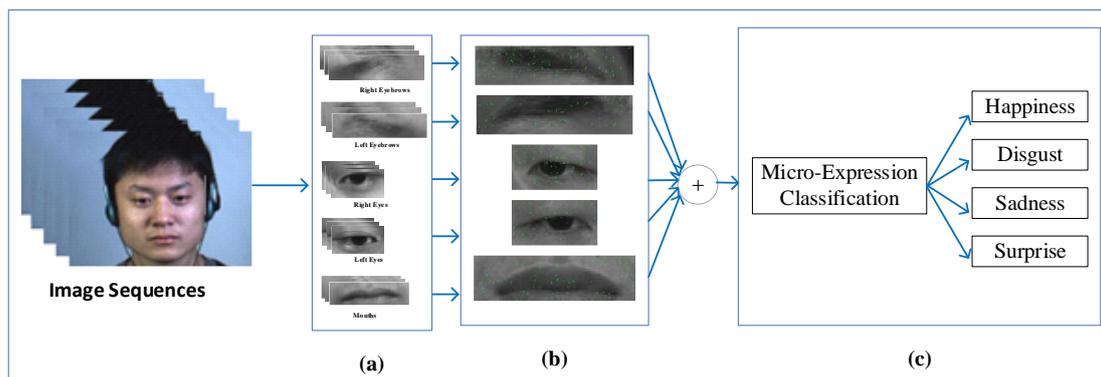


Figure. 2 Micro-expression recognition research based on POC-ABS feature extraction: (a) facial component detection and tracking, (b) feature extraction based on POC-ABS, and (c) micro-expression classification

box is always formed precisely and accurately for each facial component [34].

3.2 Feature extraction based on POC-ABS

After the areas of the facial component are formed, the next step is to get the features of motion in each area. The motion feature is obtained by matching blocks of two frames, frame current and frame reference, using the matching algorithm POC-ABS. In this section, there are three step: the creation of comparison blocks, matching between blocks using the POC-ABS method and the extraction of motion features from each motion vector, as shown in Fig. 3.

The first step is to divide the face component area into blocks of the same pixel size. In this study block sizes range from 5x5 pixels to 25x25 pixels. Each block size has a different number of blocks, the larger the block size, the smaller the number of blocks formed, as shown in Table 1 and Fig. 3 (a).

The second step is to match the blocks between the two frames using the POC-ABS method. The frame being compared is a neutral face frame with a frame when micro-expression occurs. Match each block, between the two frames, in the same block order. If the two blocks are identical or not shifted, then the highest POC value is 1 and there is no motion vector arrow. However, if there is a difference or movement, then the highest POC value is less than 1 and a motion vector will be generated. As explained in section 2, when the two blocks are equal, the highest POC is 1 (or close to 1) in the center of coordinates (0,0) then there is no motion vector. But when two blocks move in translation, the highest POC is not 1 and is in coordinates (x, y). This will produce a motion vector from the center point of the block to the coordinates (x, y), as shown in Fig. 3 (b). For the example above, as shown in Fig. 3 (b), the matching stage between the two frames in the mouth area. The blocks that are compared between the two

frames are in the same order position. A comparison of block number 28 indicates movement. The highest POC value (value 0.4687) is at position $x=2$ and $y=2$ from the center coordinate position (0,0), so that the motion vector arrow is directed to the top right.

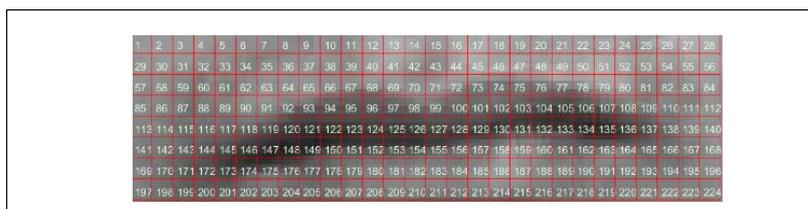
In the third step, after a comparison between all blocks, many motion vectors will be generated, as shown in Fig. 3 (c). The motion vector value of each block will be used as a motion feature for each area of the face component. The value of this motion vector is the horizontal (XBx) and vertical (YBx) components, the magnitude of the vector (MBx), and the direction of the vector produced based on its angle (TBx). As shown in Fig. 3 (c), the total number of blocks in the mouth area for the 5x5 pixel block size is 224 blocks. Each block has 4 features. In blocks numbers 1 and 224, there are no motion vectors or not, all features are 0. While in block number 28, there is a motion vectors, then the value of each feature is $XB_{28} = 2$, $YB_{28} = 2$, $MB_{28} = 0.46$ and $TB_{28} = 45$. It can be seen that not all matching blocks have moves in translation. The vector of motion is the product of a block move. So that the extraction of the function performed at this point uses only motion vectors on these blocks, while the matching results from blocks which have no motion are not used.

