Micro-expression recognition based on motion detection method

U D Rosiani¹*, P Choirina² and M N Shoumi¹

¹Department of Information Technology, State Polytechnic of Malang, Malang, Indonesia
²Faculty of Science and Technology, Universitas Islam Raden Rahmat Malang, Malang, Indonesia

*rosiani@polinema.ac.id

Abstract. Micro-expressions are emotional representations that occur spontaneously and cannot be controlled by humans. The micro-expression movements are temporary with fast duration and have subtle movements with little intensity. This is difficult to detect with the human eye. Previous studies have shown that micro-expression movements occur in several areas of the face. This study aims to determine the subtle movements in several areas of the face using the motion detection method. We compared the performance of two motion detection methods: the optical flow method and the Block Matching Algorithm (BMA) method. The optical flow method uses the Kanade-Lucas Tomasi (KLT) method and the BMA method uses the Phase Only Correlation (POC) algorithm. Observations were carried out based on region, where the face area was divided into several observation areas: eyebrows, eyes and mouth. Both methods perform motion detection between frames. The KLT method tracks the movement of the observation points on the frame movement. Meanwhile, the POC method matches the blocks between frames. If the two blocks are the same, no motion vector is generated. However, if the two blocks are different, it is assumed that there is a translational motion and a motion vector is generated. Experiments were conducted using a dataset from CASME II with emotional classes of disgust, happiness, surprise, and sadness. The classification accuracy of the POC method is 94% higher than the KLT method of 84.8% which uses the SVM classification.

1. Introduction

Facial expressions are usually considered to be physiological reactions to the emotions within. Human facial expressions are divided into two parts: macro-expression and micro-expression. Macro expression is defined by easily recognizable movements of the facial muscle over many areas of the face which occur between ½ seconds and 4 seconds. Meanwhile, micro-expression movements take the form of smooth muscle movements in some areas of the face that are quicker, from 1/25 seconds to 1/5 seconds [1]. Other people are often not aware and do not see this movement [2]. Similar to macro expressions, micro expressions can be grouped into six basic expression: happy, surprise, anger, sad, disgust, and fear [3]. In many studies on the recognition of micro-expressions, the approach to research is divided into two classes: appearance-based and motion-based [4].

The main approaches to appearance-based methods, such as Gabor wavelet [5] and local binary pattern (LBP) [6], focus on the use of texture representations [7]. However, the texture feature makes it difficult to capture the fast, low-intensity display changes that occur in micro-expressions. Meanwhile,
motion features that are usually derived from optical flow are considered to be a robust and accurate approximation to the recognition of micro expressions [8]. This method tries to recognize facial activity by analyzing the movement itself. Since micro-expressions can be well characterized by motion features, the performance of the motion approach is better than that of the appearance-based approach [9].

In this study, the performance of two motion detection methods for recognition of micro-expressions is compared: optical flow method and Block Matching Algorithm (BMA) method. The first method, optical flow method, is part of method in pixel-based motion aims to observe and detect non-rigid motion of facial components [10]. Liu et al., used Main Directional Mean Optical-flow (MDMO) based on robust optical flow in micro-expression video clips by partitioning the area on the face (ROI) [9]. MDMO produces better performance than LBP-TOP and HOOF using SMIC, CASME and CASME II dataset. Another optical flow method is Kanade-Lucas-Tomasi (KLT), capable of detecting subtle motion based on point feature tracking. KLT calculates the number of differences in the square of the intensity between features on two frames. Asmara et al., applied the KLT method for the recognition of micro expressions based on tracking the points formed by the facial landmark method, namely DRMF [11]. This study resulted in an accuracy rate of 79.3% in CASME II with SVM.

The second method, BMA metod, is a part of block-based motion which estimates block-based motion quickly [12]. One of the matching methods in the comparison between blocks is Phase-Only Correlation (POC) which is better than other block matching methods [13]. The POC method can be used to detect translation motion that occur in an image based on the displacement of two images in the cross-spectrum power phase [14]. In this study, we compared the performance of KLT and POC method for micro expression recognition. Observations were made based on region, where the face area is divided into several observation areas: eyebrows, eyes and mouth. Region based observation is better than whole face observation [15].This study aims to obtain the best accuracy with minimal computation time from the comparison of these methods, because the movements generated by micro expressions occur very quickly.

