*Research Article*

# Facial Expression Recognition Using a Novel Modeling of Combined Gray Local Binary Pattern

**An H. Ton-That** [1] **and Nhan T. Cao** [2,3]

*[1]Information Technology Faculty, University of Finance and Marketing, Address: 778 Nguyen Kiem Street, Phu Nhuan District, Ho Chi Minh City 700000, Vietnam*
*[2]University of Information Technology, Ho Chi Minh City, Vietnam*
*[3]Vietnam National University, Ho Chi Minh City 700000, Vietnam*

Correspondence should be addressed to Nhan T. Cao; nhanct@uit.edu.vn

Facial Expression Recognition (FER) is an active research field at present. Deep learning is a good method that is widely used in this field but it has extreme hardware requirements and it is hard to apply in normal terminal devices. So, many other methods are being researched to apply FER in such devices and systems. This work proposes fresh modeling of Combined Gray Local Binary Pattern (CGLBP) for extracting features in facial expression recognition to enhance the recognition rate that can apply FER in the kind of devices and systems. The work included the main steps such as the technique of cropping an input face image from a camera or dataset, the approach of dividing face images into nonoverlap regions for extracting LBP features, applying the fresh modeling of Combined Gray Local Binary Pattern (CGLBP) for extracting features, using uniform feature to reduce the lengths of descriptors, and finally using Support Vector Machine (SVM) for emotion classification. Four popular facial emotion datasets are used in experiments and their results demonstrate that the recognition rate of the proposed method is better in comparison with two types of existent features: Local Binary Pattern (LBP) and Combined Local Binary Pattern (CLBP). The accuracy of experiments performed on four facial expression datasets with different sizes is from about 95% to more than 99%.

## 1. Introduction

Facial expressions are facial changes in response to a person's internal emotional states, intentions, or social communication as in [1]. Facial expression recognition is an active research field that is defined as extracting the emotional state of a person from his or her facial expression. FER is particularly useful for enhancing naturalness in human-machine interaction robotics and contributes to intelligent human-machine interface systems that can be applied in behavioral science, clinical practice, and day-to-day life. For applying, many FER systems have been built and three state-of-the-art FER systems among them as in [2] show that the systems classify basic emotions accurately and they are mostly on par with human raters for standardized images.

Deep learning is a good method that is widely used in this field and they overcome the limitations of conventional approaches, but it requires excessive hardware specifications and so the application of the deep learning method is rather limited, especially applied in real-time systems or normal terminal devices. For that reason, other methods without excessive hardware specification requirements have been still proposed for facial expression recognition. For example, [3] proposed a method of using a combination of Local Binary Pattern (LBP) features and Histogram Oriented Gradient (HOG) features to create the feature vector and after that, applying K-NN for classification. And Ref. [4] presented an algorithm with the combination of the oriented FAST and rotated BRIEF (ORB) features and LBP features extracted from facial expressions and using a Support Vector Machine classifier.

There are commonly two main methods to enhance the accuracy of an automatic facial expression recognition system. The first is discovering suitable features that more efficiently characterize different facial expressions. The second is finding classifiers that better classify facial expressions. In addition, determining a method of face image preprocess for extracting the feature that is most appropriate for the kind of features and classifiers used is very useful for improving the recognition rate.

To select the kind of feature for FER, there are commonly two main approaches that are geometric features and appearance features. Geometric features focus on the shape and locations of facial components, which are extracted to represent the face geometry. Whereas appearance features show the appearance changes (skin texture) of the face, which are extracted by applying image filters to either the whole face or specific facial regions. In facial expression modeling, appearance features are often chosen because they are good at initialization, tracking, and encoding changes in skin texture.

There have been many local image descriptors developed to analyze image texture, such as Local Binary Pattern (LBP) [5, 6], Completed Local Binary Pattern (CLBP) [7], Local Directional Gradient Pattern (LDGP) [8], Local Gradient Hexa Pattern (LGHP) [9], Rotation and Scale-invariant Hybrid Descriptor (RSHD) [10], Local Quadruple Pattern (LQPAT) [11], Center Symmetric Local Binary Pattern (CSLBP) [12], Bag–Of–Filters Local Binary Pattern (BoF-LBP) [13], R-Theta Local Neighborhood Pattern (RTLNP) [14], Improved Local Ternary Patterns (ILTP) [15], and Multi Threshold Uniform-Based Local Ternary Patterns (MT-ULTP) [16].

The descriptor researchers try to meet requirements for capturing texture information and reducing the length of the descriptor. Moreover, depending on the specific research domain, different image descriptor is applied. For example, LBP and CLBP are used in facial recognition and facial emotion recognition [3, 4, 17, 18]; LDGP, LGHP, LQPAT, CSLBP, CSLBP, BoF-LBP, and RTLNP are used in facial image recognition and retrieval [8, 9, 11–14]; RSHD is used in color image recognition and retrieval [10]; ILTP is used in bark texture classification [15]; or MT-ULTP is used in cell phenotype classification [16].

This work proposes novel modeling of features for facial expression recognition called the Combined Gray Local Binary Pattern (CGLBP). The feature combines the Local Binary Pattern (LBP) and Local Gray Level Difference (LGLD) of face images. This feature is extracted based on an effective method of face acquisition employed as in [17]. For emotion classification, a Support Vector Machine (SVM) has been used.

## 2. Previous Related Works

### 2.1. Face Acquisition.
The face acquisition process commonly has two main steps the basic step and the enhancement step. The basic step usually uses a face detector to discover the face region of an input face image from a camera or dataset and then eliminate its redundant regions.

The enhancement step aims to improve the face region for extracting facial expression features with cropping methods, image normalization methods, or image filter techniques. After that, the face images are rescaled and used for feature extraction. The preprocessing of face acquisition is shown in Figure 1.

The face detector in the basic processing method that is used much in literature is the robust real-time face detector [19] proposed by Viola and Jones. This detector carries out a fast face recognition method based on Haar-like features and a cascade of classifiers. Experimental results approved that this method obtained good performance including both high accuracy and fast calculation. However, to attain better performance of facial emotion recognition, facial images obtained from the robust real-time face detector can eliminate redundant image regions by geometrically standardized or cropping methods before being used for extracting features. This normalization method is usually performed based on eye locations, lip states, nasolabial furrows, or crows-feet wrinkles, as in [20, 21].

For enhancement preprocessing, Reference [22] proposed a cropping method with 2 steps: first detect face landmarks and then remove the forehead region based on the horizontal distance between the eye centers as shown in Figure 2(a).

This work uses an enhancing processing technique presented in [17]. The face images obtained from the robust real-time face detector are cut by the cropped technique to get rid of unnecessary information or pixels with little information as in Figure 2(b).

Observations of the human face usually show that the forehead part of the human face height accounts for about a quarter of it and does not hold much essential facial emotion information. Therefore, to improve the accuracy rates of FER, two-thirds of the upper forehead part will be cut off and one-third of the lower forehead part close to the eyebrows is retained.

The objective of the cropping method is to take a smaller face image region but a higher rate of essential information about emotion. So, the method not only reduces the processing time in the feature extraction step and recognition step but also enhances the accuracy of facial expression recognition and saves computational resources.

Figure 3 illustrates the steps for face acquisition: Figure 3(a) is a face image from a dataset of emotional face images; Figure 3(b) shows a part of the face image (in red squares) obtained from 3(a) by using the robust real-time face detector; Figure 3(c) presents a part of the face image (in yellow square or small square) obtained from 3(b) by using the cropping method; and Figure 3(d) is the face image obtained from the process of face acquisition (or from 3(c)) for extracting features.

