MIXTURE FEATURE EXTRACTION BASED ON LOCAL BINARY PATTERN AND GREY-LEVEL CO-OCCURRENCE MATRIX TECHNIQUES FOR MOUTH EXPRESSION RECOGNITION

Ricardus Anggi Pramunendar[1], Dwi Puji Prabowo[2], Yuslena Sari[3]
[1]Department of Informatics Engineering, Universitas Dian Nuswantoro, Semarang, Indonesia
[2]Department of Visual Communication Design, Universitas Dian Nuswantoro, Semarang, Indonesia
[3]Department of Information Technology, Faculty of Engineering, Universitas Lambung Mangkurat, Banjarmasin, Indonesia
e-mail: ricardus.anggi@dsn.dinus.ac.id[1], dwi.puji.prabowo@dsn.dinus.ac.id[2], yuzlena@ulm.ac.id[3]

Corresponding Author: Ricardus Anggi Pramunendar

ABSTRACT

Some academics struggle to recognize facial emotions based on pattern recognition. In general, this recognition utilizes all facial features. However, this study was limited to identifying facial emotions in a single facial region. In this study, lips, one of the facial features that can reveal a person's expression, are utilized. Using a combination of local binary pattern feature extraction (LBP) and grey level co-occurrence matrix (GLCM) methods and a multiclass support vector machine classification approach for feature extraction in facial images. The concept begins with image segmentation to create an image of a mouth. Experiments were also conducted for various tests, and the outcomes of these experiments revealed a recognition performance of up to 95%. This result was obtained through experiments in which 10% to 40% of the data were evaluated. These findings are beneficial and can be applied to expression recognition in online learning media to monitor the audience's condition directly.

Keywords: *Expression recognition, LBP, GLCM, Multiclass SVM.*

I. INTRODUCTION

ACCORDING to research by Ekman et al. [1], a person's facial expression has multiple meanings. The six basic human facial expressions include happiness, surprise, anger, disgust, sadness, and fear. However, students' facial expressions based on learning activities or given when using learning media are classified into four categories: neutral, smiling, confused, and sleepy [2]. Several researchers employed computer vision technology in various ways to conduct these studies. Wu et al. [3] performed the analysis using the cascade classifier method based on the AdaBoost method. According to these studies, the same individual remains challenging to identify. Shekaina et al. [4] conducted a similar study in which they performed facial expression recognition using data from real-world scenarios. Using the Gabor wavelet feature extraction method and the decision tree classification method to classify two mouth expressions, this study achieved an accuracy of 90%. (Smiling and not smiling).

Mandalapu and Preeti [5] also use the wavelet method to detect yawning and normal mouths using the Haar-wavelet feature and the support vector machine (SVM) classification method. The results indicate a level of accuracy that is not excessively high: 81% for yawning mouths and 80% for normal mouths. The proposal is less effective at describing image objects with anisotropic elements or line-based characteristics. Due to the lack of non-geometry in the employed method, the regularity of the curved side cannot be exploited.

Several factors, including image preprocessing, influence facial expression research because different exposure levels can be created when photographing the same object. It can also be utilized as a feature of the classification algorithm. The technique used for feature extraction also affects the performance of facial expression recognition. It is less accurate to extract features using haar cascade, PCA, and Gabor Wavelet. Additional research is required to improve the success rate of identifying facial expressions. This research proposes a combination of the Local Binary Patterns (LBP) and Gray-Level Cooccurrence Metrics (GLCM) feature extraction techniques and the SVM classification technique. This method is based on the LBP method, which compares a pixel's relative intensity to its neighbouring pixels' intensity. Since LBP is insensitive to photometric variations of the same object, it is regarded as quite reasonable and resistant to numerous lighting disturbances in the image [6]. The GLCM method was selected because its quantization step aids in reducing noise, and statistical functions are used to determine the image's features [7].

In contrast, the SVM-based classification method is used because it seeks to find a separator function or hyperplane to divide the data set into two distinct classes, so it has a specific ability to distinguish between two images [8]. The contribution of this study is the analysis results derived from the combination of the LBP and GLCP feature extraction methods applied to the SVM classification method. In addition, the best parameters for this proposed combination model were also obtained.

