

# Region-Based Facial Expression Recognition in Still Images

Gawed M. Nagi\*, Rahmita Rahmat\*, Fatimah Khalid\* and Muhamad Taufik\*

**Abstract**—In Facial Expression Recognition Systems (FERS), only particular regions of the face are utilized for discrimination. The areas of the eyes, eyebrows, nose, and mouth are the most important features in any FERS. Applying facial features descriptors such as the local binary pattern (LBP) on such areas results in an effective and efficient FERS. In this paper, we propose an automatic facial expression recognition system. Unlike other systems, it detects and extracts the informative and discriminant regions of the face (i.e., eyes, nose, and mouth areas) using Haar-feature based cascade classifiers and these region-based features are stored into separate image files as a preprocessing step. Then, LBP is applied to these image files for facial texture representation and a feature-vector per subject is obtained by concatenating the resulting LBP histograms of the decomposed region-based features. The one-vs.-rest SVM, which is a popular multi-classification method, is employed with the Radial Basis Function (RBF) for facial expression classification. Experimental results show that this approach yields good performance for both frontal and near-frontal facial images in terms of accuracy and time complexity. Cohn-Kanade and JAFFE, which are benchmark facial expression datasets, are used to evaluate this approach.

**Keywords**—Facial Expression Recognition (FER), Facial Features Detection, Facial Features Extraction, Cascade Classifier, LBP, One-Vs-Rest SVM

## 1. INTRODUCTION

Facial expressions are considered to be an important and natural way to enrich human-to-human interactions and they add more meaning to human communication. The first step of facial expression analysis goes back to the early 1970s when Ekman and Friesen [1] defined and categorized the human emotions that can be displayed by the face into six basic expressions, which are referred to as happiness, sadness, surprise, fear, anger, and disgust. All of which are universally distinguishable irrespective of the diversity of the human race, ethnicity, or culture. Moreover, they developed the Facial Action Coding System (FACS), which describes human face behaviors and it codes the facial expressions based on the movements of the face muscles. These codes are called action units (AUs). FACS and these six prototypic expressions form the cornerstone and a de-facto standard in facial expression analysis and recognition research. They have also become widely used by most of the researchers in these areas.

---

Manuscript received October 24, 2012; first revision January 24, 2013; accepted March 6, 2013.

**Corresponding Author: Gawed M. Nagi**

\* Faculty of Computer Science and Information Technology, University Putra Malaysia (UPM), Selangor, Malaysia (gnagi2000@yahoo.com)

A study done by Mehrabian [2] reported that facial expressions play a crucial role in understanding the verbal messages in human-to-human interactions and that they contribute to 55% of understanding social communication, while verbal and vocal components contribute to 7% and 38%, respectively. In other words, facial expressions deliver information more than spoken words of importance.

The real start of automatic facial expression recognition systems was launched in the early 1990s. At the time, computing power had rapidly become cheaper and widely available. Face detection and tracking research, which are required for FERS [3], had already made significant progress. Since the early 1990s, FERS have attracted many researchers from several disciplines, such as the fields of robotics, psychology, and computer vision. Furthermore, it has gained increasingly potential applications in areas such as human-computer interaction systems, image retrieval, face modeling, patient monitoring systems (e.g., for pain and depression detection), face animation, and so forth.

Although facial expression recognition has been active over the last two decades and substantial advancements have been made with it that have yielded promising results, it is still challenging and far from being satisfactory in terms of speed and accuracy due to the richness, complexity, and variability of human face expressions [4].

In this paper, we present a region-based approach for automatic facial expression recognition in still images using extracted region-based features, which represent the most prominent facial features with important and discriminant information. The proposed approach involves two main tasks. The first is to detect and extract the regions of interest of facial features using boosted cascade classifiers with building a cascade classifier to detect the closed/open mouth region, particularly in scared and surprised facial images. Then, the second task is applying LBP as a face descriptor to these extracted region-based feature images to produce an effective feature vector, which in its turn is used as input for the one-vs.-all SVM classifier to finish the facial expression recognition.

The main focus of this work is to propose an approach that is suitable to work in real facial expression recognition scenarios by obtaining a balance between the speed and accuracy required in such scenarios. The rest of this paper is structured as follows: In Section 2, an overview of some related works is presented. In Section 3 the automatic facial expression recognition system is described in combination with its main components. Finally, the conclusion is drawn in Section 4.

## 2. RELATED WORKS

Over the last two decades, intensive research has been conducted to develop more robust person-independent facial expression recognition systems that work in real-time and in difficult scenarios with lighting changes, facial makeup, different ages and races, low resolution images, and facial occlusions. Gabor filters are optimally used for orientation and frequency localization and have been utilized in many computer vision applications, including image processing and facial expression recognition. Barlette et al. [5] developed an automatic real-time facial actions recognition system. The system was based on video sequences from spontaneous and posed databases in which 20 action units are tracked and classified using Gabor filters for features extraction and a combination of SVM and AdaBoost classifiers for classification.