### 2.2. Face Feature Extraction.
Commonly, facial representation consists of two ways: geometric features as in [23], and appearance features as in [17]. First, this work divides the face image into nonoverlap regions and then extracts LBP features as in Figure 4 presented in [17].
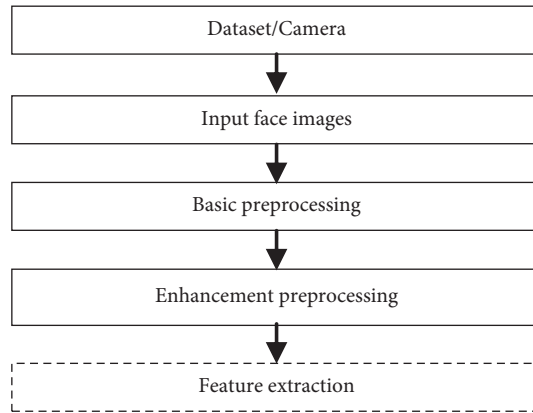
FIGURE 1: The process of face acquisition preprocessing.



(a)

(b)
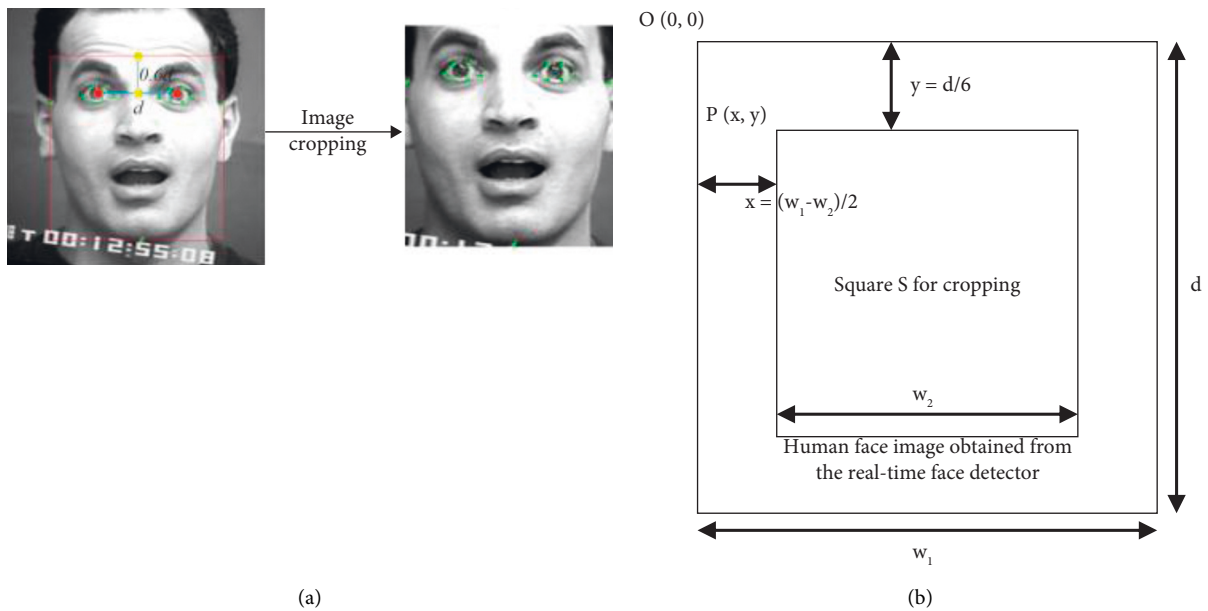
FIGURE 2: (a) A cropping method in [22]. (b) The cropped technique of face region in [17].



(a) (b) (c) (d)
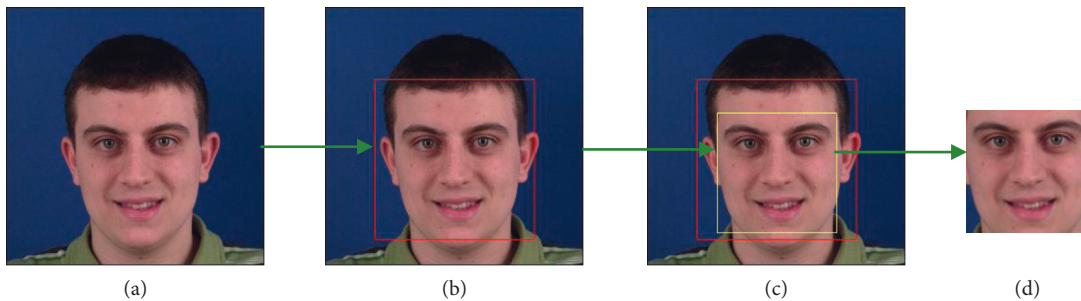
FIGURE 3: The process of face acquisition.

Then LBP histogram (or LBP feature) is calculated for each region. Finally, the feature vector of the face image is formed by concatenating the LBP histogram of each region from left to right and up to down as shown in Figure 5.

## 3. The Feature Modeling of Novel Combined Local Binary Pattern

*3.1. Local Binary Pattern (LBP).* The LBP operator was recommended in [5, 6] and used as a complementary

FIGURE 4: A face image with nonoverlap regions.



FIGURE 5: LBP feature vector of one facial image.
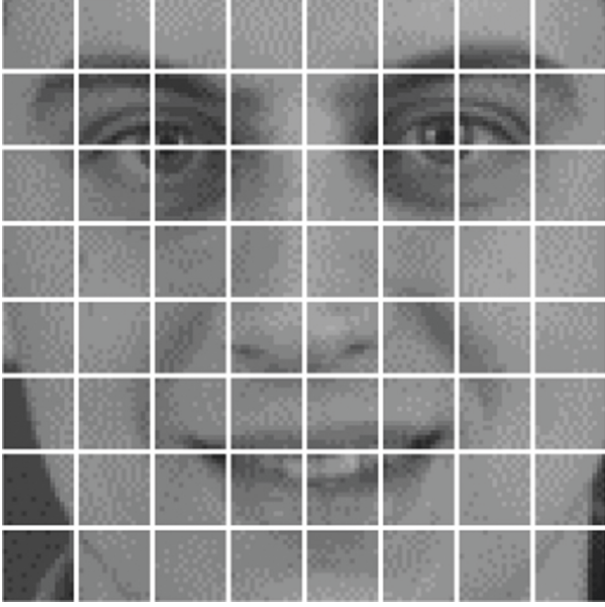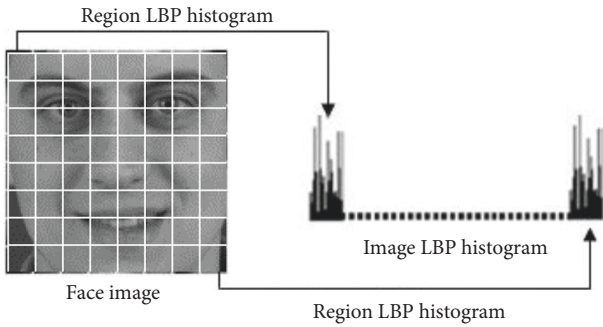
measure for local image contrast. The operator uses the value of a pixel and the circular eight neighbors of it to calculate the LBP code of the pixel. If the value of a pixel of a neighbor is greater than or equal to the value of the center pixel, it is labeled 1, otherwise is 0. The result (the eight pixels from the pixel in the upper left clockwise) is an 8-digit binary number of the pixel (also called LBP code). Then, the histogram is calculated using a decimal number converted from the binary number. Based on the operator, each pixel of the image is labeled by one LBP code value from 0 to 255 depending on its $3 \times 3$ neighborhood values. An example of the calculation of binary numbers and decimal numbers using the basic LBP operator is demonstrated in Figure 6.