3.3 Micro-expression classification

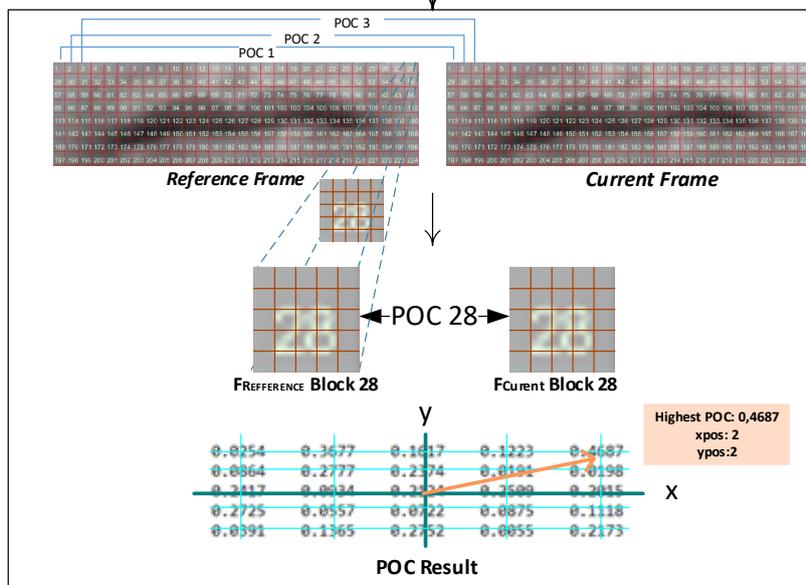
The final step of this system is to classify the emotions in micro-expressions. Four micro-expression classes were used for all experiments (disgust, happiness, surprise, and sadness) via a multi-classification approach. The motion feature from the POC-ABS feature extraction method is recognized using Supports Vector Machine (SVM)[[35] with Radial Base Function (RBF) [6] and K-Nearest Neighbor (KNN) with Euclidean distance metric[36]. In this study, ten-fold cross-validation was used to train and test all samples of each class. Each sample consists of the motion

Table 1. Facial component area and block size (pixel)

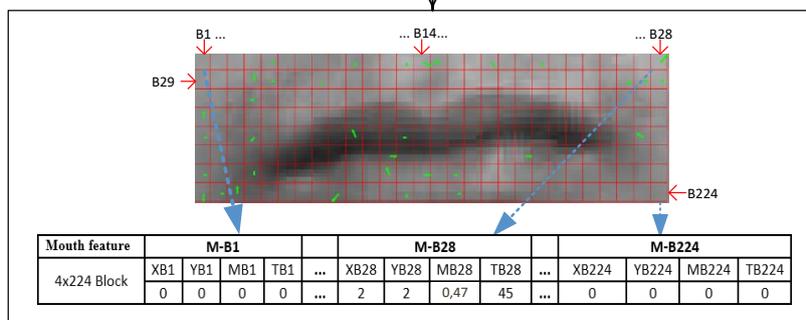
Facial Component	Facial Component Area Size	Block Size										
		5x5	7x7	9x9	11x11	13x13	15x15	17x17	19x19	21x21	23x23	25x25
Left eyebrow	28x111	110	64	36	20	16	7	6	5	5	4	4
Right eyebrow	28x111	110	64	36	20	16	7	6	5	5	4	4
Left eye	43x81	128	66	36	28	18	10	8	8	6	3	3
Right eye	43x81	128	66	36	28	18	10	8	8	6	3	3
Mouth	39x139	224	100	60	36	30	8	16	14	6	6	5
Sum of Blocks		700	360	204	132	98	42	44	40	28	20	19



(a)



(b)



(c)

Figure. 3 Feature extraction based on POC-ABS: (a) dividing equal blocks sized, (b) block matching using the POC-ABS algorithm, and (c) motion feature extraction