Additionally, Giorgana [6] presented a fully automatic FERS to recognize the emotions of joy, surprise, sadness, and neutrality by utilizing Gabor filters for feature extraction and AdaBoost for feature selection. A recognition rate of 87.14% has been reported using the one-vs.-one support vector machine (SVM) and Error-Correcting Output Codes (ECOC) with normalized face images that are 96 x 96 in size.

Ryan et al. [7] used Constrained Local Models (CLMs) to extract and represent facial features, and support vector machines (SVMs) to classify expressions. In this work, CLM has been proven to be more robust against occlusions and appearance changes in comparison to active appearance models (AAMs). In [8], Lu et al. proposed the pixel-pattern-based texture feature (PPBTF) as a new method for feature representation. It is insensitive to illumination and is less time consuming to use. In Lu's method, AdaBoost and SVM are used for enhancing the discrimination and classification respectively, and an average recognition rate of 92.26% has been stated.

Higher-order local autocorrelation (HLAC) has been successfully applied to face recognition [9, 10] and to gesture recognition [11] due to its robustness in object translation and it is also computationally inexpensive. In this context, Lajevardi et al. [12] employed HLAC and HLAC-like features (HLACLF) for feature extraction and the Naive Bayesian classifier for classification. A more than 93% classification accuracy has been reported with Cohen-Kanade and JAFFE databases.

Feng et al. (2005) proposed an approach for facial expression recognition in which the local binary pattern (LBP) is employed as a facial expression descriptor and where linear programming (LP) is used for facial expression classification. A 93.8% facial expression recognition rate has been reported using the JAFFE dataset. However, this approach has not been evaluated on a large database.

A comprehensive study conducted by Shan et al. (2009) proved the superiority of LBP and of boosting LBP in the facial expression recognition area with several classifiers (e.g., SVM, LDA, and LP), as well as with using low-resolution facial images and compressed video sequences. Zhao & Zhang (2011) proposed LBP for facial expression representation and kernel discriminant isometric mapping (KDIsomap) for producing the low-dimensional discriminant embedded data representations from the extracted LBP features. The best accuracies of 94.88% and 81.59% are achieved on the Cohn-Kanade and JAFFE databases, respectively. Due to its computation cost and its sensitivity to noise, Isomap is not a choice for facial expression recognition in real systems.

As a conclusion of the literature listed above, the most critical factor in facial expression recognition systems is the way that the informative and discriminative facial features that change with the different facial expressions are extracted and represented.

### **3. SYSTEM DESCRIPTION**

Generally, automatic FERS involves the following three main steps: face detection, features extraction and representation, and facial expression classification. Fig.1 depicts our basic system structure.

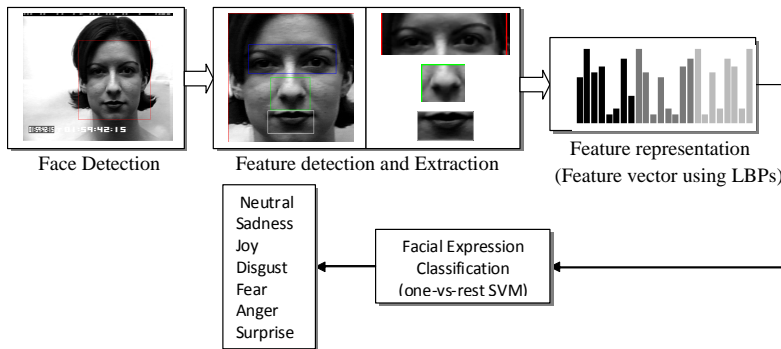


Fig. 1. The basic structure of the facial expression recognition system

### 3.1 Facial expression datasets

The experiments are conducted on the two standard facial expression databases (the so-called Cohen-Kanade [13] and JAFFE [Japanese Female Facial Expression] [14] databases). A brief description of those two databases is shown in Table 1. It should be noted that the test images are different from the training images as they cover the intensity of each expression of the six main facial expressions.

Table 1. A Brief Description of the Databases that were Used

	Cohn-Kanade Database	JAFFE Database
Number of Subjects	100	10
Number of Expressions	23 facial displays	7 (6 basic facial expressions + 1 neutral)
Type of Expression	Posed and spontaneous	Posed
Race	Caucasian + African-American + Asian and Latino	Japanese
Gender	Female (69%) + Male (31%)	Female
Samples per Expression	Not all subjects perform all the basic expressions	2,3, or 4
Total Number	Approximately 2,000 image sequences	213
Number of Images for Training	424	173
Number of Images for Testing	120	40

### 3.2 Automatic face detection

Face detection is the prior step, not only for facial expression tracking and recognition, but also for many human-computer interaction systems, such as face gender and age recognition. The more accurate the face detection is, the better the recognition performance is. Although a lot of research has been conducted in the object and face detection area, Viola and Jones' [15] cascade classifier, which is based on Haar-like features, is the most commonly used technique for face

detection, particularly in real-world applications. This is due to its simplicity and robustness. Several extensions and developments have been made to the Viola-Jones approach [16-19].

In this work, OpenCV implementation of the Viola-Jones algorithm [20] is used with updates for face detection. The face detection rate was above 90% and faces are automatically detected and rescaled to  $120 \times 120$  pixels. Additionally, automatic eye detection is employed for the elimination of false alarms.

