Facial expression recognition in JAFFE and KDEF Datasets using histogram of oriented gradients and support vector machine

S K Eng*, H Ali, A Y Cheah, Y F Chong

School of Mechatronics Engineering, Universiti Malaysia Perlis (UniMAP), Kampus Pauh Putra, Arau, 02600, Perlis, MALAYSIA

*Corresponding author: kheng@unimap.edu.my

Abstract. This paper presents the used of histogram of oriented gradient (HOG) for facial expression recognition using support vector machine (SVM). In this work, the facial expression images are firstly preprocessed by face detection and cropped images. Then, HOG method is adopted as feature extraction on facial image. The ability of HOG to preserve the local information and orientation density distribution in facial images suitable as shape descriptor for facial expression. It divides the image into cell or patch that has magnitude and orientations. The extracted HOG was then concatenated into histogram bin to form one feature vector before feed into SVM classifier. Both JAFFE and KDEF datasets were employed to evaluate the performance of proposed method. Based on results, the average recognition rates of JAFFE and KDEF datasets are 76.19% and 80.95% respectively. The results show that the performance of expression surprise has outperformed compared to others expression while expression fear contributes the lowest recognition rate. Thus, utilization of HOG features with SVM classifier have shown the promising results in recognizing facial expression.

1. Introduction
Facial expression refers to the explicit transformation of human face due to the automatic response of the emotional states in fact it is induced by the activations of facial muscles that temporarily deformed both permanent and transient features. Temporal dynamics of facial muscle typically brief normally ranging between 5 seconds [1]. Facial expression plays an important role in human communication since 55% of facial modality involve in human-to-human interaction [2]. Nowadays, with advanced technology, facial expression recognition can be applied to education field, where the emotion state of the students in the classroom can be detected. The facial expression recognition might help in establishing a healthy learning and teaching environment.

Over the past decades, facial expression recognition becomes the prominent research in the field of computer vision, pattern recognition and image analysis. Many approaches constantly proposed by previous researchers. Some of them utilized facial components [3], facial points [4], facial landmarks [5], shape and texture [6], empirical mode decomposition [7], higher order spectra [8], neural-Adaboost [9] and local binary pattern [10] in recognizing facial expression.

Even though facial expression recognition has reached a certain level of success, however its development of a robust facial expression recognition is still ongoing and challenging as there are still many unsolved aspects due to various unpredictable facial variations and complicated exterior environment conditions. Therefore, this paper proposes histogram of oriented gradient of facial images.
using support vector machine to recognize the seven facial expressions. Although this paper similar to [11], however the pre-processed facial image, HOG [12] framework and the datasets used (JAFFE and KDEF) in the proposed method are varies one with [11]. In this work, face detection using Viola Jones method [13] is adopted and the detected face region is automatically cropped. Then, the pre-processed images were subjected to HOG as feature extraction. The rationale of using HOG is that it has ability to preserve the local information using orientation density distribution and gradient of the edge in facial images. Thus, extracted HOG features will used as input before fed to SVM classifier.

The outline of the paper is organized as follows: Section 2 presents brief methodology used in facial expression. It consists of pre-processing of facial image, extracting facial features and classifier. Section 3 provides the experimental results and discussion of facial expression recognition using HOG technique and SVM in classifying seven facial expressions. Section 4 concludes the proposed method.

2. Methodology
The flowchart of the proposed method is depicted in Figure 1. The details of each block are described in subsequent sections.

2.1. Facial Expression Database
Two benchmark databases were used in this work: JAFFE and KDEF database.

2.1.1. JAFFE Database. Japanese Female Facial Expression (JAFFE) database was taken from publicly available data which consists of 213 facial expression images of 10 subjects of Japanese female. Each subject performed six basic emotions plus neutral (30 angry, 29 disgust, 33 fear, 30 happiness, 31 sad, 30 surprises and 30 neutral) in which each expression contains 3 to 4 images per subjects. The image has the resolution in grayscale. All the facial images have been taken under strict controlled conditions of similar lighting and no occlusion such as hair or glasses. All the expression in frontal view and the resolution of the original image are 256 x 256 pixels.