To capture the dominant features of large-scale structures of an image better, the operator was extended to use neighborhoods of different sizes as in [6]. With circular neighborhoods and bilinear interpolation, the pixel values allow any radius and number of pixels in the neighborhood. The notation (P, R) represents a neighborhood of P equally spaced sampling points on a circle of radius of R to form a

circular and symmetric neighbor set. Figure 7 shows examples of the extended LBP operators.

Generally, to extract facial expression features, the basic LBP operator is used instead of the extended LBP operator for the following reasons. The first reason is although the extended LBP operator can capture dominant information with large-scale textures of an image, it is time-consuming to process in comparison to basic LBP. The second reason is in facial expression recognition, facial images are usually normalized to be not large in size (such as $36 \times 48$, $48 \times 48$, $55 \times 75$, $64 \times 64$, $110 \times 150$ pixels) then the face images are split into smaller regions. It means that facial emotion is analyzed based on changes in these smaller regions (as shown in Figure 4).

The 256-bin histograms of LBP features that are called the non-uniform pattern operators usually take much time for processing. So, a uniform LBP pattern operator was proposed in [6]. A Local Binary Pattern is uniform if it comprises at most two bitwise transitions from 0 to 1 or vice versa when the binary string is considered circular. Some examples of uniform patterns are 11111111, 00001111, 01111000, 11000001, and 00011110. Experimental results of uniform LBP operator in [6] showed that in texture images, uniform LBP patterns account for 87.2% of all patterns in the (8, (1) neighborhood and for 66.9% of all patterns in the (16, (2) neighborhood.

The notation symbolizes a uniform LBP operator. The superscript *u2* means the uniform pattern with the number of transitions; the subscript describes the operator using a (P, R) neighborhood. Uniform LBP uses only uniform patterns and labels all remaining patterns with a single label. A local binary pattern code is computed for a pixel in an image by comparing it with its neighbors as in equation (1):

$$LBP_{P,R} = \sum_{p=0}^{P-1} s\left(g_p - g_c\right)2^p, \ s(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0, \end{cases} \quad (1)$$

where $g_c$ is the gray value of the central pixel, $g_p$ is the gray value of its neighbors, $P$ is the total number of involved neighbors, and $R$ is the radius of the neighborhood. A histogram of a labeled image $f_k(x, y)$ can be defined as in equation (2):

$$H_i = \sum_{x,y} I\left(f_k(x, y)\right) = i, \quad i = 0, \ldots, n - 1, \quad (2)$$

where $n$ is the number of different labels produced by the LBP operator and

$$I(A) = \begin{cases} 1, & A \text{ is true,} \\ 0, & A \text{ is false.} \end{cases} \quad (3)$$

This histogram comprises information about the distribution of the local micropatterns, e. g., corners, edges, spots, or flat areas, over the whole image. To represent the face efficiently, the features extracted should retain spatial information. For this reason, the face image can be divided into $m$ small regions $R_0, R_1 \ldots, R_m$ as shown in Figure 5, and a spatially enhanced histogram is expressed as
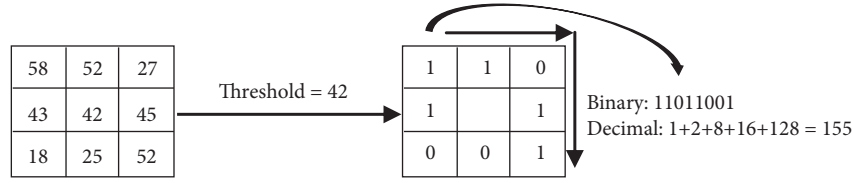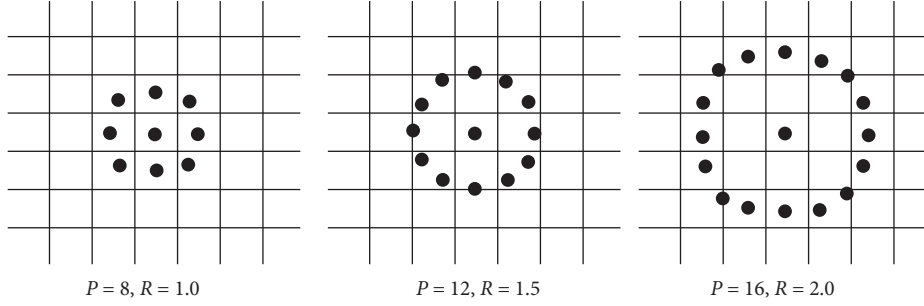
FIGURE 6: The basic LBP operator.



$P = 8, R = 1.0$                   $P = 12, R = 1.5$                   $P = 16, R = 2.0$

FIGURE 7: Example of the extended LBP.

$$H_{i,j} = \sum_{x,y} I\left(f_k(x, y) = i\right) I\left((x, y) \in R_j\right), \tag{4}$$

where $i = 0\ldots, n\text{-}1, j = 0\ldots, m\text{-}1$.

### 3.2. Completed Local Binary Pattern.

The local binary pattern has been used as an effective feature for facial expression recognition. However, the local binary pattern only uses the conventional sign components and ignores the magnitude component. However, it is also discovered that the information contained in magnitude difference or local gray level difference can provide a significant performance improvement.

As said by [7], the difference between $g_c$ and $g_p$ can be computed as $d_p = g_p - g_c$, where $g_c$ is the central pixel and its $P$ circularly and evenly spaced neighbors $g_p$, $P = 0, 1, \ldots, P\text{-}1$.

The local difference vector $[d_0\ldots, d_{P\text{-}1}]$ describes the image local structure of $g_c$ and can be decomposed into two components as in equation (5):

$$d_p = s_p * m_p \text{with} \begin{cases} s_p = \text{sign}(d_p), \\ m_p = |d_p|, \end{cases} \tag{5}$$

where the sign of $d_p$ and $m_p$ is the magnitude of $d_p$. Equation (5) is called the local difference sign-magnitude transform and it transforms the local difference vector $[d_0\ldots, d_{P-1}]$ into a sign vector $[s_0\ldots, s_{P-1}]$ and a magnitude vector $[m_0\ldots, m_{P-1}]$.

The completed local binary pattern includes the CLBP_S ($s_p$) operator and the CLBP_M ($m_p$) operator with the same format. To create a CLBP descriptor, the histograms of CLBP_S and CLBP_M can be combined by using one of the following two methods: concatenation or jointly. In the first method, the CLBP_S histogram and the CLBP_M histogram are calculated separately, then concatenated as one histogram. This CLBP descriptor is denoted by CLBP_S_M. In the second method, a joint 2D histogram of the CLBP_S and CLBP_M are calculated. This CLBP descriptor is denoted by CLBP_S/M. The feature has been researched for facial expression recognition and presented in [18].

### 3.3. The Novel Modeling of Combined Local Binary Pattern.

The approach using CLBP_S_M in Section 3.2. takes much time to calculate $m_p = |d_p|$. In this work, an improved approach to modeling the completed local binary pattern is proposed, called Combined Gray Local Binary Pattern (CGLBP or CLBP_S_G). This modeling is based on the CLBP_S operator and images local gray level difference called CLBP_G for brief. Figure 8 shows an example of the transformation of CLBP_S and CLBP_G on a $3 \times 3$ sample block.

The sign component of CLBP_S_G is the same as the sign component of CLBP_S_M, where the CLBP_G operator is defined as in equation (6):

$$\text{CLBP\_G}_{P,R} = \sum_{p=0}^{P-1} t(g_p, c) 2^P, \ t(x, c) = \begin{cases} 1, & x \geq c, \\ 0, & x < c, \end{cases} \tag{6}$$

where the threshold $c$ is to be determined adaptively. Here, $c$ is taken to be the mean value of $g_p$ over the CLBP_G operator.