### 3.3 Facial features detection and extraction

Detecting and describing the discriminative facial features are crucial for many applications in computer vision. Cascade classifiers based on Haar-like features are used with OpenCV to detect and extract the most important and discriminative facial features that are needed for facial expression classification. These features (e.g., eyes, eyebrows, nose, and mouth) are automatically detected and their regions are extracted and stored in separate image files. The aim is to get rid of all unnecessary extrinsic areas of face space and to make the computation more suitable for real-life applications.

#### 3.3.1 Open-mouth classifier

As far as we know, no classifier in literature has been developed that is able to detect open mouths precisely, in particular for surprised and scared face images. Therefore, the open-mouth cascade classifier was developed based on Haar-like features to detect open mouths in such cases, as well as for detecting closed mouths. The OpenCV library is utilized to build, train, test, and evaluate a new open-mouth classifier.

Regarding the size of the training set, Lienhart et al. [21] proposed 5,000 positive and 3,000 negative images as a training set whereas Viola & Jones [22] trained their classifier using 5,000 faces and 1,0000 non-faces that were randomly selected from non-face images. Similarly, Lefkovits [23] generally analyzed the performance of cascade classifiers and argued that 5,000 face images would be enough for the training of face detectors but there is no optimal number for backgrounds (i.e., negative images). He used 3,000 positive images and 27,000 negative ones for face detector training.

In this study, 5,000 positive images derived from 1,000 original open and closed mouths were cropped from Cohen-Kenade, JAFFE, and BU\_3DFE [24] databases with the following parameters:

- Random rotation =  $\pm 10$  degrees
- Random scaling =  $\pm 10$  %
- Random mirroring and shifting up to  $\pm 1$  pixel

Besides the positive images, 3,000 negative images (backgrounds) were collected from the web. These images included non-facial images and facial-regions but did not contain any mouths. Fig. 2 shows some samples of these positive (e.g., open and closed mouths) and negative images (facial and non facial images). The training process with 20 cascade stages takes two days on an Intel Pentium 4 3.0GHz with 2GB memory machine. The performance of the trained open-mouth classifier is illustrated in Table 2.

“Hits” represents the number of correctly detected open mouths, “Missed” represents the

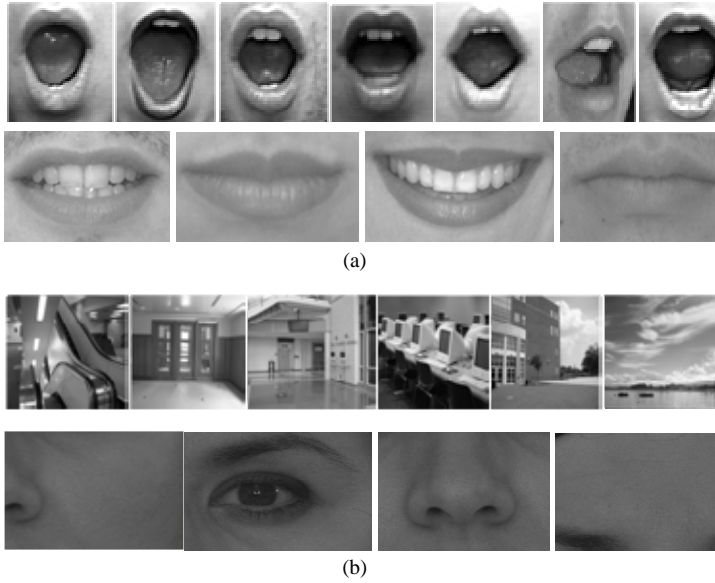


Fig. 2. Samples of positive (a) and negative (b) samples

Table 2. The Performance of the Proposed Open-Mouth Cascade Classifier

	Hits	Missed	False
Total	828	172	131
Number of stages	20		
Number of weak classifiers	82		

number of false negatives (i.e., missed detections) where the object (open mouth) is there but the detector failed to detect it, and “False” represents the number of false positives where the detector alerted that an object exists but there is actually no object.

The scanning window size is set to  $24 \times 24$ , which is used in most cascade classifiers in real-time object detection. Indeed these small scanned windows generate a large number of features and hence the idea of boosting (e.g., AdaBoost) is used to select the most appropriate features. Boosting algorithms combine a set of weak learners into one powerful strong classifier. In fact, selecting the training parameters of cascade classifiers is considered a challenge since there is no optimal scheme to choose these parameters.

The open-mouth classifier is built with 20 stages and each stage of the cascade eliminates 50% of the false positives, which means that 50% of the negative samples are eliminated in the early stages of the cascading process. Table 3 shows the detection rate of the proposed classifier in the last five stages whereas Fig. 3 illustrates the ROC curves of classification on the Cohn-Kanade database where the performance of two open-mouth classifiers was built in 10 stages and in 20 stages.