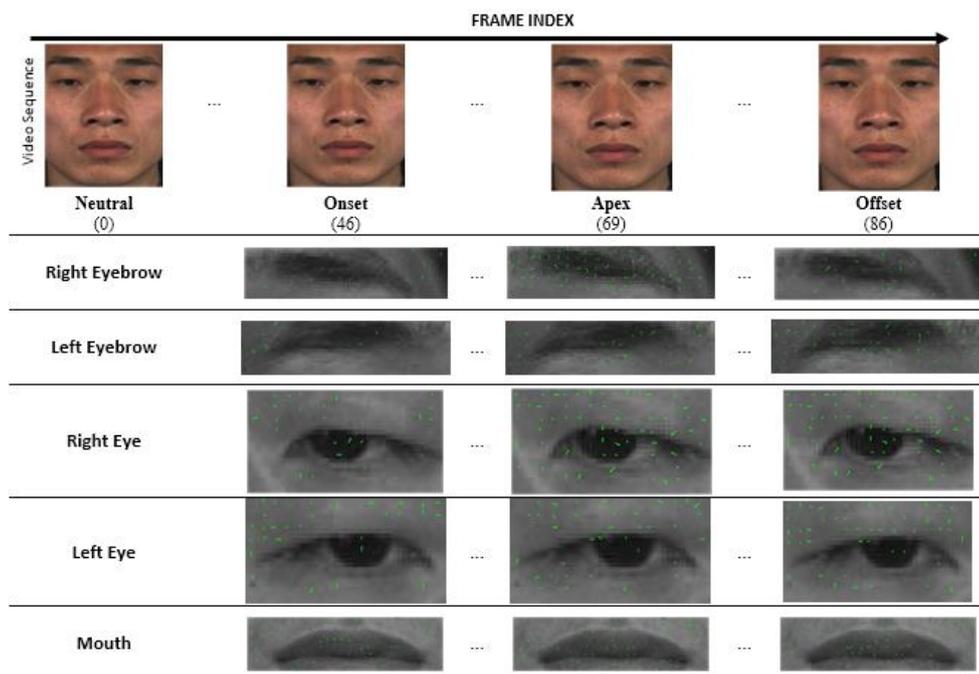


Figure. 4 Subtle motion estimation on facial component areas using POC-ABS

feature of all facial components as shown in Fig. 2 (b) and Fig. 3 (c).

In past micro-expression studies such as [16, 37-39], the selection of SVM and KNN as baseline model was motivated by its success. SVM uses a hyperplane to separate data groups into their respective classes. More than one hyperplane could separate the classes, and the one with the largest margin is chosen as the best / most correctly classified. The RBF kernel is a kernel function that is commonly used for analysis when data is not linearly separated.

KNN is a classification method based on the distance of the Nearest Neighbor. The concept of the KNN is to find the value of k training data which is the closest distance to new data whose label is unknown (testing data). In this study, to obtain the amount of this distance, it is calculated by applying the euclidean distance. Detail experiments and accuracy of classification are presented in the next sections.

4. Experiment setup and result

To recognize micro-expressions faster and accurately, and to identify where there is expression leakage using the proposed method, POC-ABS method, several experiments are required to get the right results. The experiment began with the detection of micro-expression subtle motion in areas of facial components, from the onset, apex to offset stages. Next, to determine the computation time, a time measurement was performed when the POC-ABS

extracted the subtle micro-expression motion features. To recognize micro-expressions, we use POC-ABS motion features with four emotion classes which are classified using two classification engines. Comparison of the accuracy of the two classification machines over several block sizes will provide the recommended block size and classification engine for the most accurate recognition of micro-expressions. To detect the expression leakage, a motion vector ratio analysis of each area of the facial component is used. It is assumed that the leakage area of each micro-expression is the highest movement ratio in the region of the facial components.

4.1 Micro-expression dataset

Research micro-expression requires a dataset of high quality. The CASME II (Chinese Academy of Sciences. Micro-expression II) dataset [40] is one of the most commonly used datasets. Expressions in the CASME II dataset were collected from 26 people (from one ethnicity), in the form of video recordings with 200 fps quality. There are many expression labels on the set, with uneven distribution. The sample in CASME II consisted of six basic emotions, namely happiness, disgust, oppression, surprise, fear, and sadness. This study uses a series of images ranging from onset to offset status with four emotional labels (happy, disgusted, surprised, and sad).

4.2 Detection of subtle motions in micro-expression

In this section, we will discuss the detection of subtle micro-expression motion using proposed method, POC-ABS method. Experiments were carried out on areas of facial components (right eyebrow, left eyebrow, right eye, left eye, and mouth) by comparing the frame during neutral conditions with other frames during micro-expression movements, from frame onset, apex, and offset. The onset and offset states are when the states shift after the neutral states (onset) and the neutral states meet again (offset). Meanwhile, the Apex state is the expression peak between onset and offset.

The matching process between the neutral frame (reference frame) with the micro-expression frames (current frame) will result a MV arrows that indicate a translational movement between both frames. The results of motion estimation detection in the five areas of the face components are shown in Fig. 4. The face images shows the sequence of image from the neutral state, initial state, peak state, to the offset state. The number below the caption of each face image is an index of the position of each image. The next part is an MV arrow image, with a green line, which shows the translational motion in each area of the facial component. If the two blocks are identical, no MV arrow is generated. If the two blocks are different, there will be an MV arrow. In the Apex state, the number of MV arrows is higher than in other states, this shows that there are many translation movements that occur at the peak of the expression.

4.3 Computational time

In this section we examine the computational time of the POC-ABS and POC-FS methods in the CASME II dataset. Block matching is performed on 5x5 pixel blocks. The hardware used is an Intel core i76700T, 4GB RAM, 2.8 GHz CPU, 64-bit windows.

The average duration per video and per-frame for block matching process execution, shown in Table 2. The fastest time are shown in bold and italics. The all frame series time is the processing time from onset state until offset time. The per frame time is the processing time when matching each face component

Tabel 2. Average computational time

Method	Average Duration per Video	
	All Frame Series (s)	Per Frame (s)
POC-FS	9,444	0,236
POC-ABS	<i>1,448</i>	<i>0,036</i>

area between reference frame and current frame. The results show that the processing time of the POC-ABS method is faster than the POC-FS method. It cannot be denied that extracting features using POC-ABS is much faster than compared using POC-FS.

4.4 Comparison of micro-expressions classification accuracy in various block sizes

The performance of micro-expression recognition for each block size is discussed in this section. The POC-ABS matching motion feature is extracted from five areas of the face component (right eyebrow, left eyebrow, right eye, left eye, and mouth). Features from all areas will be combined and become a series of features for each matching frame. In this experiment, the POC-ABS Block size that will be used starts from 5x5 pixels to 25x25 pixels. Furthermore, it will compare the accuracy of micro-expression recognition for each block size using two classification machines (SVM and KNN).

