Face expression recognition using Local Gabor Binary Pattern Three Orthogonal Planes (LGBP-TOP) and Support Vector Machine (SVM) method

R R K Dewi¹, F Sthevanie² and A Arifianto³
Informatics, Telkom University, Bandung, Indonesia
E-mail:¹rosekumala@student.telkomuniversity.ac.id,²sthevanie@telkomuniversity.ac.id,³anditya@telkomuniversity.ac.id

Abstract. The video expression recognition system has been created before using the Local Gabor Binary Pattern Three Orthogonal Planes (LGBP-TOP) extraction method and the Support Vector Machine (SVM) classification method. However, the recognizable facial expressions use the entire area of the face image, while the expression can be recognized from the change of face fiducial point on the eyes and lips only. In this study, the introduction of facial expressions was developed using LGBP-TOP and SVM methods by focusing on facial and lip images only. Therefore, an algorithm is needed to extract the eye and lip area of the face image using 3x3 blocks and 4x4 blocks, which will then be used as input on the LGBP-TOP method. After the image of the eyes and lips extracted its features, the extraction results are classified using the SVM method. The results obtained is the recognition of facial expressions using the eye and lip area get 80% accuracy and better than using the entire area of the face, eye area only, and lip area only.

1. Introduction
Facial expressions are facial movements or facial gestures that indicate clearly the emotions a person experience [1]. At this time, facial expressions can be recognized using expression recognition technology by measuring fiducial points on the geometric structure of the face [2]. The facial expression recognition approach uses a picture or video approach. In the video approach, facial expressions are recognized based on the order of parts of the image [1]. Facial expression recognition can be used as learning for robots in recognizing human facial expressions and making hospital alarms to recognize abnormal facial expressions of patients [2].

There have been several previous studies conducted in recognizing facial expressions. In Xie Liping's research, the construction of facial expression recognition systems using the entire face area was carried out using the LGBP-TOP and SVM methods. All face areas are used as input to the Gabor method which is then divided into 4x4 blocks before included in the LBP method. The level of accuracy generated in the division of the face area into 4x4 blocks is 83% [1]. In Almaev's research, a comparison was made between the LBP, LBP-TOP, LGBP, and LGBP-TOP methods. The data set used is derived from the MMI database and CK database. The study was carried out by calculating 2AFC at each AU detection. The results obtained from this study are LGBP-TOP is a better method when compared with the LBP, LBP-TOP, and LGBP methods [3]. The use of the Gabor filtering method can clarify the borders on the image so that it will increase the resolution and orientation of the image [1]. The use of TOP on LBP
may expand the description of temporal information in imagery [3]. Then the SVM method is able to classify high-dimensional data and have a high generalization [2].

A previously built video expression recognition system uses the extraction method of Local Gabor Binary Pattern Three Orthogonal Planes (LGBP-TOP) and Support Vector Machine (SVM) classification method using all face areas [1]. However, Joumana calligraphers argue that facial expressions can be recognized only on the eyebrows, eyes, and lips [4]. Therefore, in this study is built an introduction to facial expressions with the LGBP-TOP and SVM methods that focus on the eyes and lips on a video image that is able to know the part of the eye or lips or both of which produce a high degree of accuracy.

2. LGBP-TOP
The process blocks of the LGBP-TOP method can be illustrated in figure 1. The first step, the video obtained is made into a sequence of images which is then cutting the face area and normalized. Then the resulting sequence of images is consonated by using 40 gabor filters to 40 GMSs. The result of convolution gabor taken the eye and lip area that made the input on LBP-TOP. Finally, the LGBP-TOP histogram is combined into the final feature of feature extraction.

![Figure 1. Process blocks of LGBP-TOP.](image)

2.1. Preprocessing
In the preprocessing stage of input data in the form of video is done face detection using Viola-Jones face detection algorithm so that the image sequences of the face area.