Furthermore, a comparison was conducted on the same facial images to compare the performance of the proposed open-mouth classifier to the mouth classifier that was developed by Castrillon-Santana et al. [25], which is publicly available [26]. The output of both classifiers is

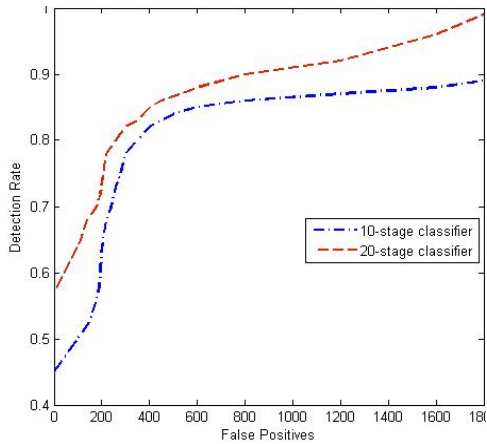


Fig. 3. The ROC of the trained classifier with different numbers of stages

Table 3. The Detection Rate of the Proposed Classifier

Stage no.	16	17	18	19	20
Hit rate	91.4	90	93.2	92.2	93



Fig. 4. Open-mouth detection performance with Castrillon-Santana's classifier (upper) and the proposed cascade classifier (lower)

illustrated in Fig. 4.

It is obvious that the borders of the detected open mouth are not well fitted to the open mouth region, as shown in Fig. 4, because facial expressions are posed and are extreme simulations of spontaneous ones.

3.4 Facial feature representation

Facial feature representation and classifier design are the most vital aspects in FERS. Facial feature representation can be accomplished by using either geometric-based methods or appearance-based methods. On one hand, geometric based methods deal with the shape and location of

the prominent facial components, such as the eyes, mouth, nose, and eyebrows. Those facial features reflect the visual variations caused by the different expressions. On the other hand, appearance based methods deal with the image filters, such as LBP and Gabor wavelets, which apply to the whole face or part of it, to generate the feature vector.

#### 3.4.1 Local Binary Patterns (LBP)

In this work, LBP is employed as a texture descriptor. Ojala et al. [27] first described LBP. Due to its resilience against illumination changes, computational simplicity, and discrimination performance, LBP has been used in many studies in different areas where texture analysis is required. A recent survey [28] presents a number of variations and extensions to the LBP texture operator along with their applications. In short, the LBP operator labels every pixel in the region of interest by thresholding the  $3 \times 3$  neighborhood with the center pixel value. An 8-digit binary number is given and converted into its equivalent decimal, as shown in Fig. 5.

For each pixel  $p(x,y)$ , the LBP code is calculated as follows:

$$LBP_{P,R} = \sum_{P=0}^{P-1} S(g_P - g_C) 2^P \quad (1)$$

Where:

- $P$  denotes the sampling points (circular neighborhood pixels).
- $g_P, g_C$  denote the gray value of the sampling and center pixels.
- $S(x) = \begin{cases} 1 & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$  (2)

The LBP operator was improved to deal with various size of neighborhoods and to avoid the limitation of the  $3 \times 3$  neighborhood of the original LBP operator. Ojala et al. [29] presented an extension of the original LBP operator by using uniform patterns and they empirically proved that uniform patterns represent the majority of image patterns (around 90%) when using the (8,1) neighborhood and 70% with the (16,2) neighborhood. A Local Binary Pattern is labeled as being “uniform” when there are no more than 2 bitwise transitions of 0 to 1 or vice versa. For example, the patterns 00000000 (0 transition), 11110000 (1 transition), and 11100111 (2 transitions) are uniform and patterns like 00110011 (3 transitions) and 10010010 (5 transitions) are not.

In this paper, uniform patterns are employed to extract the LBP code for each pixel of region-based images and 58 output labels are produced for 8 sampling-point neighborhoods ( $p$ ) according to uniform LBP mapping ( $p^2 - p + 2$ ). Once the LBP operator labels the image, a histogram of the labeled image  $f_l(x,y)$  is computed as follows:

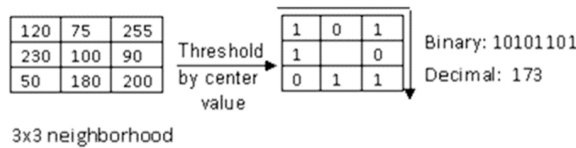


Fig. 5. The original LBP operator



$$H_i = \sum_{x,y} I\{f_i(x,y) = i\}; i = 0, \dots, n-1 \quad (3)$$

where  $n$  indicates the number of different labels produced by the LBP operator, which is 58 herein, and:

$$I\{A\} = \begin{cases} 1, & A \text{ is true} \\ 0, & A \text{ is false} \end{cases} \quad (4)$$

The resultant histogram gives information about the distribution of the local micro-patterns, such as the edges, spots, and flat areas over the whole image.

### 3.5. Facial expression recognition

#### 3.5.1 Multi-class SVMs

Prior to applying the SVM, data scaling is advisable to improve classification. Many advantages of data scaling, such as preventing the attributes of high numeric ranges from dominating those of low numeric ranges and avoiding the complexity of calculations, which are reported in [3]. Data is scaled in both training and testing datasets into the range [0, 1].