As shown in Table 3, this accuracy measurement is used to evaluate the block size and performance of each classifier. The 10-fold cross-validation method is used to determine the accuracy of recognition of each classifier. The highest accuracy values are shown in bold and italics. The highest accuracy value is 95.6 percent of KNN and 94.3 percent of SVM, both of which are 5x5 pixels in the same block size. Although the lowest accuracy value for SVM is 79.8 percent and 52 percent for KNN. The SVM classification's accuracy value is quite stable, although it tends to fall, which is 87.8 percent. Whereas the KNN classification with an average accuracy of just 69.65 percent is not too good, it is below SVM's average accuracy. It can be seen that the 5x5 block changes from the highest value, then it drops to 52 percent in the size of the 9x9 block, and returns up to 73 percent in the size of the 25x25 block.

4.5 Leakage motion detection of facial components

Upon achieving the highest accuracy of classification resulting in a block size of 5x5 pixels, the next step is to determine where the leak in the expression is concealed while the micro-expression is present. This leak is a movement on the face that cannot be captured and occurs spontaneously. This movement occurs quickly and only in limited facial areas[3], [41]. Matching results between the two frames, determined using the same POC-ABS process, are used to obtain motion estimation. The MV arrows suggest which motions are occurring. Motion estimation in each region of the facial

Table 3. Comparison of recognition accuracy in various block sizes (pixel) using SVM and KNN classification (%)

Classifier	Block Size										
	5x5	7x7	9x9	11x11	13x13	15x15	17x17	21x21	19x19	23x23	25x25
KNN	<i>0,956</i>	0,792	0,521	0,57	0,602	0,625	0,708	0,727	0,728	0,702	0,73
SVM	<i>0,943</i>	<i>0,935</i>	<i>0,921</i>	<i>0,901</i>	<i>0,884</i>	<i>0,873</i>	<i>0,863</i>	<i>0,853</i>	<i>0,853</i>	<i>0,798</i>	<i>0,835</i>

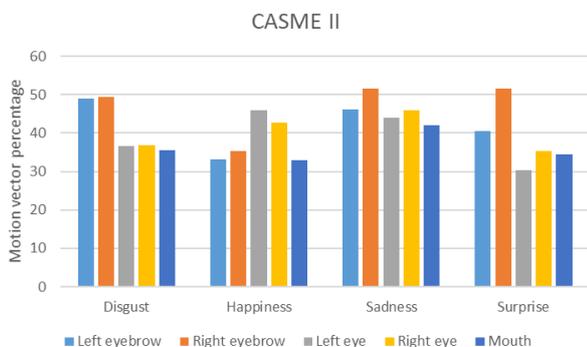


Figure. 5 The motion vector ratio of facial components for all expressions

component area are used to evaluate leakage in micro-expression.

In this section we will analyze the number of arrows generated from each match between frames. Each face component area has a different block number, as shown in table 1. In order to display the movement tendencies in each area of the facial part, it is important to measure the ratio of the number of motion vectors to the total number of blocks in each block, as shown in Eq. (6).

$$P_{vector} = \frac{\text{Sum of vector arrow}}{\text{Sum of block}} \tag{6}$$

The ratio of MV for each expression is shown in Table 4. The average number of arrows on the right eyebrow is 57 arrows to the sadness, then the percentage of vectors is 52 percent of a total of 110 blocks (as shown in Table 1). The right eye area, the average number of arrows is 59, which represents 46 percent of all 128 blocks. Similarly, 46 percent of left eyebrows, 44 percent left eye, and 42 percent left mouth. For each division of lines, the two highest percentage values are marked with italics and bold as the sign with the most dominant motion, respectively on the right eyebrow and right eye.

Based on the graph in Fig. 5, when the expression is disgust and surprise, there are many movements in both the left and right eyebrows. The motion dominant at right eye and left eye when happiness. While the sad expression is dominated by the movement at the right eyebrow, left eyebrow, and right eye. Thus, the location of the leak is found in the eye region from these four expressions (right eyebrow, left eyebrow, right eye, and left eye).

5. Discussion

Our experiment is done using the motion feature obtained from the motion estimation method proposed, POC-ABS method. This method can produce more detailed observations of movements in the facial component areas (right eyebrows, left eyebrows, right eye, left eye, and mouth). This method can properly detect subtle micro-expression motion from onset to offset state. At the onset state, it was seen that some movement was detected, there were several MV arrows, as shown in Fig. 4. The number of MV arrows then increased substantially in all areas of facial components in the Apex state, the peak state of expression at the index 69th frame. However, when the state was offset, at the 86th index frame, the number of MV arrows decreased, especially in the right eyebrow and mouth area. This method can detect existing movements well.

Likewise in no movement block detection. This method can detect an average of more than 50% of blocks with no MV or is assumed to be identical or no shift, as shown in Table 4. For each expression, the MV arrow ratio for all areas of the facial components is almost all below 50%, this means that more than half of the area of the facial components does not move. This shows that the POC-ABS method can detect subtle movements during micro-expressions just in certain areas of the face not in all areas properly [42].