Based on the idea of choosing an effective threshold $c$ for CLBP_G in facial expression recognition, we tested the following different thresholds:

The mean value of $g_p$ of the whole face image

The mean value of $g_p$ of the region

The mean value of $g_p$ of the CLBP_G operator

Experimental results on four datasets indicated that the threshold in the last case obtains the best accuracy in facial expression recognition. It can be explained as follows:

| 25 | 48 | 76 |
|----|----|----|
| 19 | 32 | 41 |
| 36 | 87 | 9 |

| -7 | 16 | 44 |
|----|----|----|
| -13 | | 9 |
| 4 | 55 | -23 |

| 0 | 1 | 1 |
|----|----|----|
| 0 | | 1 |
| 1 | 1 | 0 |

| 25 | 48 | 76 |
|----|----|----|
| 19 | | 41 |
| 36 | 87 | 9 |

(a)    (b)    (c)    (d)

FIGURE 8: The transformation of sign and magnitude components of the improved modeling of CLBP: (a) A $3 \times 3$ sample block, (b) local difference, (c) sign component, and (d) gray level component.

One main weakness of the LBP operator is that it is very sensitive to noise. The label of patterns will be changed even if only one noise pixel occurred.

Facial emotions mainly are expressed in some specific areas of the face (such as eyes, mouth, and eyebrows) as shown in divided regions in Figure 4.

Thus, choosing threshold is the mean value of $g_p$ of the CLBP_G operator that can not only make the CLBP_S_G operator more tolerant of noise but also combines the difference of magnitude component.

Both the CLBP_S operator and CLBP_G operator have the same binary string format, so they can be used to create the CLBP_S_G or CLBP_S/G descriptors by concatenation or jointly combination method respectively.

## 4. Experiments and Results

*4.1. Experimental Process.* The work experimented on four typical image datasets based on the proposed approach. The first is the Japanese Female Facial Expression (JAFFE) dataset with 213 images, the second is Cohn-Kanade+ (CK+) dataset having 2,040 images, the third is the FG-NET database (or FEEDTUM) containing 2,268 images, and the last is MUG dataset consisting of 5,132 images. Cropped face images are normalized to the size of 64x64 pixels, and then they are divided into regions of 8x8 pixels. To evaluate the accuracy of the proposed method, the 3-fold cross-validation technique is used. Figure 9 illustrates the experimental process.

*4.2. The Japanese Female Facial Expression (JAFFE) Dataset.* The 213 gray images of ten Japanese female facial expressions are contained in the JAFFE dataset as in [24] or [25] in which each person represents seven different facial expressions including anger, disgust, fear, joy, neutral, sadness, and surprise. Most facial expressions of each subject have 3 different images. In particular, three cases have two images, and six cases have four images. Original images have a resolution of $256 \times 256$ pixels and grayscale values. In this work, all 213 images are selected as experiment samples. A few examples of facial expression images from the JAFFE dataset could not be shown in this paper because of complying with the conditions for using the dataset as in [26].

In Figure 10 shows that the accuracy rates of non cropped images (full images acquired from the robust real-time face detection algorithm or percentage 100% of $w_2/w_1$) for all three kinds of features (LBP, CLBP_S_M, CLBP_S_G) are almost less than the accuracy rates of cropped images (from 80% to 90% of $w_2/w_1$) on JAFFE dataset. The CLBP_S_G feature obtains the highest accuracy at 97.21% with a cropping percentage of $w_2$ to $w_1$ being 86% and almost better than LBP and CLBP_S_M features in the remaining percentages of $w_2$ to $w_1$.

Table 1 shows the confusion matrix of the JAFFE dataset based on the CLBP_S_G feature with the cropping percentage of w2 to w1 being 86%.

*4.3. The Cohn-Kanade (CK+) Dataset.* The Cohn-Kanade dataset is one of the most comprehensive datasets in the current facial expression research community as in [27]. The CK + dataset comprises a variety of subjects: 100 university students with an age range between 18 and 30, 65% are female, 15% are African-American, and 3% are Asian or Latino. Subjects were instructed to perform a series of 23 facial displays, six of which were based on the description of basic emotions consisting of anger, disgust, fear, joy, sadness, and surprise. Image sequences expressed from neutral to strong emotion have a resolution of $640 \times 490$ pixels and 8-bit grayscale values. A few facial expression images (character S55) from the CK + dataset are shown in Figure 11.

In Figure 12 shows an anger emotion image (character S52) of the CK + dataset and its CLBP_S_G feature histogram.

Figure 13 shows that the accuracy rates of non-cropped images for all three kinds of features (LBP, CLBP_S_M, CLBP_S_G) are almost less than the accuracy rates of cropped images (from 80% to 90% of w2/w1) on CK + dataset. The CLBP_S_G feature obtains the highest accuracy at 99.95% with the cropping percentage of w2 to w1 being 85% and almost better than LBP and CLBP_S_M features in the remaining percentages of w2 to w1.

Table 2 shows the confusion matrix of the CK + dataset based on the CLBP_S_G feature with the cropping percentage of w2 to w1 being 85%.

*4.4. The FEEDTUM Dataset.* The FEEDTUM or FG-NET dataset in [28]. Facial expressions and emotions from the Technical University Munich is an image dataset containing face images presenting a number of subjects performing the
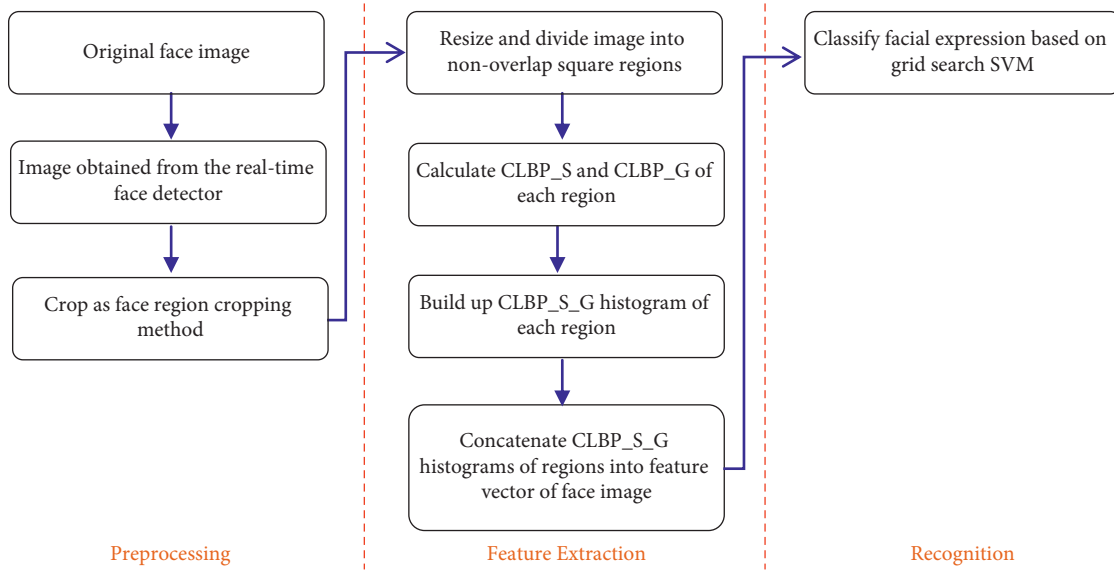
Figure 9: Experimental process.



| | 80% | 81% | 82% | 83% | 84% | 85% | 86% | 87% | 88% | 89% | 90% | 100% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LBP | 95.32 | 95.78 | 94.85 | 95.79 | 96.25 | 95.32 | 96.27 | 95.79 | 94.85 | 95.79 | 95.79 | 95.8 |
| CLBP_S_M | 96.25 | 95.78 | 95.78 | 96.25 | 95.79 | 96.27 | 95.8 | 96.72 | 95.33 | 95.78 | 95.78 | 95.34 |
| CLBP_S_G | 96.72 | 96.72 | 96.26 | 96.73 | 96.73 | 96.74 | 97.21 | 97.19 | 95.79 | 96.72 | 96.72 | 96.75 |

Figure 10: Comparison of experimental results of LBP, CLBP_S_M, and CLBP_S_G for JAFFE dataset.