Mainly, SVM, which is a state-of-the-art multiclass method, has been developed for binary classification (two classes). However, SVM is utilized for multi-class classification by combining several binary SVMs and by constructing a multi-class SVM. A One-vs.-rest SVM, which is also referred to as one-against-all SVM classifier, was adopted in this work. It is the most common technique used for multi classification in which one SVM is built for each class and trained to discriminate the examples in a single class from those in the rest of classes. One-vs.-rest SVM performance is empirically approved as compared with other multi classification techniques, such as the one-vs.-one SVM [30]. The public machine learning library, Spider [31], is utilized to implement the one-vs.-rest SVM for multi classification with the Radial Basis Function (RBF) kernel and the 10-fold cross validation.

#### 3.5.2 Experiments and discussion

The region-based features of the eyes and eyebrows, nose, and mouth are extracted and separated into three independent images. The areas around those discriminative features are ignored, as shown in Fig. 6. In fact, we try to answer the question, “Can one facial feature (e.g., eyes, eyebrows, nose, or mouth) or a combination of such features be informative enough for facial expression discrimination?”

In histogram-based methods, partially overlapping regions can be found and they do not affect the overall performance of facial expression recognition systems.

Five experiments were conducted to investigate the facial expression recognition performance across a different combination of region-based feature images. In the first experiment, the proposed approach was run on all region-based feature images in which the LBP operator is applied to the eyes and eyebrows, nose, and mouth region-based images. Then, LBP histograms were concatenated to form the so-called feature vector, which is stored in a single row (i.e., a single



Fig. 6. Cropped face with the detection of the main features and the features are extracted into separate images

row per face subject) in the feature-data file. The constructed feature-data file forms the training data. For testing, the same steps are applied on the tested input image to generate the single row testing data file. Afterwards, the one-against-rest SVM technique is employed for classifying with the radial basis function (RBF) kernel. The one-against-rest SVM takes the training data along with the class labels and the testing data as input. The classification results are shown in Table 4 with the Cohn-Kanade database. It is obvious that neutral and surprised faces are recognized with a high accuracy, whereas expressions of sadness and fear are less recognizable.

The best facial expression recognition rate of 90% was achieved with only 0.38 seconds being consumed for each image on a standard PC (Pentium 4 3.0GHz, 2GB RAM) in a Matlab (R2011a) environment when using the Cohn-Kanade database. Using the JAFFE database, the expressions of fear, sadness, and anger showed a low recognition rate compared to the expressions of neutral, joy, and surprise, as illustrated in Table 5.

In the second experiment and for comparison purposes, a general approach that is adopted by some researchers in the facial expression recognition field [4, 32-35] was implemented. With this approach the detected face is divided into small regions of  $6 \times 6$  blocks, as shown in Fig. 7, and then a single feature vector is generated by concatenating the extracted LBPs histograms.

Table 4. The Confusion Matrix of Facial Expression Recognition Using the One-vs.-Rest SVM and the RBF Kernel Based on the Cohn-Kanade Database

	Neutral %	Joy %	Sadness %	Surprise %	Anger %	Fear %	Disgust %
Neutral	85.5	3.4	7	0	2.1	0.6	1.4
Happiness	6.2	83.7	4.4	2.2	1.4	2.1	0
Sadness	7.8	3	77	0	3.5	5.6	3.1
Surprise	1.2	3.7	0	90	2	3.1	0
Anger	9	0	2.3	0	81.2	4.2	3.3
Fear	2.3	3.2	4.4	12	0	78.1	0
Disgust	0	3.4	3.8	0	3	5.8	84

Table 5. Recognition Rate of Facial Expression Recognition Using the One-vs.-Rest SVM and the RBF Kernel Based on the JAFFE Database

	Neutral	Joy	Sadness	Surprise	Anger	Fear	Disgust
Recognition rates (%)	87	85.3	72.2	84	71	70	76

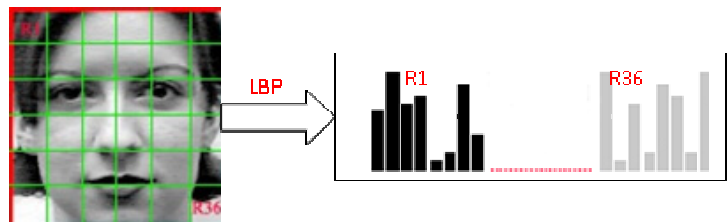


Fig. 7. Applying LBPs to all regions of the detected face

Table 6. Recognition Rates of Facial Expressions Using the General Approach Based on the Cohn-Kanade Database

	Neutral	Joy	Sadness	Surprise	Anger	Fear	Disgust
Recognition rates (%)	84	81.3	75	87.4	79	80	85

After that, the SVM is employed for classification. The results are illustrated in Table 6.

The results clearly indicate that applying the LBP operator to region-based feature images offers better recognition performance than when applying the LBP operator to the entire face region, even with a weighting scheme. Fig. 8 shows the performance of both our method and the general method in a visual way. Although the accuracy of the proposed approach is slightly better than the one obtained by the so-called “general approach,” the time complexity, which is a big concern in real applications, has been reduced to some extent.