2. Proposed methods

In this study, there are four stages: 1) pre-processing; 2) motion detection; 3) feature extraction; and 4) classification. Figure 1, shows a block diagram of a micro-expression recognition based on motion detection method. In this section, we describe the whole process the comparison of motion detection methods is Phase-only Correlation (POC) and Kanade Lucas Tomasi (KLT) aims to find the best accuracy of the two methods.

![Figure 1. The block diagram for micro-expression recognition based on motion detection method.](image-url)

2.1. Pre-processing

This section discusses the detection method of the location of components on the face, namely the right eyebrow, left eyebrow, right eye, left eye, and mouth. In the preprocessing stage, we only use the first frame as a reference for the reference frame. This occurs because the changes that result in micro-expressions are very small and can minimize the time in the sequence.

This section discusses the detection method of the location of components on the face, namely the right eyebrow, left eyebrow, right eye, left eye, and mouth. In the preprocessing stage, we only use the
first frame as a reference for the reference frame. This occurs because the changes that result in micro-expressions are very subtle and can minimize the time in the sequence.

We used the Discriminative Response Map Fitting (DRMF) method to locate the facial components accurately in a very fast time [16]. The DRMF method can adjust facial shapes with impressive performance and is well suited for handling very large occlusion and illumination conditions [11]. The detection of facial components with DRMF has 3 stages. First, the detection of face location using the Viola-Jones method. Second, computed starting point feature set by extracting patch responses by low dimensional projection. Furthermore, this method repeats the initial feature point by connecting with the target frame.

![Image](image1.png)

**Figure 2.** The flow of face component detection.

The result of this method is a set of points totaling 49 and scattered over the area of the face component which has a coordinate value as in Figure 2. Furthermore, we form the ROI (Region of Interest) on the face component and the landmark points of the DRMF which are used as reference points, for other frames. The parameters of the ROI required include the X and Y coordinate values, width, and height for each face component. We used the reference points used in previous studies as in Table 1 [16] for the formation of ROI for each facial component.

| Facial Component | Landmark Points | X   | Y   | Width | Height |
|------------------|-----------------|-----|-----|-------|--------|
| Left Eyebrow     | 1-5             | x (1)| y (2)| 110   | 28     |
| Right Eyebrow    | 6-10            | x (6)| y (9)| 110   | 28     |
| Left Eye         | 20-25           | x (20)| y (22)| 81    | 34     |
| Right Eye        | 26-31           | x (26)| y (30)| 81    | 34     |
| Mouth            | 32-41           | x (32)| y (34)| 139   | 39     |

The POC method uses the 5x5 block area feature for each ROI as shown in Figure 4. Whereas the KLT method uses the point feature for each ROI component. As in Figure 3, the points generated in each ROI are formed by the function of the Matlab is detectMinEigenFeatures() [17]. This function is used to forming corner points (interest points) on an object based on the minimum eigenvalue algorithm.

![Image](image2.png)

**Figure 3.** Facial point feature using Eigenvalues for KLT.

![Image](image3.png)

**Figure 4.** Block area feature for POC.
2.2. Motion detection for micro-expression recognition

2.2.1. Kanade Lucas Tomasi (KLT). The optical flow method calculates pixel movement between frames over time. Optical flow can track the motion of each pixel in the image. The calculation used for all pixels is dense optical flow and for some pixels, it uses the sparse optical flow method. KLT is a representative sparse optical flow algorithm for the feature tracking process. We use this method to trace the points that have been formed in each sequence frame. The grayscale value of the same spatial point pixel does not change in all frames. It is called constant grayscale values which are used to assume the optical flow method. Figure 5 explains that when a pixel in position \((x, y)\) at time \(t\) moves to the position \((x + dx, y + dy)\) at time \(t + dt\), the grayscale values of the two pixels are same.

![Figure 5. Kanade Lucas Tomasi (KLT) feature tracking algorithm [18].](image)

First, a set of points formed in the previous process (see Figure 3), the first frame is initiated as a starting point before tracking. Furthermore, the KLT method traces the reference frame. KLT calculates the sum of the differences in the squares of intensity between features in the current and previous frames. The motion of a feature point is defined as a displacement that minimizes the amount of difference. This process is repeated in a sequence of images until all feature points can be traced [11]. Motion vectors are formed when the location of the point on the current frame moves with the movement of the reference frame. All point locations in each frame that have been through the tracking process are stored as feature extraction data.