2.1.2. KDEF Database. Karolinska Directed Emotional Faces (KDEF) database is another publicly available dataset consists a set of 4900 facial expression images. It contains 70 individuals, each displaying 7 different emotional expressions, each expression being photographed (twice) from 5 different angles.

2.2. Image Pre-Processing
Pre-processing of the facial image is considered as an important step in facial expression recognition. The image needs to be preprocessed, so that the final image will have extracted face region with uniformity of size and shape.

2.2.1. Face Detection. Face detection is a method used to determine the region of the face. In this work, Viola Jones face detection method [13] has been adopted. Figure 2 shows an example of face detection. Then, the original image is automatically cropped based on the face region obtained. Figure 3 shows the results of cropped images for both datasets. Further, the KDEF images which in RGB format, need to convert into grayscale images for further process. Finally, histogram equalization is performed on the images to enhance contrast of grayscale images and have uniform distribution of intensities that will aid in the feature extraction process.
Figure 2. Example of face detection using Viola Jones method \[13\]

(a) JAFFE dataset  
(b) KDEF dataset

Figure 3. Results of face detection and cropped images

2.3. Feature Extraction Based on Histogram of oriented Gradient

HOG \[12\] descriptor is widely used for human detection, object detection and pedestrian identification. HOG is computed using magnitude and orientation. It is based on the accumulation of gradient directions over the pixel of a small spatial region referred as “cell” and in the subsequent construction of a 1D histogram whose concatenation supplies the features vector to be considered for further purposes. Let \( L \) be an intensity (grayscale) function describing the image to be analyzed. The image is divided into cells of size \( N \times N \) pixels and the orientation \( \theta_{x,y} \) of the gradient in each pixel is computed by means of the following equation:

\[
\theta_{x,y} = \tan^{-1} \frac{L(x,y + 1) - L(x,y - 1)}{L(x + 1,y) - L(x - 1,y)}
\]  

(1)

Successively, the orientations \( \theta_{i,j} = 1 \ldots N^2 \) belonging to the same cell \( j \) are quantized and accumulated into a \( M \)-bins histogram. Finally, all the achieved histograms are ordered and concatenated into a unique HOG histogram that is the final outcome of this algorithmic step, i.e. the features vector to be considered for the subsequent processing. Figure 4 shows the HOG features extraction process based on cell size \( N \times N \) pixels:
In this HOG framework, the preprocessed facial image is divided into cells of size $N \times N$ pixels. The orientation of all pixels is computed and accumulated in an $M$-bins histogram of orientations. Then, all the cell histograms are concatenated to construct the final features vector. Figure 5 shows the illustration for different cell size of 32 x 32 orientations with cell size 16 x 16 orientations for the cell histograms.

Normally two main parameters are used to describe characteristics of the HOG descriptor i.e. orientation bines and cell size. The dimension of the patch involved in the single histogram computation is represented by cell size. Appropriate selection of parameters plays a vital role and should be selected with care for better performance of classification algorithm. Features extracted using HOG are affected by varying cell size. In the situation of using larger cell size, spatial information of certain region from facial image is squeezed into the unit cell histogram, and the contribution turn into less significant. High resolution analysis on the other hand can be carried out by using smaller cell size, but it results in more detailed information [11].

![Figure 4. Histogram of oriented gradient [14]](image)

**Figure 4.** Histogram of oriented gradient [14]

**Figure 5.** Illustration of different cell size with 32 x 32 orientations and 16 x 16 orientations

### 2.4. Support Vector Machines

The foundation of SVMs have been developed by [15] and are gaining keen interest due to attractive features and competitive performance. SVMs are based on structural Risk minimization (SRM) principle, which has been shown to be superior, to traditional empirical risk minimization. SRM minimizes an upper bound on the expected risk as opposed to ERM that minimizes the error on training data. It gives better generalization abilities which are the ultimate goal in the statistical learning. SVMs have been used widely in classification problems such as object recognition, speaker identification and face recognition. Given a set of labelled training data (supervised learning), the algorithm computes an optimal hyperplane (the trained model) which categorizes new examples in the right class. Given the training vectors $x_i \in \mathbb{R}^n$, $l = 1, ..., l$ and the corresponding set of $l$ labels $y_i \in \{1, -1\}$, the following primal optimization problem is solved:
\begin{equation}
\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i
\end{equation}