For the computation time of the block matching process, the POC-ABS method is much faster than the POC-FS method, as shown in Table 2. The average POC-ABS computation time for both, the processing time of the entire series of frames and the processing time per frame is 6.5 times faster than the computation time of the POC-FS method. This is because the POC-FS method must look for the same block as the block being compared in a search area consisting of several blocks. Meanwhile, the way the POC-ABS method works is to compare two blocks, whether these two blocks are the same or not. If they are not the same, it is estimated that there is a shift and results in an arrow MV. So that the big time difference is because the way the POC-ABS method works is simpler than the POC-FS method.

Table 4. The motion vectors ratio for each expression in 5x5 (pixel) block size

Expression	Left eyebrow		Right eyebrow		Left eye		Right eye		Mouth	
	MV	Ratio (%)	MV	Ratio (%)	MV	Ratio (%)	MV	Ratio (%)	MV	Ratio (%)
DISGUST	54	49	54	49	47	37	47	37	79	35
HAPPINESS	36	33	39	35	59	46	55	43	74	33
SADNESS	51	46	57	52	56	44	59	46	94	42
SURPRISE	45	41	57	52	39	30	45	35	77	35

Table 5. Research of motion feature for micro-expression recognition

Year	Author	Dataset	Motion Feature	Classifier	Classification Accuracy
2016	Liong et al.[16]	CASME II	Optical Strain	SVM	63.41%
2017	Peng et al. [44]	CASMEII	OF	DTSCNN	Accuracy up to 66.67%
2017	Happy and Routray [19]	CASME II	FHOFO	SVM, KNN and LDA	F1-score was 0.5248
2017	Zhang et al. [38]	CASME II	LBP-TOP, Optical Flow	SVM, KNN	KNN: 54.51% SVM: 42.35%
2017	Muna et al. [6]	CASME II	SAD and OF	SVM	85,07%
2018	Zhu et al.[43]	CASME II	LBP-TOP and OF	SVM	53.3%
2018	Liong et al.[45]	CASME II	OSF and LBP-TOP	SVM	F-measure: 0.31
2018	Liong et al. [46]	CASME II	BI-WOOF	SVM	58.85%
2020	This study	CASME II	POC-ABS	SVM and KNN	SVM: 94,3% KNN: 95,6%

Table 6. Micro-expressions leakage areas

Author(s)	Dataset	Criteria	Micro-Expression Leakage Area			
			Happiness	Disgust	Sadness	Surprise
Yan, et al.[40]	CASME II	FACS	either AU6 or AU12 (Cheek or Lip)	one of AU9, AU10 or AU4+AU7 (Nose, Lips or Eyebrow)	-	AU1+2, AU25 or AU2 (Eyebrow or Lips)
Duan, et al.[28]	CASME II	Facial Region	eye region, mainly on eyebrow raised/down	eye region, mainly on eyebrow raised/down	eye region, mainly on eyebrow raised/down	eye region, mainly on eyebrow raised/down
This Study	CASME II	Facial Region	Left Eye and Right Eye	Left and Right Eyebrow	Left and Right Eyebrow	Left and Right Eyebrow

In the previous section, we have shown the results of the classification accuracy of micro-expression recognition. For both classifiers, the highest accuracy is at the smallest block size. The highest accuracy was 95.6 percent for KNN and 94.3 percent for SVM for inter-block matching of 5x5 pixel block sizes, as shown in Table 3. For both classifiers, the highest accuracy is at the smallest block size. The smaller the block size, the narrower the area being compared. In subtle micro-expression motion, the narrower this area is, the subtler motion is detected. As a result, more motion features are obtained. Meanwhile, the wider the area being compared, these subtle movements will be difficult to detect accurately. As a result, less motion features will be obtained. So that

the more varied the motion features, the higher the accuracy of the recognition of micro-expressions. In addition, when compared with other motion feature methods, this method can produce the highest accuracy value among other studies, as in Table 5.

In previous studies, Liang et al. [16] proposed two features: optical strain weighted and optical strain. The recognition rate reached to 63.16% for the CASME II dataset with the SVM classifier. Zhang et al. [38] combining local LBP-TOP and local Optical Flow (OF) features after extracting from the local area of face based on AUs and performed it on CASME II dataset. They claim that various local features can work better than a single global feature. Accuracy up to 62.50% has been achieved when they

combine two local features with RF classifier. To enhance micro-expression recognition, Zhu et al.[43] transferring learning from speech characteristics to micro-expression. LBP-TOP and OF are used as extractor features with different vector dimensions. The best accuracy was 53.3% achieved by OF with 50 dictionary dimensions with the SVM classifier. Based on previous research, the highest accuracy value was 85% using the combination of SAD+OF feature extraction method and SVM classification. In this study, the recognition accuracy was 94.3% for SVM and 95.6% for KNN with the block size is 5x5 pixels. This shows that the motion feature of the POC-ABS motion estimation method has the highest micro-expression recognition accuracy compared to other motion-based features.