Table 1: The confusion matrix of the JAFFE dataset.

| (%) | Anger | Disgust | Fear | Joy | Neutral | Sadness | Surprise |
|---|---|---|---|---|---|---|---|
| Anger | **96.67** | 3.33 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Disgust | 3.33 | **96.67** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Fear | 0.00 | 0.00 | **96.97** | 0.00 | 0.00 | 0.00 | 3.03 |
| Joy | 0.00 | 0.00 | 0.00 | **100** | 0.00 | 0.00 | 0.00 |
| Neutral | 0.00 | 0.00 | 0.00 | 0.00 | **100** | 0.00 | 0.00 |
| Sadness | 0.00 | 0.00 | 3.33 | 0.00 | 3.34 | **93.33** | 0.00 |
| Surprise | 0.00 | 0.00 | 0.00 | 3.33 | 0.00 | 0.00 | **96.67** |

FIGURE 11: A few examples of facial expression images of character S55 in the CK + dataset [27].



FIGURE 12: An anger emotion image of character S52 in CK + dataset and its CLBP_S_G feature histogram [27].

LBP vs. CLBP_S_M vs. CLBP_S_G -CK+ Dataset



|       | 80%   | 81%   | 82%   | 83%   | 84%   | 85%   | 86%   | 87%   | 88%   | 89%   | 90%   | 100%  |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| LBP       | 99.71 | 99.85 | 99.75 | 99.75 | 99.75 | 99.8  | 99.85 | 99.75 | 99.75 | 99.66 | 99.71 | 99.61 |
| CLBP_S_M  | 99.8  | 99.8  | 99.85 | 99.85 | 99.85 | 99.8  | 99.85 | 99.75 | 99.85 | 99.75 | 99.75 | 99.41 |
| CLBP_S_G  | 99.8  | 99.9  | 99.8  | 99.8  | 99.9  | 99.95 | 99.8  | 99.8  | 99.85 | 99.8  | 99.75 | 99.66 |

```
....... LBP
- - -   CLBP_S_M
─────   CLBP_S_G
```
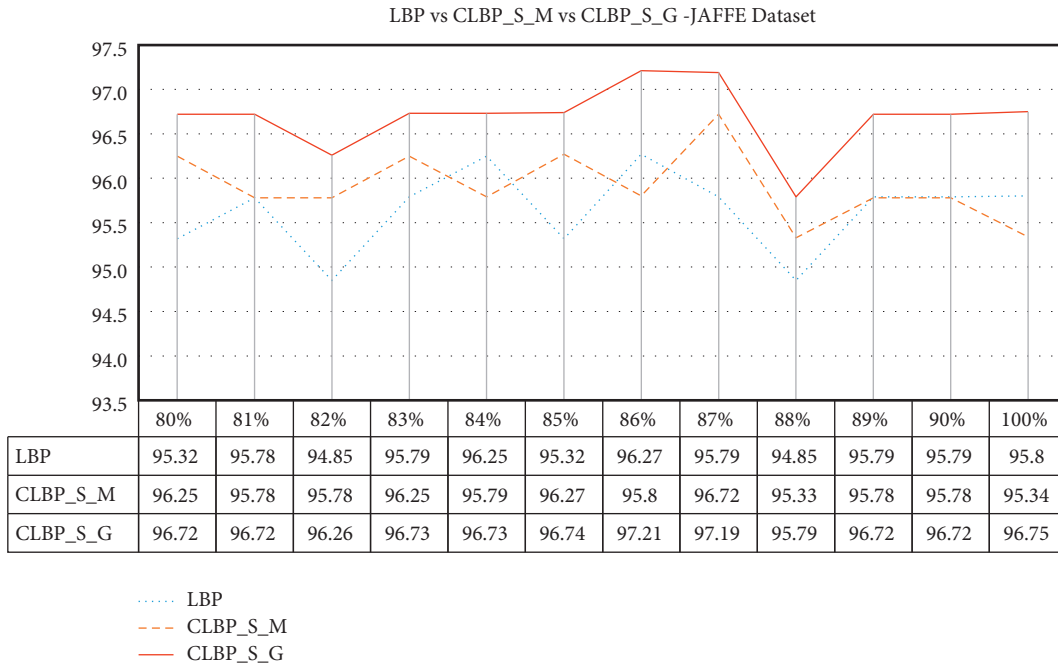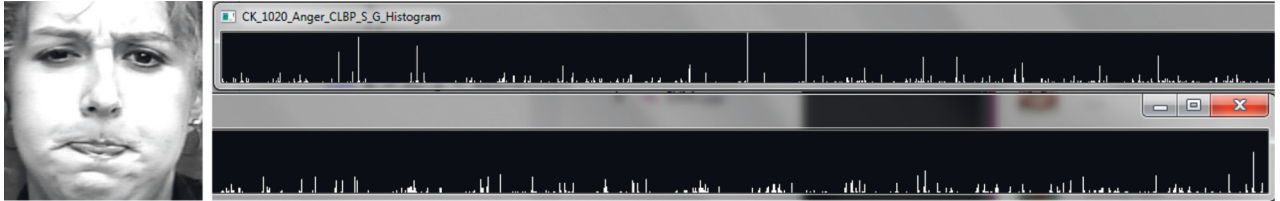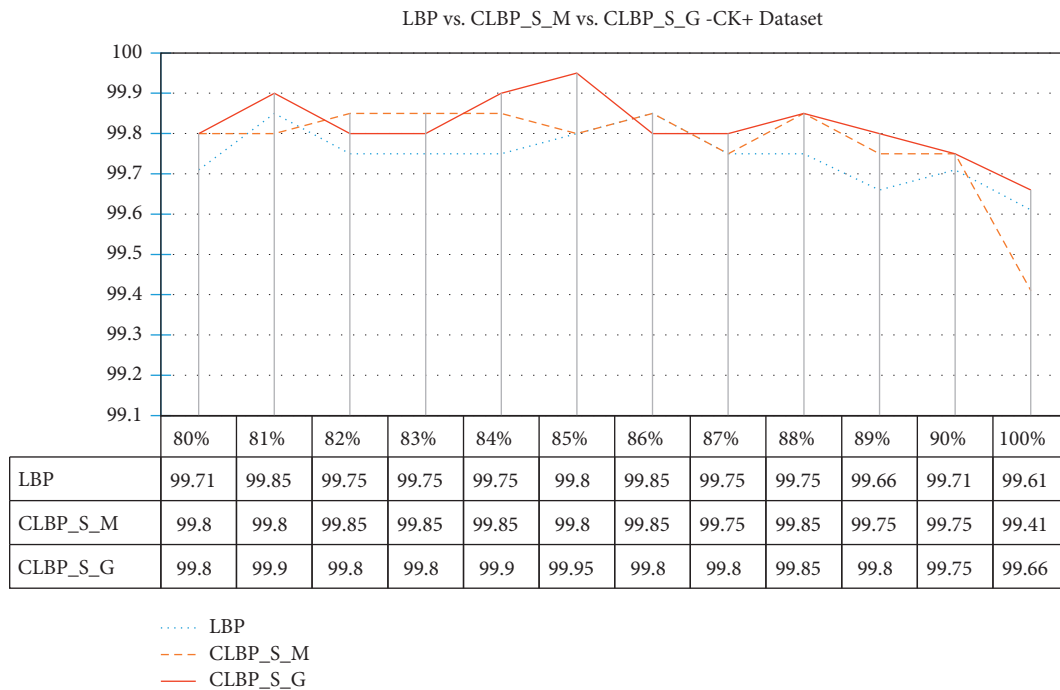
FIGURE 13: Comparison of experimental results of LBP, CLBP_S_M, and CLBP_S_G for CK + dataset.