In the third experiment, the LBP operator is applied to region-based feature images but only the images of the eyes plus the eyebrows, and of the mouth were used. The feature vector was composed and the classification results are shown in Table 7. Similarly, in the fourth experiment, the LBP operator is applied to images of the eyes plus the eyebrows, and of the nose. The results of classification are illustrated in Table 8. It is obvious that the regions of the eyes plus the eyebrows, and the mouth are important and that they are the most discriminant facial features.

In the last experiment, the LBP operator is applied to mouth and nose region-based images. It is observed that these two features solely give a poor recognition rate of facial expressions. Table 9 shows the classification of facial expressions using these two features. From the last three experiments, the areas of the eyes and the eyebrows, and the mouth have been shown to have a

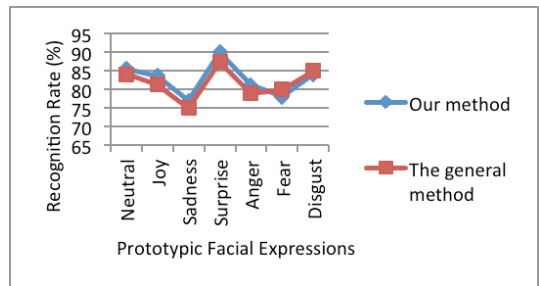


Fig. 8. The performance of our method and the general method

Table 7. Recognition Rates of Facial Expressions Using Only Images of the Eyes + Eyebrows, and Mouth as Features, Based on the Cohn-Kanade Database

	Neutral	Joy	Sadness	Surprise	Anger	Fear	Disgust
Recognition rates (%)	60	58	53.2	65	45	48	42

Table 8. Recognition Rates of Facial Expressions Using Only Images of the Eyes + Eyebrows, and Nose as Features, Based on the Cohn-Kanade Database

	Neutral	Joy	Sadness	Surprise	Anger	Fear	Disgust
Recognition rates (%)	45	42	40.2	36.5	32.7	34	38

Table 9. Recognition Rates of Facial Expressions Using Only Images of the Mouth and Nose as Features, Based on the Cohn-Kanade Database

	Neutral	Joy	Sadness	Surprise	Anger	Fear	Disgust
Recognition rates (%)	42	39.2	34	40	35	32.8	33

good discriminative ability for different facial expressions.

### 3.5.3 Cross Validation (CV)

Assessing the generalization performance of learning methods with unknown test data is critical and important in practical applications, since it guides the choice of the learning method and gives a measure of quality for the selected method [36]. K-fold cross validation is the simplest and most commonly used technique for measuring the prediction accuracy and generalization performance in many computer vision and machine learning applications and to also prevent the problem of overfitting. With this technique the data set is partitioned into approximately equal k subsets. In each round, one subset is used as a test set while all the other k-1 subsets are used for training so that all of the examples in the dataset are eventually used for both training and testing, as show in Fig. 9. Then the average error across all k folds is computed:

$$E = \frac{1}{k} \sum_{i=1}^k E_i \quad (5)$$

where  $E_i$  denotes the estimated error of the  $i$ th fold.

In this work, 10-fold cross validation ( $k=10$ ) is adopted to ensure both the efficiency and effectiveness of this technique. Table 10 shows the cross validation (CV) accuracy for the one-vs.-rest SVM as compared to the k-Nearest Neighbor (kNN) on 300 images of Cohn-Kanade database.

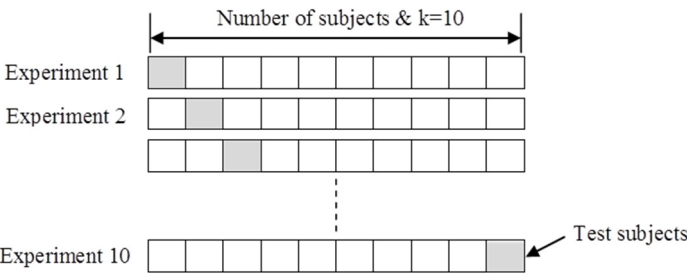


Fig. 9. Visualization of 10-k cross validation

Table 10. The Cross Validation (CV) Accuracy for the One-vs.-Rest SVM and kNN

Classifier	Parameters	Average CV Error	Recognition Rate
One-vs.-rest SVM	rbf = 5	0.2219	77.81
kNN	rbf = 5 , k=3	0.23781	76.21
	rbf = 5 , k=5	0.2235	77.65

4. CONCLUSION AND FUTURE WORK

Using local features based methods to represent facial images has been achieving promising results and attracting many researchers in the computer vision community recently. Since most of the developed methods of facial expression recognition are not being further investigated in real environments, this paper proposed an approach to provide the best possible tradeoff between accuracy and computational cost in identifying facial expressions under real scenarios. The proposed approach achieved 90% accuracy with a low computational cost (0.38s per image).

In short, this approach detects and extracts the main facial features regions (i.e., the regions of the eyes, eyebrows, nose, and mouth) from a detected cropped face using the boosted cascade classifiers. These region-based features are stored into separated images and all unnecessary surrounding areas are cancelled.