The arrow of MV resulting from the matching process using the POC-ABS method is used to observe motion leaks when micro-expression occurs. In this study, the MV ratio of each face component has compared. The highest MV ratio in a certain facial component area shows the most dominant movement among others. The motion estimation on the four emotional labels shows that the motion in the eye area (right eyebrow, left eyebrow, right eye, and left eye) dominates all expressions compared to the motion in the mouth area, as shown in Table 4. The movements in the right and left eyebrows dominate the expression: Disgust and surprise. The movements in the right eyebrow, left eyebrow and right eye dominate the sad expression. Whereas in happy expressions, the movement of the right eye and the left eye dominate, this is shown in Fig. 5. For example, vector motion in all areas of facial components for sad expressions is shown in Fig. 4. Face images are displayed on the onset, apex, and offset frames. By using a block size of 5x5 pixels, it can be seen that there are many motion vectors in the right eyebrow, left eyebrow, and right eye. In general, there is an increase in the number of arrows from apex to offset. Also in the same frame, the number of arrows for the right eyebrow, left eyebrow, and right eye is more dominant than any other facial component.

When compared with previous studies, as shown in Table 6, Duan also stated where the eye area was the most influential in each expression [28]. Wen-Jiang Yan's research, which uses FACS observations or based on AU movements, also states that the eye area also affects the expression of disgust and shock. However, a little different in the expression of happiness, the area of the cheeks, and lips that affect this expression [40].

6. Conclusion

In this paper, a novel approach of motion estimation, the POC-ABS method, could recognize the subtle motion of micro-expressions faster and accurately. This method also could identify where the expression leakage during the micro-expression occurs is.

At the apex state of expression, the POC-ABS could show the number of motion vectors increased compared with the onset and offset state. In general, the movement in the facial component area is less than half the area, so this is consistent with the nature of micro-expression i.e. the intensity of low muscle and dominance of movement only occurs in certain areas. In the computational time measurement, the average time needed by the POC-ABS method to match between frames is 0.03 seconds. This method is sufficient for real-time and practical implementation.

With the SVM and KNN classification, recognition of micro-expressions using the POC-ABS method is more accurate than other motion feature-based micro-expression recognition methods. The smaller the block sizes being compared, the higher the accuracy value.

By using the MV arrow ratio value, this method could identify where the expression leakage during the micro-expression occurs is. The analysis showed that the point of leakage at the time of micro-expression was dominated by movement in the upper face area, especially in the right and left eyebrow areas.

For further research, more detailed observations are needed in reading the direction of the facial component's movements and this method can be used in further research, the realtime recognition of micro-expressions.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, U.D. Rosiani; methodology, U.D. Rosiani, N. Muna and E. Mulyanto; software, U.D. Rosiani and P. Choirina; validation, U.D. Rosiani, S. Sumpeno, and M.H. Purnomo; formal analysis, U.D. Rosiani, P. Choirina, S. Sumpeno, and M.H. Purnomo; investigation, U.D. Rosiani, P. Choirina, S. Sumpeno, and M.H. Purnomo; resources, U.D. Rosiani, and P. Choirina; data curation, U.D. Rosiani, and P. Choirina; writing—original draft preparation, U.D. Rosiani; writing—review and editing, U.D. Rosiani,; visualization, U.D. Rosiani;

supervision, S. Sumpeno, and M.H. urnomo; project administration, U.D. Rosiani. All authors read and approved the final manuscript.

Acknowledgments

This work is supported by the Department of Technology Information, Politeknik Negeri Malang and the research group of Multimedia Computational Laboratory at Institut Teknologi Sepuluh Nopember Surabaya.