TABLE 2: The confusion matrix of CK + dataset.

| (%)      | Anger | Disgust | Fear | Joy  | Neutral | Sadness | Surprise |
|----------|-------|---------|------|------|---------|---------|----------|
| Anger    | **100** | 0.00  | 0.00 | 0.00 | 0.00    | 0.00    | 0.00     |
| Disgust  | 0.00  | **100** | 0.00 | 0.00 | 0.00    | 0.00    | 0.00     |
| Fear     | 0.00  | 0.00    | **100** | 0.00 | 0.00 | 0.00    | 0.00     |
| Joy      | 0.00  | 0.00    | 0.00 | **100** | 0.00 | 0.00    | 0.00     |
| Neutral  | 0.00  | 0.00    | 0.00 | 0.00 | **100** | 0.00    | 0.00     |
| Sadness  | 0.00  | 0.00    | 0.00 | 0.00 | 0.00    | **100** | 0.00     |
| Surprise | 0.00  | 0.00    | 0.00 | 0.00 | 0.28    | 0.00    | **99.72** |

six different basic emotions defined by Eckman and Friesen as in [29]. The dataset contains material gathered from 18 different individuals. Each individual performed all seven expressions three times. Sequences are starting from the neutral state passing into the emotional state. The images were saved in 8-bit JPEG format with a resolution of 320 × 240 pixels.

This dataset has many challenges such as the emotional expressions of many subjects being weak and highly confused with other expressions, or there are a lot of subjects

FIGURE 14: A few examples of facial expression images of character 0002 in the FEEDTUM dataset [28].
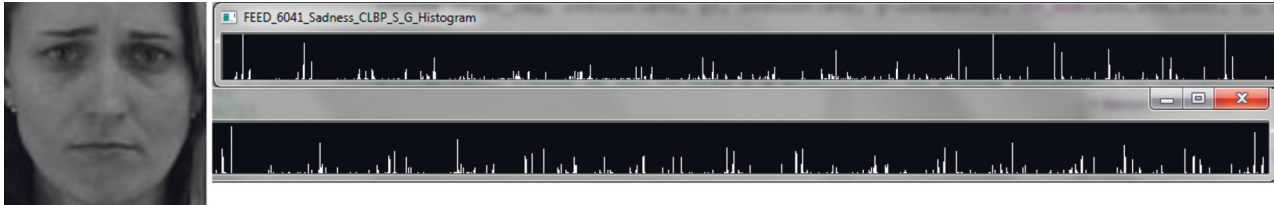


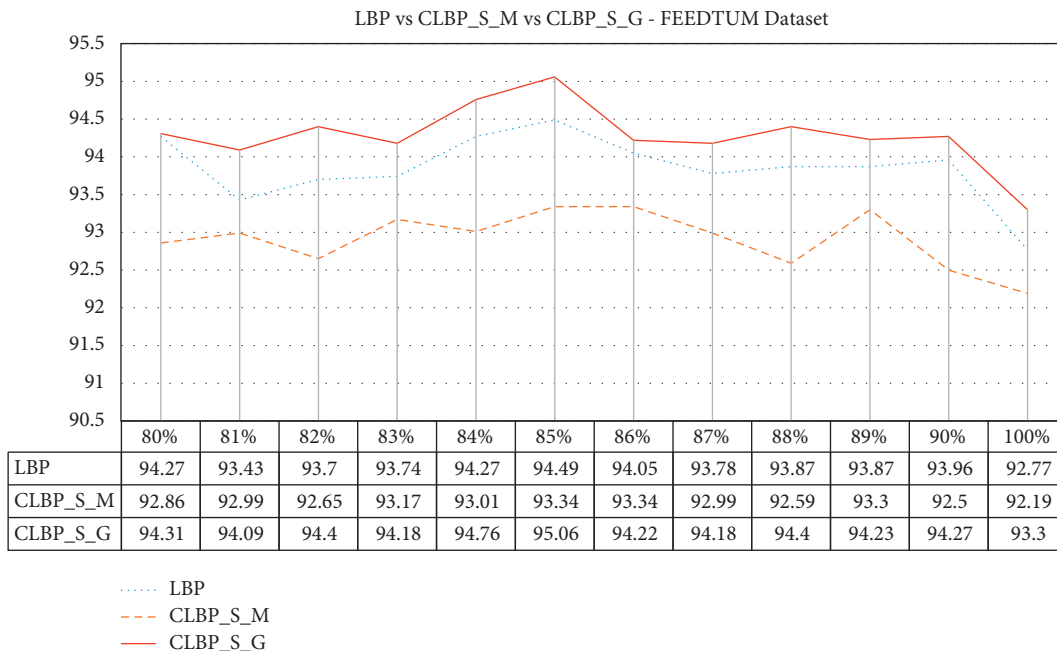FIGURE 15: A sad emotion image of character 0016 in the FEEDTUM dataset and its CLBP_S_G feature histogram [28].



LBP vs CLBP_S_M vs CLBP_S_G - FEEDTUM Dataset

|  | 80% | 81% | 82% | 83% | 84% | 85% | 86% | 87% | 88% | 89% | 90% | 100% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LBP | 94.27 | 93.43 | 93.7 | 93.74 | 94.27 | 94.49 | 94.05 | 93.78 | 93.87 | 93.87 | 93.96 | 92.77 |
| CLBP_S_M | 92.86 | 92.99 | 92.65 | 93.17 | 93.01 | 93.34 | 93.34 | 92.99 | 92.59 | 93.3 | 92.5 | 92.19 |
| CLBP_S_G | 94.31 | 94.09 | 94.4 | 94.18 | 94.76 | 95.06 | 94.22 | 94.18 | 94.4 | 94.23 | 94.27 | 93.3 |

······ LBP
--- CLBP_S_M
— CLBP_S_G

FIGURE 16: Comparison of experimental results of LBP, CLBP_S_M, and CLBP_S_G for the FEEDTUM dataset.

with beards, hair covering their forehead or wearing glasses, or subjects turning their heads sideways in different directions. In this work, all 18 subjects (9 men and 9 women) with a total of 2268 images are selected. Figure 14 displays a few examples of facial expression images of character 0002 in the FEEDTUM dataset.

Figure 15 shows a sad emotion image of character 0016 in the FEEDTUM dataset and its CLBP_S_G feature histogram.

In Figure 16 shows that the accuracy rates of non cropped images for all three kinds of features (LBP, CLBP_S_M, CLBP_S_G) are almost less than the accuracy rates of cropped images (from 80% to 90% of $w_2/w_1$) on the FEEDTUM dataset. The CLBP_S_G feature obtains the highest accuracy at 95.06% with the cropping percentage of $w_2$ to $w_1$ being 85% and completely better than LBP and CLBP_S_M features in the remaining percentages of $w_2$ to $w_1$.

Table 3 shows the confusion matrix of the FEEDTUM dataset based on the CLBP_S_G feature with a cropping percentage of $w_2$ to $w_1$ being 85%.