Since no precise open-mouth detector has been reported in literature and since an open mouth is an important discriminate feature in fear and surprise facial expressions, an open-mouth classifier was built and evaluated to detect open mouths with more precision in such expressions. Applying LBPs to the entire face manifold would significantly increase the amount of time and the required storage. So in this paper, LBP as a texture descriptor was applied only to the separated region-based feature images and the extracted histograms were concatenated into a feature vector to represent the facial texture of each face image with an expression.

For classification, the one-vs.-rest SVM was adopted with RBF since this technique is more reasonable for automatic multi classification systems. The results of the experiments showed that the proposed approach provides a better recognition rate in comparison to applying the LBP operator to the whole face region. Additional experiments showed that eyes plus the eyebrows, and the mouth are the most discriminative features for facial expression classification, while the other surrounding extrinsic areas of the face can be removed to reduce the time complexity in real-world facial expression recognition systems.

In summation, a new approach was introduced for FER based on the fusion of results from a

set of separated region-based feature images that have been represented by LBP and that were independently matched by multiclass SVM.

## REFERENCES

- [1] Ekman P. and Friesen W. V., "Constant across cultures in the face and emotion," *Journal of personality and social psychology*, 1971, pp.124-129.
- [2] Mehrabian, A. "Silent messages: Implicit communication of emotions and attributes (2nd edition)," Belmont, CA: Wadsworth Publishing Company, 1981.
- [3] Hsu C.-W., Chang C.-C., Lin C.-J., "A Practical Guide to Support Vector Classification," Tech. Rep., Taipei, 2003.
- [4] Shan C., Gong S., and McOwan P., "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, 2009, pp.803-816.
- [5] Bartlett M., G Little wort C., Lainscsek C ., Fasel I., Frank M., and Movellan J., "Fully automatic facial action recognition in spontaneous behavior," *Automatic Face and Gesture Recognition* , 2006, pp.223-228.
- [6] Giorgana G., "Facial Expression Recognition for Domestic Service Robots," *Proceedings of SIMPAR 2010 Workshops*, 2010, pp.251-262.
- [7] Ryan A., Jeffery F. Cohn, and Adam Rossi, "Automated Facial Expression Recognition System," in *IEEE 43rd International Carnahan Conference on Security Technology*, Zurich, 2009, pp.172-177.
- [8] Lu H., Huang Y., Chen Y. -W., and Yang D., "Real-time facial expression recognition based on pixel-pattern-based texture feature," *Electronic Letters*, 43 (17), 2007, pp.916-918.
- [9] Nomoto N. and Otsu N., "A New Scheme for Image Recognition using Higher-order Local Autocorrelation and Factor Analysis," *MVA2005*, 2005.
- [10] Kurita T., Otsu N., and Sato T., "A face recognition method using higher order local autocorrelation and multivariate analysis," *Proc. 11th IAPR International Conf. on Pattern Recognition*, 1992, pp.213-216.
- [11] Kurita T., and Hayamizu S., "Gesture Recognition using HLAC Features of PARCOR Images and HMM based Recognizer," *IEEE Int. Conf. on Automatic Face and Gesture Recognition* , Japan,1998.
- [12] Lajevardi, S., Hussain, Z., "Automatic facial expression recognition: feature extraction and Selection," *Signal, Image and Video Processing*, 2011, pp.1-11.
- [13] Kanade T., Cohn J. F., and Tian Y., "Comprehensive Database for Facial Expression Analysis," in *Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, Grenoble, France, 2000, pp.46-53.
- [14] Lyons M. J., Akamatsu S., Kamachi M., and Gyoba J., "Coding Facial Expressions with Gabor Wavelets," in *Third IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, Japan, 1998, pp.200-205.
- [15] Viola P., and Jones M., "Rapid Object Detection Using a Boosted Cascade of Simple Features," *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* , Kauai, Hawaii, Vol.1, 2001, pp.511-518.
- [16] Lienhart R., and Maydt J., "An extended set of haar-like features for rapid object detection," *Proceedings of the IEEE International Conference on Image Processing*, Rochester, New York, Vol.1, 2002, p. 900-903.
- [17] Li S.Z., and Zhang Z., "FloatBoost Learning and Statistical Face Detection", *PAM I*, 2004, pp.1112-1123.
- [18] Mita T., Kaneko T., and Hori O., "Joint Haar-like features for face detection," *Proceedings of the International Conference on Computer Vision*, Washington, DC, USA, 2005, pp.1619-1626.
- [19] Sri-Kaushik Pavani, David Delgado, and Alejandro F. Frangi, "Haar-like features with optimally weighted rectangles for rapid object detection," *Pattern Recognition*,2010, pp.160-172.
- [20] "Face Detection using OpenCV", Available at: <http://opencv.willowgarage.com/wiki/FaceDetection>
- [21] Lienhart R., Kuranov A., and Pisarevsky V., "Empirical Analysis of Detection Cascades of Boosted