This paper's structure follows: Prior research pertinent to this study's topic is presented in Section 2. In Section 3, the proposed model and experimental design are presented. Following that, Section 4 describes the observed outcomes and discussion. The overall study's conclusions are presented in Section 6.

II. RELATED RESEARCH

Some researchers have conducted facial expression recognition experiments, while others have used the mouth to capture the full spectrum of facial expressions. Wu et al. [3] classified several facial expressions, including interested, neutral, and exhausted, during the learning process. Some facial features were removed, including the eyebrows, forehead, eyes, and mouth. The classification method employing the Adaboost-based cascade classifier algorithm achieves an accuracy performance of 74.1 percent. May-Ping Loh et al. [2] classified neutral, sleepy, confused, and smiling facial expressions using the Gabor Wavelet method as a feature extraction method and a neural network as a classification method. May-Ping Loh et al. [2] derived a classification performance level of 83.75 % from this study.

Mandalapu and Preeti [5] conducted a study employing oral imagery using the Haar-wavelet-based feature extraction method and the SVM classification method. This study aimed to classify facial images displaying yawning and typical lips. This study found that 86% of participants had a standard mouth shape, and 81% had a yawning mouth. Based on the recommendation of Mandalapu and Preeti [5], segmentation is performed to obtain the image of the lips; therefore, image processing techniques are required to achieve optimal performance. Wahyuningrum et al. [9] also recommended employing a mouth image. Wahyuningrum et al. [9] were able to achieve a facial expression recognition performance of 82.67 % using the backpropagation neural network (BPNN) and feature extraction based on principal component analysis (PCA). This value is the average of recognizing five types of facial expressions: a sweet smile, a smile with a closed mouth, an open-mouth smile, a mocking smile, and a forced smile. The accuracy of the research conducted by Shekaina et al. [4] to detect a smile on the face was 90 percent. Applying the Gabor wavelet method and the decision tree classification method to obtain the unique characteristics of the mouth image yielded these results.

The error rate, which is still relatively high according to previous research, substantially affects facial expressions. Therefore, additional research is required to improve the success rate of facial expression recognition.

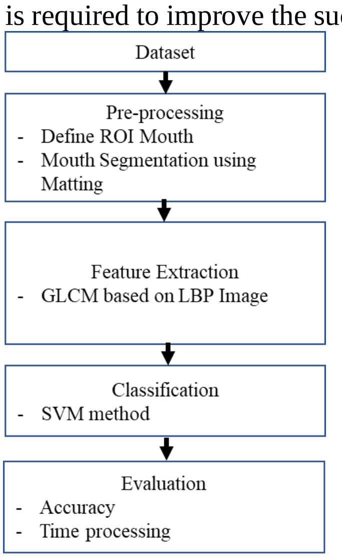

Figure 1. Proposed Method

III. PROPOSED RESEARCH

Figure 1 depicts the research proposal for the experimental research model utilized in this study. The proposal includes preprocessing or image data processing using the Matting method for image segmentation, followed by a method for feature extraction using GLCM based on LBP methods. The classification of three types of mouth models, namely open mouth, closed mouth, and smiling mouth, was then performed using the SVM method. The classification results are evaluated using the confusion matrix method to determine the level of precision.

A. Data collection

In this study, video recordings were used for the experiment, which was then extracted into (*.png) images containing the content of students engaging in online learning. The camera is positioned directly in front of the students in an enclosed room with standard lighting. The obtained data were extracted from the student's face template and sliced into various experimental aspects, with the mouth template used in this investigation. Figure 2 is a collection of images of mouths that will be utilized.

Figure 2. Mouth dataset

B. Data Processing

According to the roadmap in this study, previous researchers will preprocess the mouth feature image for the dataset; the mouth feature will be segmented using the Matting method and a threshold value, which is necessary given the poor image quality of the dataset. Images are captured under various lighting conditions; therefore, preprocessing is essential to address this issue. Tests were conducted on a total of 201 data based on comparing the percentage of testing and training. The training and testing data are separated into four categories: 10% testing and 90% training, 20% testing and 80% training, 30% testing and 70% training, and 40% testing and 60% training.

C. Feature Extraction

In this study, the feature extraction method comprises the local binary pattern (LBP) and grey-level co-occurrence matrix (GLCM) methods, which are combined for data classification.