subject to \( y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, ..., l \)

where \( \xi_i \) is the misclassification error for the \( i \)th training vector; \( \xi \) is the total misclassification error, \( w \) is the normal vector to the hyperplane. Training vectors are mapped into higher dimensional space by the function \( \phi \). Furthermore, \( K(x_i, x_j) = \phi(x_i)^T \phi(x_j) \) is called the kernel function. The choice of kernel function is also significant when designing SVM-based classifier. The 2 basic kernel function are usually used:

Linear function: \( K(x_i, x_j) = x_i^T x_j \)

Radial basis function: \( K(x_i, x_j) = (y x_i^T x_j + \rho)^d, y > 0 \)

The classical SVM approach is suitable only for a two classes problem but, unfortunately, FER involves multi-class handling. Multi-class problems can be addressed by using “one-against-one” method as proposed by [16], the approach is based on the construction of an SVM classifier for each pairwise of classes and a voting system aided to elect the predicted class when an unseen item is tested. More specifically, let \( k \) be the number of classes, then \( k(k-1)/2 \) classifiers \( K_{ij} \) are trained from the available data of the \( i \)th and \( j \)th classes. The final prediction is returned by a voting system: the unseen example is analysed by each classifier \( K_{ij} \) that gives a vote to one or to the other class. Finally, the class with the largest number of votes is chosen.

Let considered 3-class problem. In this way, three classifiers came out, example, \( C_i(x), (i = 1,2,3) \). Calculate the testing data sample one by one with three classifiers, and pick the maximum value \( C_i(x), (i = 1,2,3) \), so the testing data sample belongs to the \( i \)th class. Figure 6 shows the examples of SVM classifier for one-against-one and one-against-all.

![Figure 6. Examples of SVM classifier for one-against-one and one-against-all [17]](image)

3. Results & Discussion

The proposed method is evaluated using two publicly available datasets, JAFEE and KDEF. In JAFEE, a total of 213 images has been employed. While, in KDEF, a total of 980 images have been adopted. All the images have been subjected to image preprocessing such as face detection and facial cropped. Then, the HOG features were extracted and used as input to predict the expression using SVM classifier.

In this experiment, the data was split into training and testing set based on conventional method with 70% training and 30% testing strategy. Figure 7 shows the recognition rates based on extracted HOG features using SVM classifier on JAFFE and KDEF datasets. Based on Figure 7, it can be seen that the expression surprise achieved the highest recognition rate in both datasets which are 88.89% (JAFFE) and 97.62% (KDEF). This can be inferred that, the HOG as shape descriptor able to counts occurrences of gradient orientations in localized portions of an image and able to model the shape of the surprise facial muscles via edge information effectively compared to others facial expression. In contrast,
expression fear shows the lowest recognition rate in both JAFFE and KDEF datasets which is 55.56% and 52.38% respectively. On the other hand, the expression disgust in JAFEE and KDEF datasets shows the recognition rate of 66.67% and 76.19%, respectively. While, expression happy, neutral and sadness are comparable with recognition rates of 77.78% in JAFEE dataset.

![Graph showing recognition rates of seven facial expressions using SVM classifier in JAFFE and KDEF datasets.](image)

**Figure 7.** Recognition rates of seven facial expression using SVM classifier in JAFFE dataset and KDEF dataset

As summary, the recognition rate of expression surprise gave the best results in both JAFFE and KDEF datasets. The features of expression surprise may be easily distinguishable. Meanwhile, the recognition rate of expression fear shows the lowest results in both datasets. It can be inferred that; the expression fear is difficult to recognize due to the existence of intra-class variations of different expression between the subjects. Thus, the system confused to classify the expression fear correctly.