*4.5. The MUG Dataset.* The MUG dataset was made by the multimedia understanding group as in [30]. It was created to overcome some limitations of other similar datasets that preexisted at that time, such as low resolution, uniform lighting, many subjects, and many takes per subject. The Internet user version dataset has 52 subjects. This version of the dataset includes 52 Caucasian subjects with an age range between 20 and 35, 22 females and 30 males (with or without beards). The original images are saved in .jpeg format with a resolution of $896 \times 896$ pixels.

The experiments used 50 subjects, in which there are 22 females and 28 males. Fifteen images of facial emotions

TABLE 3: The confusion matrix of the FEEDTUM dataset.

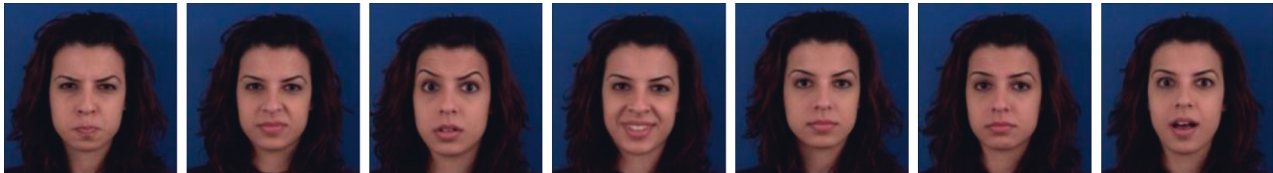| (%) | Anger | Disgust | Fear | Joy | Neutral | Sadness | Surprise |
|---|---|---|---|---|---|---|---|
| Anger | **94.14** | 2.47 | 0.00 | 0.31 | 0.31 | 2.78 | 0.00 |
| Disgust | 2.16 | **96.91** | 0.00 | 0.31 | 0.00 | 0.62 | 0.00 |
| Fear | 0.31 | 0.62 | **95.99** | 0.00 | 0.93 | 1.23 | 0.93 |
| Joy | 0.00 | 0.62 | 0.00 | **98.15** | 0.93 | 0.00 | 0.31 |
| Neutral | 1.85 | 0.31 | 0.93 | 0.00 | **92.90** | 3.40 | 0.62 |
| Sadness | 4.32 | 0.93 | 0.31 | 0.00 | 1.23 | **91.36** | 1.85 |
| Surprise | 0.31 | 0.00 | 2.16 | 0.31 | 1.23 | 0.00 | **95.99** |



FIGURE 17: A few facial expression images from the MUG dataset.
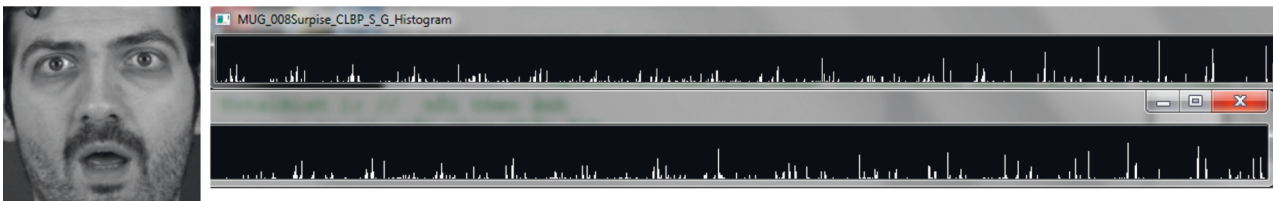


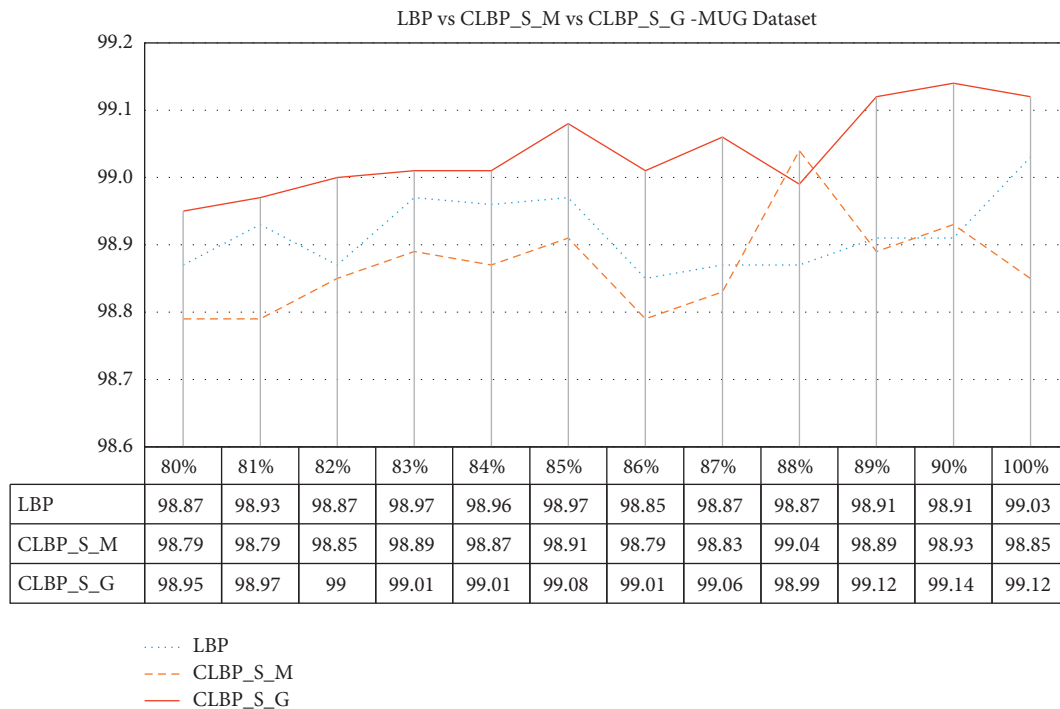FIGURE 18: A surprise emotion image of the MUG dataset and its CLBP_S_G feature histogram.



LBP vs CLBP_S_M vs CLBP_S_G -MUG Dataset

| | 80% | 81% | 82% | 83% | 84% | 85% | 86% | 87% | 88% | 89% | 90% | 100% |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| LBP | 98.87 | 98.93 | 98.87 | 98.97 | 98.96 | 98.97 | 98.85 | 98.87 | 98.87 | 98.91 | 98.91 | 99.03 |
| CLBP_S_M | 98.79 | 98.79 | 98.85 | 98.89 | 98.87 | 98.91 | 98.79 | 98.83 | 99.04 | 98.89 | 98.93 | 98.85 |
| CLBP_S_G | 98.95 | 98.97 | 99 | 99.01 | 99.01 | 99.08 | 99.01 | 99.06 | 98.99 | 99.12 | 99.14 | 99.12 |

········· LBP
- - - CLBP_S_M
——— CLBP_S_G

FIGURE 19: Comparison of experimental results of LBP, CLBP_S_M, and CLBP_S_G for the MUG dataset.

expressed from less to more of each expression are selected. Because some subjects did not express enough seven facial expressions, 5,130 human face images were selected in total, including 750 anger images, 735 disgust images, 705 fear images, 735 joy images, 750 neutral images, 705 sadness images, and 750 surprise images. A few facial expression images from the MUG dataset are shown in Figure 17.

TABLE 4: The confusion matrix of the MUG dataset.

| (%) | Anger | Disgust | Fear | Joy | Neutral | Sadness | Surprise |
|---|---|---|---|---|---|---|---|
| Anger | **99.07** | 0.13 | 0.00 | 0.00 | 0.80 | 0.00 | 0.00 |
| Disgust | 0.14 | **99.73** | 0.00 | 0.00 | 0.14 | 0.00 | 0.00 |
| Fear | 0.00 | 0.00 | **99.01** | 0.00 | 0.71 | 0.00 | 0.28 |
| Joy | 0.00 | 0.14 | 0.14 | **99.73** | 0.00 | 0.00 | 0.00 |
| Neutral | 0.13 | 0.00 | 0.00 | 0.00 | **99.87** | 0.00 | 0.00 |
| Sadness | 1.13 | 0.00 | 0.28 | 0.00 | 0.14 | **98.44** | 0.00 |
| Surprise | 0.00 | 0.00 | 1.87 | 0.00 | 0.00 | 0.00 | **98.13** |

Figure 18 shows a surprise emotion image of the MUG dataset and its CLBP_S_G feature histogram.