1) Local Binary Pattern

This research's feature extraction phase involves determining the histogram value of mouth features. The binary matrix value will be searched using the LBP extraction function, which is used because it is not affected by photometric changes of the same object because it is a measure of the intensity relative to the strength of the surrounding pixels. LBP is computed by associating the bits of an integer with their nearest neighbours. If the intensity of an adjacent pixel exceeds that of the processed pixel, the bit position will be 1. Until a binary matrix is obtained, pixel values less than the threshold value will be represented as binary 0, and pixel values greater than or equal to the threshold value will be described as binary 1. Each neighboring pixel is connected by a portion of one bit, resulting in $2^8 = 255$ neighbour combinations for a 3×3 matrix (8 neighbours). The resultant binary value is written directly from right to left as a binary string. The operations applied to each resulting binary line are recombined to obtain the data's overall texture [10].

2) Gray-Level Co-occurrence Matrix

GLCM is a grey-level dependency matrix represented as a two-dimensional histogram of grey levels that are spatially separated. It is utilized to extract various texture features. The extraction of texture features was performed from four directions (0^o, 45^o, 90^o, and 135^o) separated by one pixel. After obtaining the pixel values' composition, the matrix's order is adjusted from row to column, and the values are then added to form a GLCM matrix before normalization. After obtaining the GLCM matrix, the results of pixel value composition can be used to determine the statistical properties of an image, such as second-order moment or energy (ene), entropy (ent), contrast (con), homogeneity (hom), and correlation (cor). A co-occurrence matrix can be used to retrieve several texture characteristics provided by GLCM.

$$ene = \sum_{k=0}^{L-1}\sum_{l=0}^{L-1}\{P(k,l)\}^2 \quad (1)$$
$$ent = -\sum_{k=0}^{G-1}\sum_{l=0}^{G-1} P(k,l) \times Log\big(P(k,l)\big) \quad (2)$$
$$con = \sum_{n=0}^{G-1} n^2\{\sum_{k=0}^{G}\sum_{l=0}^{G} P(k,l)\}\{n = |k-l| \quad (3)$$
$$hom = \sum_{k=0}^{G-1}\sum_{l=0}^{G-1} \frac{1}{1+(k-l)^2} P(k,l) \quad (4)$$
$$cor = \frac{\sum_{k=0}^{G-1}\sum_{l=0}^{G-1}(k,l)(P(k,l) - \mu_k{}'\mu_l{}'}{\sigma_k{}'\sigma_l{}'} \quad (5)$$

The correlation equation is met using (6)

$$\begin{aligned} P_x(k) &= \sum_{l=0}^{G-1} P(k,l) \\ P_y(l) &= \sum_{k=0}^{G-1} P(k,l) \\ \mu_k' &= \sum_{k=0}^{G-1}\sum_{l=0}^{G-1} k * P(k,l) \\ \mu_l' &= \sum_{k=0}^{G-1}\sum_{l=0}^{G-1} l * P(k,l) \\ \sigma_k' &= \sum_{k=0}^{G-1}\sum_{l=0}^{G-1} P(k,l)(k-\mu_l')^2 \\ \sigma_l' &= \sum_{k=0}^{G-1}\sum_{l=0}^{G-1} P(k,l)(l-\mu_l')^2 \end{aligned} \quad (6)$$

The distribution of co-occurrence values is denoted as k and l at the offset (1,1) given by p(k,l), with angles 0^o, 45^o, 90^o, and 135^o. Different angles are used to make rotation invariant, and the mean and standard deviation of orientation-dependent characteristics are determined independently for each angle.

D. *Classification*

Classification evaluates data objects to assign them to one of several accessible classes. The classification algorithm determines the relationship between the predicted and target values (training) when developing the model. Different classification algorithms employ distinct methods for locating connections. The model encapsulates these relationships, which can then be applied to various data sets when the class is unknown. The model's classification in the test data set is validated by comparing the predicted value to the known target value. Support vector machine (SVM) is the classification technique utilized in this study.