4. Conclusion

This paper presented the used of histogram-oriented gradient as feature descriptor of facial image using SVM classifier to classify the seven facial expressions. In this work, the facial expression images are firstly preprocessed by face detection and cropped images. Then, HOG method is adopted as feature extraction on facial image. The ability of HOG to preserve the local information and orientation density distribution in facial images able to distinguish the characteristic of facial expression. The extracted HOG was then concatenated into histogram bin to form one feature vector before subjected as input to SVM classifier. Both JAFFE and KDEF dataset were employed to evaluate the proposed method. Based on results, average recognition rates of JAFEE and KDEF datasets are 76.19% and 80.95%, respectively. It can be seen that the expression surprise has shown the highest recognition rate while fear provide the lowest recognition rate. Further, the average recognition rates of KDEF dataset (80.95%) is slightly higher than JAFEE dataset (76.19%). The results show that the performance of expression surprise has outperformed others expression while expression fear slightly hard to recognize thus contribute the lowest rate. As overall, utilization of HOG features with SVM classifier have shown the promising results in recognizing facial expression.

However, further study needs to be conducted to improve the recognition rate of expression fear by investigating the several HOG features for different magnitude and orientation as well as utilizing different machine learning classifiers to enhance the performance of the proposed method.
References

[1] Fasel B and Luettin J 2003 Automatic facial expression analysis: A survey *Pattern Recognition* **36** 259–275

[2] Mehrbadian A 1968 Communication without words *Psychological Today* 8–9

[3] Ilbeygi M and Shah-Hosseini H 2012 A novel fuzzy facial expression recognition system based on facial feature extraction from color face images *Engineering Applications of Artificial Intelligence* **25** 130–146

[4] Pantic M and Rothkrantz L J M 2000 Expert system for automatic analysis of facial expressions *Image and Vision Computing* **18** 881–905

[5] Aifanti N and Delopoulos A 2014 Linear subspaces for facial expression recognition *Signal Processing: Image Communication* **29** 177–188

[6] Kotsia I, Zafeiriou S and Pitas I 2008 Texture and shape information fusion for facial expression and facial action unit recognition *Pattern Recognition* **41** 833–851

[7] Ali H, Hariharan M, Yaacob S and Adom A H 2015 Facial emotion recognition using empirical mode decomposition *Expert Systems with Applications* **42** 1261-1277

[8] Ali H, Hariharan M, Yaacob S and Adom A H 2015 Facial Emotion Recognition Based on Higher Order Spectral using SVM *Journal of Medical Imaging and Health Informatics* **5** 1272-1277

[9] Owusu E, Zhan Y and Mao Q R 2014 A neural-AdaBoost based facial expression recognition system. *Expert Systems with Applications* **41** 3383–3390

[10] Luo Y, Wu C and Zhang Y 2013 Facial expression feature extraction using hybrid PCA and LBP *The Journal of China Universities of Posts and Telecommunications* **20** 120–124

[11] Nazir M, Jan Z and Sajjad M 2018 Facial expression recognition using histogram of oriented gradients based transformed features *Cluster Computing* **21** 539-548

[12] Dalal N and Triggs B 2005 Histograms of oriented gradients for human detection *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **1** 886-893

[13] Viola P and Jones M 2001 Rapid Object Detection using a Boosted Cascade of Simple Features *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01)* **1** 511-518

[14] Del Coco M, Carcagni P, Leo M and Distante C 2015 A Minimax Framework for Gender Classification Based on Small-Sized Datasets *International Conference on Advanced Concepts for Intelligent Vision Systems* 415-427

[15] Vapnik V 1998 Statistical Learning Theory *John Wiley and Sons Inc New York*

[16] Knerr S, Personnaz L and Dreyfus G 1990 Single-layer learning revisited: A stepwise procedure for building and training neural network *Neurocomputing: Algorithms, Architectures and Applications, NATO ASI, Berlin: Springer-Verlag* 41-50

[17] Ying Z and Zhang G 2009 Facial Expression Recognition Based on NMF and SVM *International Forum on Information Technology and Applications, Chengdu* 612-615