![Figure 2. Preprocessing using Viola-Jones algorithm.](image)

2.2. Convolution
In general, convolution is a combination of two series of numbers that produce the third series of numbers. The convolution operation in the dwimatra plane as a discrete function is in equation (1).

\[
h(x, y) = f(x, y) * g(x, y) = \sum_{\alpha=-\infty}^{\infty} \sum_{\beta=-\infty}^{\infty} f(\alpha, \beta) g(x - \alpha, y - \beta) \tag{1}
\]

Where \( f(x, y) \) is a digital image in the form of a matrix, \( g(x, y) \) is a convolution mask or a convolution kernel, and \( h(x, y) \) is the result of the convolution operation of \( f(x, y) \) and \( g(x, y) \). In convolution theory, convolution kernels are expressed in the form of a matrix generally \( 3 \times 3 \), where the size of the kernel matrix is usually smaller when compared with the size of the image. Each element in the kernel matrix is called the convolution [5].
2.3. Gabor filtering

In the preprocessing stage of input data in the form of video is done face detection using Viola-Jones face detection algorithm so that the sequence of image of the face area. After the image sequence is done preprocessing, a number of sequences of face area images will be done convolution with 40 gabor filters. The gabor filters [2] used in this research to recognize facial expressions are in equation (2).

$$g = e^{-\frac{x'^2 + y'^2}{2\sigma^2}} \cos \left(2\pi \frac{x'}{\lambda} + \varphi \right)$$ (2)

Where $\lambda$ is a sinusoidal wavelength, $\theta$ is the orientation of the gabor filter, $\sigma$ is the Gaussian wavelength, $\psi$ is the offset phase, $\gamma$ is the spatial ratio [2].

Figure 3. Gabor filter.

Figure 4. Face image after gabor filtering.

Before being included in LBP, image sequences is divided into 4x4 blocks as shown below:

Figure 5. 4x4 block of LBP.
In the picture of the face is taken 8 blocks that show the area of the eyes and lips.

2.4. Local Binary Pattern Three Orthogonal Planes

LBP is one of the first texture descriptor known in 1994 by Ojala. LBP is a simple texture operator by doing thresholding on each neighboring pixel and considering the result as a binary number (0 or 1). The LBP method uses grayscale images with 3 x 3, where the middle position pixel is used as a threshold. Mathematically, LBP operators can be defined as follows:

\[
(x) = \begin{cases} 
0, & x < 0 \\
1, & x \geq 0 
\end{cases}
\]  

(3)

\[
LBP(P, R) = \sum_{p=0}^{P^2-1} s(g_p - g_c) 2^p 
\]  

(4)

where \((P, R)\) is the sampling point \(P\) on radius \(R\), \(g_p\) is the gray degree value in the \(p\)-neighbor pixel, \(g_c\) is the gray degree value at the center pixel [6].

The neighboring pixel value is binary with the condition that the neighboring pixel value is smaller than the threshold it will be 0, whereas if the value is greater than the threshold it will be 1. After that, the binary value of each pixel is multiplied by the specified weight, i.e. multiplied by \(2^n\), where \(n\) is a sequence of neighboring pixels starting from the top left corner by circling the pixels in the middle position starting from 0-7. The multiplication result is then summed to be the LBP value in the middle position pixel [6].

LBP-TOP combines the temporal and spatial features of image sequences using the LBP basic method which extracts the dynamic texture feature of the image sequence using three orthogonal fields. This dynamic texture feature is used to represent the spatial and temporal characteristics of the image sequence [7].

Three orthogonal fields representing each sequence of facial expressions are the XY plane on the T axis, the XT plane on the Y axis, and the YT field on the X axis. The XY-LBP field is the spatial feature information of the image sequence and the XT-LBP and YT-LBP fields are information of the temporal features of the image sequence. The entire histogram of the three orthogonal planes is merged into one to represent a feature vector for one block in the image [1].