1) *Support Vector Machine*

Support Vector Machine (SVM) is a learning system that employs a linear function in a high-dimensional feature space, which is trained with an optimization-based learning algorithm by adopting a learning bias derived from statistical learning theory. SVM can only classify data into two categories (binary classification). Nevertheless, additional research has enabled SVM to classify data into more than two categories. The SVM method is a classification technique that seeks a separator function or hyperplane to partition the data set into two distinct groups.

E. *Evaluation*

The evaluation aims to describe the outcomes of the data under consideration. During the evaluation phase, the method's precision is assessed. The assessment is performed by dividing the total quantity of test data within the data set by the number of recognized test data. Using equation (7), accuracy is computed.

$$accuracy = \frac{known\ data}{total\ data} \times 100\% \quad (7)$$

IV. Analysis And Discussion

A. *Segmentation Step*

Image data extracted into multiple frames and cropped is used to analyze facial characteristics. Prior research employed the preprocessing step method, which centred on identifying specific oral characteristics. At this point, the mouth is trimmed to differentiate it from the other facial features depicted in Figure 3.

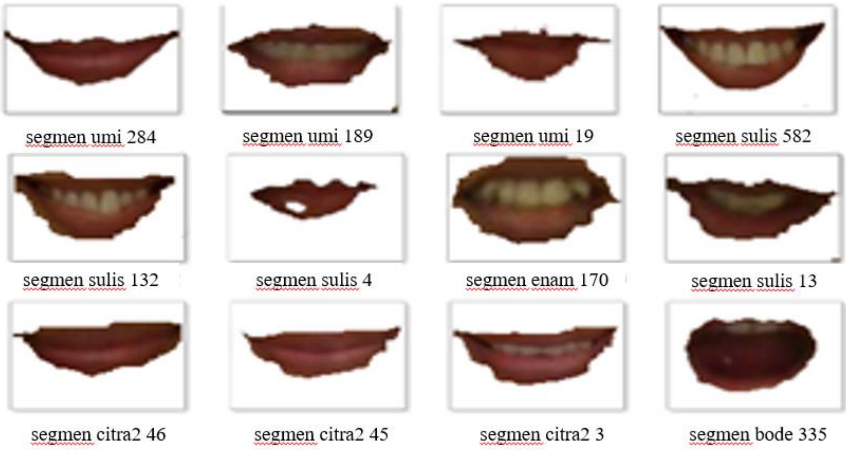

Figure 3. The mouth dataset that has been segmented

B. Feature extraction

Using multiclass SVM, we will classify some instances of data extracted using the LBP and GLCM extraction features (can be seen in Table 1). The results demonstrate a distinct difference in the value of the features generated by combining LBP and GLCM features.

TABLE I. FEATURE EXTRACTION

CITRA	CONTRAST	CORRELATION	ENERGY	HOMOGENEITY	TARGET
1	1,783	0.804	0.589	0.829	closed
2	1,743	0.801	0.607	0.843	closed
3	1,885	0.810	0.549	0.815	closed
4	1,661	0.804	0.601	0.847	closed
5	1,605	0.826	0.594	0.838	closed
…					
108	1,622	0.827	0.623	0.843	smile
109	1,708	0.809	0.636	0.846	smile
110	2,157	0.800	0.531	0.792	smile
111	2,093	0.804	0.541	0.796	smile
112	2,109	0.802	0.546	0.799	smile
154	3,828	0.748	0.312	0.663	Open
155	4,381	0.738	0.230	0.612	Open
156	4,112	0.760	0.223	0.608	Open
157	4,079	0.744	0.233	0.623	Open
158	4,123	0.744	0.224	0.616	Open

C. Model evaluation

The evaluation of the MATLAB application's functions is conducted to obtain the accuracy and average error values. In tests using 10:90, or 10% testing and 90% training, it has been demonstrated that 95% accuracy can be achieved with 72.16 seconds of training. In tests using 20:80, or 20% testing and 80% training, it has been demonstrated that 87.5% accuracy can be attained with a training duration of 128.56 seconds. In the test using 30:70, or 30% testing and 70% training, the given accuracy reaches 80.33 percent after 128.56 seconds of training. In tests using 40:60, or 40% testing and 60% training, the accuracy achieved is 75.31 percent with 233.27 seconds of training. Table 1 contains the complete list of test results.