2.2.2. Phase-Only Correlation (POC). Image registration with POC makes it possible to predict movement between images with sub-pixel accuracy. POC functions to measure the similarity between two blocks in the frame, the current block, and the reference block. The function used in POC is the inverse 2D discrete Fourier transformation. The translation and similarity between the two images can be obtained based on the position and height of the correlation peaks. If the two images are the same, the POC graph produces a peak that is 1. On the other hand, if the two images are different, the resulting graph is not 1. The graph of the POC function can be seen in Figure 6.

The main step in calculating the movement of each component of the facial area is dividing each area into blocks. The block size used is 5x5 pixels as shown in Figure 4. Next, match each block between the current frame \(I(x, y, t)\) and the reference frame \(I(x, y, t + 1)\) using the POC function. If both are equal or there is no shift, the POC value approaches 1 and a motion vector is formed on the block. On the other hand, if the two image blocks are different, no motion vector is formed.

2.3. Feature extraction

The output of the two motion detection methods, namely POC and KLT, is the value of the motion vector which consists of the x and y coordinates in each frame comparison. This value will be used as a motion feature for each component of the face area. Motion features consist of horizontal (Rx) and vertical (Ry) components, the magnitude of the vector (M), and the direction of the resulting vector.
3. Experimental results

3.1. Database
To verify the effectiveness of the comparison of motion detection methods on the recognition of micro-expressions, we conducted experiments using the Chinese Academy of Sciences Micro-expression (CASME II) dataset. CASME II is a comprehensive database of spontaneous micro expression data and contains 247 video samples with 26 Asian participant subjects with an average age of 22.03 years. Each video in CASME II gives one micro expression to the subject of five categories of micro expression: Happiness, Disgust, Repression, Surprise, Other. In this study, we only used four micro-expressions, namely: Happiness, Disgust, Surprise and Sadness.

3.2. Processing time
The recognition of micro expressions requires real-time processing, it takes fast processing time. Experiments in this study conducted in MATLAB run on Intel Core i5 CPU with RAM size 4 GHz using 64-bit Windows OS. Figure 7 describes the graph of time requirements for processing motion detection methods used in this study, namely POC and KLT. In this graph, it can be seen that the time require for the KLT method is smaller than the POC with an average time of 0.0445 seconds. Whereas the POC method requires time with an average time of 0.2041 seconds.

![Figure 6. Phase-only correlation between two images.](image)

![Figure 7. Time comparison of motion detection method on micro expression recognition.](image)
3.3. Comparison accuracy of motion detection methods in micro expression recognition

In this study, the results of feature extraction from the motion detection method will be recognized by machine learning using a Support Vector Machine (SVM) as a measure of the resulting accuracy in classifying emotions in micro expressions. The SVM used is non-linear using the Radial Basis Function (RBF) kernel. RBF is used because it can map the sample to a higher dimensional space. This study resulted in the POC method the resulting accuracy rate of 94.3% was better than the KLT method with an accuracy of 84.8%. This study produced the best accuracy with the KLT (optical flow) method compared to previous studies which are shown in Table 2.

| Year  | Author               | Dataset          | Feature                  | Classifier   | Accuracy       |
|-------|----------------------|------------------|--------------------------|--------------|----------------|
| 2016  | Liong et al., [19]   | CASME II and SMIC| Optical Strain           | SVM          | 63.41% on CASME II, 52.44% on SMIC |
| 2017  | Zhang et al., [20]   | CASME II         | LBP-TOP, Optical Flow    | KNN, SVM and RF | Random Forest 62.5% |
| 2018  | Zhu et al., [21]     | CASME II         | LBP-TOP and Optical Flow | SVM          | 53.3%          |
| 2020  | This study           | CASME II         | POC, KLT                 | SVM          | POC 94.3% KLT 84.8% |

4. Conclusion

The movements produced by micro-expressions are very subtle and occur in a short period of time. This makes this movement difficult to detect with the human eye. The two methods used in the study, KLT and POC, can detect the movements that occur. The KLT method tracks the motion at the observation points, while the POC method compares the blocks between frames. In the comparison of the performance of the two methods, the classification accuracy with the POC method is higher than the KLT method. This is because more motion features are used. The smaller the block size that is matched by the POC matching method, the more motion features are generated. Meanwhile, for the comparison of processing times, the KLT method is faster than the POC method, this is because less parts are detected in KLT method. Research using a combination of the two methods can be further developed, in order to obtain better accuracy and processing time.