![Figure 6. Three Orthogonal Planes of LBP [1].](image)

3. SVM classifier

SVM is a statistical learning method based on structural risk minimization to maximize the gap between the two data classes by finding the dividing field and ensuring the correct classification ratios [1]. SVM can separate data sets in two ways: Linearly Separable Data and Non-Linearly Separable Data [8]. Linearly, SVM can use Lagrange formula with Lagrange Multiplier is as follows:

\[
min_{w, b} L_p(w, b, \alpha) = \frac{1}{2} |w|^2 - \sum_{i=1}^{n} \alpha_i y_i (x_i \cdot w + b) + \sum_{i=1}^{n} \alpha_i 
\]  

(5)

Where \(\alpha\) is the Langrangian coefficient, \(w\) is the weight at \(x\) and \(y\) coordinates, and \(b\) is the bias of \(x\) and \(y\) coordinates. While not linearly, SVM uses the kernel to classify data. The various kernel functions on SVM are as follows:
Linear Kernel

\[ K(x_i, x) = x_i^T x \] (6)

Polynomial Kernel

\[ K(x_i, x) = (\gamma x_i^T x + r)^p \] (7)

Radial Base Function Kernel (RBF)

\[ K(x_i, x) = \exp(-\gamma|x_i - x|^2) \] (8)

Sigmoid Kernel

\[ K(x_i, x) = \tanh(\gamma x_i^T x + r) \] (9)

There are 3 orders of kernel polynomials in this study that are 1st order polynomials, 2nd order polynomials, and 3rd order polynomials. If the order in the kernel polynomial become larger, more curved the resulting kernel is. SVM can be implemented in the classification of data over two classes called multi-class SVM. The Multi Class SVM is done by combining multiple binary SVMs or combining all the data consisting of several classes into an optimization problem form. There are two types of multi-class SVM namely one-against-all method and one-against-one method [8].

4. Experiments

The dataset used in this study came from Cohn-Kanade + Facial Expression Database. A database containing image sequences of eight facial expressions, namely normal expression, despair, happiness, sadness, anger, surprise, fear, and disgust. Data is taken using two synchronized Panasonic AG-7500 camera hardware. The data contained 123 subjects aged 18 to 50 years, 69% women, 81%, Euro-Americans, 13% African-Americans, and 6% other groups. Each data starts and ends with a neutral face. Image sequences have a duration of between 10 to 60 frames [9]. However, what will be used in this study is 10 frames in each image sequences and only recognizes expressions of happiness, sadness, anger, surprise, fear, and disgust. The dataset used will be divided into two types, namely 120 datasets for system training and 30 datasets for system testing.

![Figure 7. CK+ Database; (a) Angry; (b) Disgust; (c) Fear; (d) Happy; (e) Sadness; (f) Surprise [10, 11].](image-url)
In this study using normalization 64 x 64 in facial image after face detection using Viola-Jones algorithm. One example of a histogram on every facial expression generated as a result of feature extraction is as follows:

![Histograms of facial expressions](image)

Figure 8. Histogram of expression; (a) Angry; (b) Disgust; (c) Fear; (d) Happy; (e) Sadness; (f) Surprise.

Table 1 shows that the accuracy obtained for the entire face area 73.33% and from all experiments obtained the best results by using two eye area and one lip area with 80% accuracy using one-against-one SVM in the 2nd Polynomial kernel.

|                  | Linear | Polynomial 1 | Polynomial 2 | Polynomial 3 | RBF  |
|------------------|--------|--------------|--------------|--------------|------|
| Full Face        | 63.33  | 63.33        | 73.33        | 73.33        | 63.33|
| Eye and lip area | 70     | 70           | 80           | 80           | 63.33|
| Eye area         | 53.33  | 53.33        | 56.67        | 56.67        | 43.33|
| Lip area         | 70     | 70           | 70           | 66.67        | 66.67|