TABLE II. PERFORMANCE MEASUREMENT RESULTS FROM MODEL

MEASUREMENT	PERSENTASE TESTING DATA			
	10 %	20 %	30 %	40 %
ACCURACY	95.00 %	87.50 %	80.33 %	75.31 %
ERROR RATE	5.00 %	12.50 %	19.67 %	24.69 %
TIME (SECOND)	59.7299	121.1405	194.0284	233.3775

After extracting oral features using a combination of LBP and GLCM feature extraction and classifying using the multiclass support vector machine method, it has been determined that the accuracy of the proposed method is 95%. This method that combines feature extraction with a high classification success rate can be proposed for continuous use in the recognition of facial expressions in learning. To recognize the facial expressions of the learner. It is possible that the level of accuracy provided by the classification method has not been maximized to provide a reasonable introduction limit. Still, this proposal allows it to be used in teaching and learning so that teachers can observe the learner's expression when receiving the material.

V. CONCLUSION

The digital image processing method uses the Local binary pattern feature extraction method and the Gray Level Co-occurrence Matrix, which is then classified using multiclass SVM to organize the characteristics of the learner's mouth expression so that it can be used as input for the application of the Adaptive eLearning System during the industrial revolution. 4.0. It is possible to classify the characteristics of mouth shapes with a performance of 95% accuracy using 90% training data and 10% test data by combining the feature extraction techniques of Local Binary Pattern and Gray Level Co-occurrence Matrix, which are both inputs to the SVM multiclass classification method. The next phase of research will focus on improving recognition performance without requiring massive datasets.

REFERENCES

[1] P. Ekman, P. Ellsworth, and W. V. Friesen, *Emotion in the Human Face*. Elsevier, 1972. doi: 10.1016/C2013-0-02458-9.

[2] May-Ping Loh, Ya-Ping Wong, and Chee-Onn Wong, "Facial Expression Recognition for E-learning Systems using Gabor Wavelet Neural Network," in *Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06)*, pp. 523–525. doi: 10.1109/ICALT.2006.1652491.

[3] Y. Wu, T. Wang, and X. Chu, "Affective Modeling and Recognition of Learning Emotion: Application to E-learning," *J. Softw.*, vol. 4, no. 8, Oct. 2009, doi: 10.4304/jsw.4.8.859-866.

[4] J. Shekaina and P. J L, "Multiview Smile Detection by Gabor Wavelet Decision Tree Classifier," *Int. J. Comput. Appl.*, vol. 71, no. 6, pp. 31–35, Jun. 2013, doi: 10.5120/12363-8683.

[5] S. D. Mandalapu and B. Preeti, "Driver Fatigue Detection Using Mouth and Yawning Analysis," *Coll. Enginerring*, 2008.

[6] Y. Cai, G. Xu, A. Li, and X. Wang, "A Novel Improved Local Binary Pattern and Its Application to the Fault Diagnosis of Diesel Engine," *Shock Vib.*, vol. 2020, pp. 1–15, Feb. 2020, doi: 10.1155/2020/9830162.

[7] R. A. Pramunendar, D. P. Prabowo, D. Pergiwati, Y. Sari, P. N. Andono, and M. A. Soeleman, "New workflow for marine fish classification based on combination features and CLAHE enhancement technique," *Int. J. Intell. Eng. Syst.*, vol. 13, no. 4, pp. 293–304, 2020, doi: 10.22266/IJIES2020.0831.26.

[8] M. Senekane, "Differentially Private Image Classification Using Support Vector Machine and Differential Privacy," *Mach. Learn. Knowl. Extr.*, vol. 1, no. 1, pp. 483–491, Feb. 2019, doi: 10.3390/make1010029.

[9] R. T. Wahyuningrum, R. M. Wati, and A. Rachmad, “Pengenalan Pola Senyum Menggunakan Backpropagation Berbasis Ekstraksi Fitur Principal Component Analysis (PCA),” *Rekayasa*, vol. 4, no. 1, 2005, doi: 10.21107/rekayasa.v4i1.2330.

[10] B. Kaddar, H. Fizazi, and A.-O. Boudraa, "Texture features based on an efficient local binary pattern descriptor," *Comput. Electr. Eng.*, vol. 70, pp. 496–508, Aug. 2018, doi: 10.1016/j.compeleceng.2017.08.009.