Figure 19 shows that the accuracy rates of non cropped images for all three kinds of features (LBP, CLBP_S_M, CLBP_S_G) are almost less than the accuracy rates of cropped images (from 80% to 90% of $w_2/w_1$) on the MUG dataset. The CLBP_S_G feature obtains the highest accuracy at 99.14% with a cropping percentage of $w_2$ to $w_1$ being 90% and almost better than LBP and CLBP_S_M features in the remaining percentages of $w_2$ to $w_1$.

Table 4 shows the confusion matrix of the MUG dataset based on the CLBP_S_G feature with a cropping percentage of $w_2$ to $w_1$ being 90%.

## 5. Conclusions

The research presents facial expression recognition using a novel Combined Gray Local Binary Pattern (CGLBP or CLBP_S_G) modeling with the technique of cropping face images in the preprocessing stage and the method of dividing face images into nonoverlap square regions in each stage of extracting features. The work also uses uniform features to reduce the lengths of descriptors. The experimental results on JAFFE, CK+, FEEDTUM, and MUG typical datasets show that the highest accuracy rate of the proposed feature modeling is better than that of LBP and CLBP feature modeling.

The proposed method used a small size image ($64 \times 64$ pixels) to reduce the processing time and data transmission time, but the accuracy rate is quite high. The calculating time of the CGLBP is also faster than that of CLBP since it does not calculate $m_p$ (the magnitude of $d_p$). These are advantages to apply in normal terminal devices or systems.

## Data Availability

The datasets used to support the findings of this study are included in the article.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

## References

[1] Y.-L. Tian, T. Kanade, and J. F. Cohn, *Handbook of Face Recognition*, Springer, Heidelberg, Germany, 2005.

[2] T. Küntzler, T. T. A. Höfling, and G. W. Alpers, "Automatic facial expression recognition in standardized and non-standardized emotional expressions," *Frontiers in Psychology*, vol. 12, Article ID 627561, 2021.

[3] K. S. Yadav and J. Singha, "Facial expression recognition using modified Viola-John's algorithm and KNN classifier," *Multimedia Tools And Applications*, vol. 79, no. 19-20, pp. 13089–13107, 2020.

[4] B. Niu, Z. Gao, and B. Guo, "Facial expression recognition with LBP and ORB features," *Computational Intelligence and Neuroscience*, vol. 1, 2021.

[5] T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[6] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[7] Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1657–1663, 2010.

[8] S. Chakraborty, S. K. Singh, and P. Chakraborty, "Local directional gradient pattern: a local descriptor for face recognition," *Multimedia Tools and Applications*, vol. 76, no. 1, pp. 1201–1216, 2017.

[9] S. Chakraborty, S. K. Singh, and P. Chakraborty, "Local gradient Hexa pattern: a descriptor for face recognition and retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 1, pp. 171–180, Jan. 2018.

[10] S. R. Dubey, S. K. Singh, and R. K. Singh, "Rotation and scale invariant hybrid image descriptor and retrieval," *Computers & Electrical Engineering*, vol. 46, pp. 288–302, 2015.

[11] S. Chakraborty, S. K. Singh, and P. Chakraborty, "Local quadruple pattern: a novel descriptor for facial image recognition and retrieval," *Computers & Electrical Engineering*, vol. 62, pp. 92–104, 2017.

[12] S. Chakraborty, S. K. Singh, and P. Chakraborty, "Centre symmetric quadruple pattern: a novel descriptor for facial image recognition and retrieval," *Pattern Recognition Letters*, vol. 115, pp. 50–58, 2018.

[13] S. R. Dubey, S. K. Singh, and R. K. Singh, "Boosting local binary pattern with bag-of-filters for content based image retrieval," in *Proceedings of the IEEE UP Section Conference on Electrical Computer and Electronics (UPCON)*, pp. 1–6, Allahabad, India, 2015.

[14] S. Chakraborty, S. K. Singh, and P. Chakraborty, "R-theta local neighborhood pattern for unconstrained facial image recognition and retrieval," *Multimedia Tools and Applications*, vol. 78, no. 11, pp. 14799–14822, 2019.

[15] S. Fekri-Ershad, "Bark texture classification using improved local ternary patterns and multilayer neural network," *Expert Systems with Applications*, vol. 158, Article ID 113509, 2020.

[16] S. Fekri-Ershad, "Cell phenotype classification using multi threshold uniform local ternary patterns in fluorescence microscope images," *Multimedia Tools and Applications*, vol. 80, no. 8, pp. 12103–12116, 2021.

[17] T. Nhan Cao, A. H. Ton-that, and H. I. Choi, "Facial expression recognition based on local binary pattern and support vector machine," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 28, no. 6, Article ID 1456012, 2014.

[18] N. T. Cao, A. H. T. That, and H. I. Choi, "An effective facial expression recognition approach for intelligent game systems," *International Journal of Computational Vision and Robotics*, vol. 6, no. 3, pp. 223–234, 2016.

[19] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[20] M. Everingham and A. Zisserman, "Regression and classification approaches to eye localization in face images," in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FGR'06)*, pp. 441–446, Southampton, UK, 2006.

[21] Y.-Li Tian, "Evaluation of face resolution for expression analysis," in *Proceedings of the IEEE Conference Computer Vision and Pattern Recognition Workshop (CVPRW'04)*, pp. 82–88, Washington, DC, USA, 2004.

[22] K. Li, Y. Jin, M. W. Akram, R. Han, and J. Chen, "Facial expression recognition with convolutional neural networks via a new face cropping and rotation strategy," *The Visual Computer*, vol. 36, no. 2, pp. 391–404, 2020.

[23] A. Joseph and P. Geetha, "Facial emotion detection using modified eyemap-mouthmap algorithm on an enhanced image and classification with tensorflow," *The Visual Computer*, vol. 36, no. 3, pp. 529–539, 2020.

[24] M. J. Lyons, K. Miyuki, and G. Jiro, "Coding facial expressions with gabor wavelets (ivc special issue)," 2020, https://arxiv.org/pdf/2009.05938.pdf.

[25] M. Lyons, M. Kamachi, and J. Gyoba, "The japanese female facial expression (jaffe) dataset," 1998, https://doi.org/10.5281/zenodo.3451524.

[26] "Yhwstohbyuk," https://zenodo.org/record/3451524.

[27] P. Lucey, J. F. Cohn, K. Takeo, S. Jason, and A. Zara, "the extended cohn-kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, San Francisco, CA, USA, 2010.

[28] W. Frank, B. Schuller, M. Hawellek, and G. Rigoll, *Efficient Recognition of Authentic Dynamic Facial Expressions on the Feedtum Database IEEE ICME*, pp. 493–496, IEEE Computer Society, Piscataway, NJ, USA, 2006.

[29] P. Ekman and W. V. Friesen, *Facial Action Coding System: A Technique for the Measurement of Facial Movement*, Consulting Psychologists Press, Palo Alto, California, 1978.

[30] N. Aifanti, A. Christos Papachristou, and Dolopoulos, "The mug facial expression database," in *Proceeding of the 11th Interantional Workshop On Image Analysis For Multimedia Interactive Services (WIAMIS)*, Desenzano, Italy, 2010.