References

- [1] P. Ekman, ‘Lie catching and microexpressions’, in *The Philosophy of Deception*, C. W. Martin, Ed. Oxford University Press, 2009, pp. 118-133.
- [2] U. D. Rosiani, A. Atmoko, S. Sumpeno, and M. H. Purnomo, “The synthesis of Javanese woman’s facial image on anger expression based on emotion regulation”, In: *Proc. of 2015 4th International Conf. on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME)*, Nov. 2015, pp. 185–190.
- [3] W.-J. Yan, Q. Wu, J. Liang, Y.-H. Chen, and X. Fu, “How Fast are the Leaked Facial Expressions: The Duration of Micro-Expressions”, *J Nonverbal Behav*, Vol. 37, No. 4, pp. 217–230, 2013.
- [4] Q. Wu, X. Shen, and X. Fu, “The Machine Knows What You Are Hiding: An Automatic Micro-expression Recognition System”, in *Affective Computing and Intelligent Interaction*, Springer, Berlin, Heidelberg, 2011, pp. 152–162.
- [5] S. Polikovsky, Y. Kameda, and Y. Ohta, “Facial micro-expressions recognition using high speed camera and 3D-gradient descriptor”, In: *Proc. of 3rd International Conf. on Imaging for Crime Detection and Prevention (ICDP 2009)*, pp. 1–6, 2009.
- [6] N. Muna, U. D. Rosiani, E. M. Yuniamo, and M. H. Pumomo, “Subpixel subtle motion estimation of micro-expressions multiclass classification”, In: *Proc. of 2017 IEEE 2nd International Conf. on Signal and Image Processing (ICSIP)*, pp. 325–330, 2017.
- [7] H. Dong and S. Chen, “Research of moving target tracking algorithm for video”, *International Journal of Intelligent Engineering and Systems*, Vol. 4, No. 1, pp. 18–25, 2011.
- [8] T. R. Saeed, S. Mutashar, A. N. Abed, M. H. Al-Muifraje, A. L. Shurajji, and S. Abd-Elghany, “Human Motion Detection through Wall Based on Micro-Doppler Using Kalman Filter Combined with Convolutional Neural Network”, *International Journal of Intelligent Engineering and Systems*, Vol. 13, No. 6, 2020
- [9] H. Zhuang, M. Yang, Z. Cui, and Q. Zheng, “A method for static hand gesture recognition based on non-negative matrix factorization and compressive sensing”, *IAENG International Journal of Computer Science*, Vol. 44, No. 1, pp. 52–59, 2017.
- [10] B. Setiyono, M. Hariadi, and M. H. Purnomo, “Investigation of Superresolution using Phase based Image Matching with Function Fitting”, *Research Journal of Engineering Sciences*, Vol. 2278, p. 9472, 2012.
- [11] T. S. Shinde and A. K. Tiwari, “Efficient direction-oriented search algorithm for block motion estimation”, *IET Image Processing*, Vol. 12, No. 9, pp. 1557–1566, 2018.
- [12] K. M. Goh, C. H. Ng, L. L. Lim, and U. U. Sheikh, “Micro-expression recognition: an updated review of current trends, challenges and solutions”, *Vis Comput*, Vol. 36, No. 3, pp. 445–468, 2020.
- [13] Y.-H. Oh, J. See, A. C. L. Ngo, R. C. W. Phan, and V. M. Baskaran, “A Survey of Automatic Facial Micro-Expression Analysis: Databases, Methods, and Challenges”, *Frontiers in Psychology*, Vol. 9, 2018.
- [14] M. Shreve, S. Godavarthy, V. Manohar, D. Goldgof, and S. Sarkar, “Towards macro- and micro-expression spotting in video using strain patterns”, in *2009 Workshop on Applications of Computer Vision (WACV)*, pp. 1–6, 2009.
- [15] M. Shreve, S. Godavarthy, D. Goldgof, and S. Sarkar, “Macro- and micro-expression spotting in long videos using spatio-temporal strain”, in *Face and Gesture 2011*, pp. 51–56, 2011.
- [16] S.-T. Liong, J. See, R. C.-W. Phan, Y.-H. Oh, A. C. Le Ngo, K. Wong, and S.-W. Tan, “Spontaneous subtle expression detection and recognition based on facial strain”, *Signal Processing: Image Communication*, Vol. 47, pp. 170–182, 2016.
- [17] F. Xu, J. Zhang, and J. Z. Wang, “Microexpression Identification and Categorization Using a Facial Dynamics Map”, *IEEE Transactions on Affective Computing*, Vol. 8, No. 2, Art. No. 2, 2017.
- [18] S.-J. Wang, S. Wu, X. Qian, J. Li, and X. Fu, “A main directional maximal difference analysis for spotting facial movements from long-term videos”, *Neurocomputing*, Vol. 230, pp. 382–389, 2017.
- [19] S. L. Happy and A. Routray, “Fuzzy Histogram of Optical Flow Orientations for Micro-