- Classifiers for Rapid Object Detection,” Intel Corporation , Technical report, 2002, pp.297-304.
- [22] Viola, P., & Jones, M. (2004). Robust real-time face detection. *International Journal of Computer Vision*, 57(2), 137-154.
  - [23] Lefkovits, S. (2009). Assessments of building classifiers for face detection. *Proceedings of International Conference on Recent Achievements in Mechatronics, Automation, Computer Sciences and Robotics MACRO* (pp.175-186).
  - [24] Yin L., Wei X., Sun Y., Wang J., and Rosato M., “ A 3D facial expression database for facial behavior research,” In *Proceedings of 7th International Conference on Automatic Face and Gesture Recognition*, 2006, pp.211-216.
  - [25] Castrillon M.-Santana, Deniz-Suarez O., Anton-Canals L., and Lorenzo-Navarro J., “Face and facial feature detection evaluation: Performance Evaluation of Public Domain Haar Detectors for Face and Facial Feature Detection,” in *VISAPP*, 2008.
  - [26] “Haar Cascades “, Available at: <http://alereimondo.no-ip.org/OpenCV/34>
  - [27] Ojala T., Pietikäinen M., and Harwood D., “A comparative study of texture measures with classification based on featured distribution,” *Pattern Recognition*, 1996, pp.51-59.
  - [28] Huang D., Shan C., Ardabilian M., Wang Y., and Chen L., “Local binary patterns and its application to facial image analysis: a survey,” *IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews* 41 (4), 2011, pp.1-17.
  - [29] Ojala T., Pietikäinen M., and Mäenpää T., “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, pp.971-987.
  - [30] Altun H. and Polat G., “New frameworks to boost feature selection algorithms in emotion detection for improved human-computer interaction,” *Brain Vision and Artificial Intelligent. Lecture Notes in Computer Science* , 2007, pp.533-541.
  - [31] “The Spider”: <http://people.kyb.tuebingen.mpg.de/spider/>
  - [32] Feng X., Hadid A., and Pietikäinen M., “A coarse-to-fine classification scheme for facial expression recognition,” in *Proc. Int. Conf. Image Analysis and Recognition (ICIAR)*, 2004, pp.668-675.
  - [33] Liao S., Fan W., Chung A. C. S., and Yeung D. Y., “Facial expression recognition using advanced local binary patterns, tsallis entropies and global appearance features,” in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, 2006, pp.665-668.
  - [34] Zhao G. and Pietikäinen M., “Experiments with facial expression recognition using spatiotemporal local binary patterns,” in *Proc. Int. Conf. Multimedia and Expo (ICME)*, 2007, pp.1091-1094.
  - [35] Gritti T., Shan C., Jeanne V., and Braspenning R., “Local features based facial expression recognition with face registration errors,” in *Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition (FG)*, 2008.
  - [36] Hastie T., and Tibshirani R., and Friedman J., “The Elements of Statistical Learning; Data Mining, Inference, and Prediction”, Springer, 2nd edition, 2009.



### **Gawed M. Nagi**

Gawed M. Nagi, received his B.Sc. in computer science from Baghdad University, IRAQ and his Masters degree from New Mexico State University (NMSU), USA in 1996 and 2002 respectively. Currently he is a PH.D. student in the faculty of Computer Science and Information Technology at University Putra Malaysia (UPM). His research interests include Image processing, machine vision and Computer Graphics.



**Rahmita O. K. Rahmat**

Rahmita O. K. Rahmat, received the B.Sc. and M.Sc. degrees in Science Mathematics from University Science Malaysia (USM), in 1989 and 1994, respectively. During 1989 to 1990 she work as a research assistant in department of physics in University Science Malaysia experimenting on Ozone layer measurement at the Equatorial region, before working as tutor in Universiti Putra Malaysia (UPM). She received her PhD in Computer Assisted Engineering from University of Leeds, U.K. At this moment she is working in Faculty of Computer Science and Information Technology as lecturer. Among her focus research area are Computer Graphics and Applications, Computer Assisted Surgery and Computational Geometry.



**Fatimah Khalid**

Fatimah Khalid, obtained the B. Sc. in Computer Science from University Technology Malaysia (UTM) in 1992. During 1993 to 1995 she works as a System Analyst at University Kebangsaan Malaysia (UKM) and continued her Masters Degree from UKM. After getting her Masters Degree in 1997, she started involved in teaching at Sal College until 1999 and continued at the University Putra Malaysia in June 1999. She received her PhD in System Science and Management from the National University of Malaysia in 2008. At this moment, she is working in Faculty of Computer Science and Information Technology as a lecturer. Among her focus research area are Computer Vision and Image Processing, Content based Retrieval System and Computer Graphic Applications.



**Muhamad Taufik**

Muhamad Taufik, is a lecturer in the Department of Multimedia, Faculty of Computer Science and Information Technology, University Putra Malaysia (UPM). He obtained his first degree in Computer Science from UPM in 1990, masters degree in Computer Science from Universiti Teknologi Malaysia (UTM) in 1992, and doctor of philosophy from UPM in 2006. He joined UPM in 1990 as a tutor. He became a lecturer at Department of Computer Science in 1992. His research interest includes Information Retrieval, and Cross Language Information Retrieval.