References

[1] Yan W J, Wu Q, Liang J, Chen Y H and Fu X 2013 How fast are the leaked facial expressions: The duration of micro-expressions Journal of Nonverbal Behavior 37(4) 217-230
[2] Liong S T, See J, Wong K, Le Ngo A C, Oh Y H and Phan R 2015 Automatic apex frame spotting in micro-expression database 2015 3rd IAPR Asian conference on pattern recognition (ACPR) 665-669
[3] Takalkar M A and Xu M 2017 Image based facial micro-expression recognition using deep learning on small datasets 2017 international conference on digital image computing: techniques and applications (DICTA) 1-7
[4] Lu H, Kpalma K and Ronsin J 2018 Motion descriptors for micro-expression recognition. Signal Processing: Image Communication 67 108-117
[5] Zheng H 2017 Micro-expression recognition based on 2d gabor filter and sparse representation Journal of Physics: Conference Series 787 012013
[6] Ahonen T, Hadid A and Pietikainen M 2006 Face description with local binary patterns: Application to face recognition IEEE transactions on pattern analysis and machine intelligence 28(12) 2037-2041
[7] Wu Q, Shen X and Fu X 2011 The machine knows what you are hiding: an automatic micro-expression recognition system international conference on affective computing and intelligent
[8] Liu Y J, Li B J and Lai Y K 2018 Sparse MDMO: Learning a discriminative feature for spontaneous micro-expression recognition *IEEE Transactions on Affective Computing*

[9] Liu Y J, Zhang J K, Yan W J, Wang S J, Zhao G and Fu X 2015 A main directional mean optical flow feature for spontaneous micro-expression recognition *IEEE Transactions on Affective Computing* 7(4) 299-310

[10] Happy S L and Routray A 2017 Fuzzy histogram of optical flow orientations for micro-expression recognition *IEEE Transactions on Affective Computing* 10(3) 394-406

[11] Asmara R A, Choirina P, Rahmad C, Setiawan A, Rahutomo F, Yusron R D R and Rosiani U D 2019 Study of DRMF and ASM facial landmark point for micro expression recognition using KLT tracking point feature *Journal of Physics: Conference Series* 1402(7) 077039

[12] Zhang P, Wei P and Yu H Y 2012 Biogeography-based optimisation search algorithm for block matching motion estimation *IET image processing* 6(7) 1014-1023

[13] Setiyono B, Hariadi M and Purnomo M H 2012 Investigation of Superresolution using Phase based Image Matching with Function Fitting *Res. J. Eng. Sci.* 2278 9472

[14] Reddy B S and Chatterji B N 1996 An FFT-based technique for translation, rotation, and scale-invariant image registration *IEEE transactions on image processing* 5(8) 1266-1271

[15] Duan X, Dai Q, Wang X, Wang Y and Hua Z 2016 Recognizing spontaneous micro-expression from eye region *Neurocomputing* 217 27-36

[16] Rosiani U D, Ririd A R T H, Choirina P, Sooai A G, Sumpeno S and Purnomo M H 2018 Micro Expression: Comparison of Speed and Marking Accuracy in Facial Component Detection 2018 International Conference on Computer Engineering, Network and Intelligent Multimedia (CENIM) 221-226

[17] Shi J 1994 Good features to track *1994 Proceedings of IEEE conference on computer vision and pattern recognition* 593-600

[18] Piao J C and Kim S D 2017 Adaptive monocular visual–inertial SLAM for real-time augmented reality applications in mobile devices *Sensors* 17(11) 2567

[19] Liong S T, See J, Phan R C W, Oh Y H, Le Ngo A C, Wong K and Tan S W 2016 Spontaneous subtle expression detection and recognition based on facial strain *Signal Processing: Image Communication* 47 170-182

[20] Zhang S, Feng B, Chen Z and Huang X 2017 Micro-expression recognition by aggregating local spatio-temporal patterns *International Conference on Multimedia Modeling* 638-648

[21] Zhu X, Ben X, Liu S, Yan R and Meng W 2018 Coupled source domain targetized with updating tag vectors for micro-expression recognition *Multimedia Tools and Applications* 77(3) 3105-3124