- expression Recognition”, *IEEE Transactions on Affective Computing*, pp. 1–1, 2017.
- [20] S. Metkar and S. Talbar, *Motion Estimation Techniques for Digital Video Coding*. Springer India, 2013.
- [21] K. Takita, T. Aoki, Y. Sasaki, T. Higuchi, and K. Kobayashi, “High-accuracy subpixel image registration based on phase-only correlation”, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, Vol. 86, No. 8, pp. 1925–1934, 2003.
- [22] M. Miura, S. Sakai, S. Aoyama, J. Ishii, K. Ito, and T. Aoki, “High-accuracy image matching using phase-only correlation and its application”, In: *Proc. of 2012 SICE Annual Conf. (SICE)*, pp. 307–312, 2012.
- [23] F. Francis-Lothai and D. B. Bong, “A fingerprint matching algorithm using bit-plane extraction method with phase-only correlation”, *International Journal of Biometrics*, Vol. 9, No. 1, pp. 44–66, 2017.
- [24] Y. Douini, J. Riffi, M. A. Mahraz, and H. Tairi, “Solving sub-pixel image registration problems using phase correlation and Lucas-Kanade optical flow method”, in *2017 Intelligent Systems and Computer Vision (ISCV)*, pp. 1–5, 2017.
- [25] L. H. Chien and T. Aoki, “Robust motion estimation for video sequences based on phase-only correlation”, pp. 441–446, 2004.
- [26] G. Meneghetti, M. Danelljan, M. Felsberg, and K. Nordberg, “Image Alignment for Panorama Stitching in Sparsely Structured Environments”, in *Image Analysis*, Cham, pp. 428–439, 2015.
- [27] P. Zhang, P. Wei, and H.-Y. Yu, “Biogeography-based optimisation search algorithm for block matching motion estimation”, *IET Image Processing*, Vol. 6, No. 7, pp. 1014–1023, 2012.
- [28] X. Duan, Q. Dai, X. Wang, Y. Wang, and Z. Hua, “Recognizing spontaneous micro-expression from eye region”, *Neurocomputing*, Vol. 217, pp. 27–36, 2016.
- [29] A. A. Goshtasby, *2D and 3D Image Registration: For Medical, Remote Sensing, and Industrial Applications*. Wiley-Interscience, 2005.
- [30] R. A. Asmara and M. Hariadi, “Accelerating phase based motion estimation with hierarchical search technique using parallel threading in graphical processing unit (GPU)”, *Int. J. Comput. Sci. Network Security*, Vol. 9, No. 8, pp. 140–146, 2009.
- [31] R. Yaakob, A. Aryanfar, A. A. Halin, and N. Sulaiman, “A Comparison of Different Block Matching Algorithms for Motion Estimation”, *Procedia Technology*, Vol. 11, pp. 199–205, 2013.
- [32] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, “Robust Discriminative Response Map Fitting with Constrained Local Models”, In: *Proc. of 2013 IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 3444–3451, 2013.
- [33] N. H. Barnouti, M. H. N. Al-Mayyahi, and S. S. M. Al-Dabbagh, “Real-Time Face Tracking and Recognition System Using Kanade-Lucas-Tomasi and Two-Dimensional Principal Component Analysis”, In: *Proc. of 2018 International Conf. on Advanced Science and Engineering (ICOASE)*, pp. 24–29, 2018.
- [34] U. D. Rosiani, A. R. T. H. Ririd, P. Choirina, A. G. Sooai, S. Sumpeno, and M. H. Purnomo, “Micro Expression: Comparison of Speed and Marking Accuracy in Facial Component Detection”, In: *Proc. of 2018 International Conf. on Computer Engineering, Network and Intelligent Multimedia (CENIM)*, pp. 221–226, 2018.
- [35] R. Chen, Y. Cao, H. Sun, and W. Yang, “A modified method for face recognition using SVM”, *International Journal of Intelligent Engineering and Systems*, Vol. 1, No. 1, pp. 22–29, 2008.
- [36] A. G. Sooai, P. Batarius, Y. C. H. Siki, P. A. Nani, N. M. R. Mamulak, E. Ngaga, U. D. Rosiani, S. Sumpeno, M. H. Purnomo, and S. D. B. Mau, “Comparison of Recognition Accuracy on Dynamic Hand Gesture Using Feature Selection”, In: *Proc. of 2018 International Conf. on Computer Engineering, Network and Intelligent Multimedia (CENIM)*, pp. 270–274, 2018.
- [37] Y. Wang, J. See, Y.-H. Oh, R. C.-W. Phan, Y. Rahulamathavan, H.-C. Ling, S.-W. Tan, and X. Li, “Effective recognition of facial micro-expressions with video motion magnification”, *Multimedia Tools and Applications*, Vol. 76, No. 20, pp. 21665–21690, 2017.
- [38] S. Zhang, B. Feng, Z. Chen, and X. Huang, “Micro-Expression Recognition by Aggregating Local Spatio-Temporal Patterns”, in *MultiMedia Modeling*, pp. 638–648, 2017.
- [39] I. P. Adegun and H. B. Vadapalli, “Facial micro-expression recognition: A machine learning approach”, *Scientific African*, Vol. 8, p. e00465, 2020.
- [40] W.-J. Yan, X. Li, S.-J. Wang, G. Zhao, Y.-J. Liu, Y.-H. Chen, and X. Fu, “CASME II: An Improved Spontaneous Micro-Expression

- Database and the Baseline Evaluation”, *PLoS One*, Vol. 9, No. 1, 2014.
- [41] M. Iwasaki and Y. Noguchi, “Hiding true emotions: micro-expressions in eyes retrospectively concealed by mouth movements”, *Scientific Reports*, Vol. 6, p. 22049, 2016.
- [42] E. I. Andelin and A. S. Rusu, “Investigation of Facial Micro-expressions of Emotions in Psychopathy – A Case Study of an Individual in Detention”, *Procedia - Social and Behavioral Sciences*, Vol. 209, pp. 46–52, 2015.
- [43] X. Zhu, X. Ben, S. Liu, R. Yan, and W. Meng, “Coupled source domain targetized with updating tag vectors for micro-expression recognition”, *Multimedia Tools and Applications*, Vol. 77, No. 3, pp. 3105–3124, 2018.
- [44] M. Peng, C. Wang, T. Chen, G. Liu, and X. Fu, “Dual temporal scale convolutional neural network for micro-expression recognition”, *Frontiers in psychology*, Vol. 8, p. 1745, 2017.
- [45] S.-T. Liong, J. See, R. C.-W. Phan, K. Wong, and S.-W. Tan, “Hybrid Facial Regions Extraction for Micro-expression Recognition System”, *Journal of Signal Processing Systems*, pp. 1–17, 2017.
- [46] 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, 2018.