It can be seen in table 1 that the accuracy when using only eye and lip area is higher than using the entire face, this proves that the eye and lip are the most crucial part in identifying facial expression and the addition of the less important parts such as the nose and cheeks may decrease the accuracy. In table 2 the confusion matrix results obtained from this study using eye and lip area is as follows:
### Table 2. Confusion matrix.

|       | Angry | Disgust | Fear | Happy | Sad | Surprise |
|-------|-------|---------|------|-------|-----|----------|
| Angry | 2     | 1       | 0    | 1     | 1   | 0        |
| Disgust | 0   | 4       | 0    | 1     | 0   | 0        |
| Fear | 0     | 0       | 1    | 3     | 0   | 1        |
| Happy | 0     | 0       | 0    | 5     | 0   | 0        |
| Sad | 0     | 0       | 0    | 0     | 5   | 0        |
| Surprise | 0   | 0       | 0    | 0     | 0   | 5        |

From Table 2, it can be seen that out of 5 angry facial expressions, only 2 that are correctly classified, while the other is recognized as disgust, happy, and sad. After that out of 5 disgust facial expressions only 1 is misclassified as happy. Also from 5 fear facial expressions, 4 are misclassified as happy and surprise. This is due to an error in the Viola-Jones algorithm in detecting faces. Therefore, it is expected that in subsequent studies using manual area cutting.

### 5. Conclusion

This research can be used as a reference in doing further research. Analysis of system test results achieved 80% accuracy because there are facial expressions that have similarities that facial expression of fear and shock and then facial expressions angry and disgusted. Therefore, this research can be improved in the next research. The use of LGBP-TOP and SVM method that focuses on the facial image of the eyes and lips accuracy is higher than compared with feature extraction on all facial images. The accuracy using only the eye and lip is 80% and the accuracy using the entire face is 73,33%. In the next study the preprocessing of face recognition system may use manual cutting of the eye and lips area so that there is no information wasted or added from the eyes and lips compared with using the division of the face into blocks.

### References

[1] Xie L, Wei H, Yang W and Zhang K 2014 Video-based facial expression recognition using histogram sequ-ence of local gabor binary patterns from three orthogonal planes 2014 33rd Int. Chinese Control Conf IEEE p 4772–4776

[2] Su L. and Balazsi M Recognizing facial expressions in videos

[3] Almaev T R and Valstar M F 2013 Local gabor binary patterns from three orthogonal planes for automatic facial expression recognition 2013 Int. 2013 Humaine Association Conf. on Affective Computing and Intelligent Interaction IEEE p 356–361

[4] Medlej J Human anatomy fundamentals mastering facial expressions

[5] Rinaldi M 2014 Pengolahan Citra Digital dengan Pendekatan Algoritmitk (Bandung: Informatika)

[6] Priyanto H 2017 Pengolahan Citra Digital Teori dan Aplikasi Nyata. (Bandung: Informatika)

[7] Yanjun G 2014 Micro-expression recognition based on local binary patterns from three orthogonal planes and nearest neighbor method 2014 China IEEE

[8] Krisantus S 2007 Penerapan teknik support vector machine untuk pendeteksian intrusi pada jaringan 2007 (Bandung: Institut Teknologi Bandung)

[9] Lucey P, Cohn J F, Kanade T, Saragih J, Ambadar Z and Matthews I 2010 The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression In Computer Vision and Pattern Recognition Workshops (CVPRW) 2010 Computer Society Conf. IEEE p 94–101

[10] Kanade T, Cohn J F and Tian Y 2000 Comprehensive database for facial expression analysis 2000 Proc. of the Fourth IEEE Int. Conf. on Automatic Face and Gesture Recognition (FG’00)
Grenoble p 46-53

[11] Lucey P, Cohn J F, Kanade T, Saragih J, Ambadar Z and Matthews I 2010 The extended Cohn-Kanade Dataset (CK+): a complete expression dataset for action unit and emotion-specified expression Proc. of the Third Int. Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010) San Francisco p 94-101