

# Facial Expression Recognition Based on Perceived Facial Images and Local Feature Matching

Hayet Bouhrara<sup>1</sup>, Liming Chen<sup>2</sup>, Chokri Ben Amar<sup>3</sup>,  
and Mohamed Chtourou<sup>1</sup>

<sup>1</sup> University of Sfax, Control & Energy Management Laboratory,  
Sfax Engineering School, POB 3038 Sfax, Tunisia  
hayet.bouhrara@laposte.net, mohamed.chtourou@enis.rnu.tn

<sup>2</sup> University of Lyon, Laboratoire d'Informatique en Image et Systèmes  
d'information, Central School of Lyon, 69134 Ecully, France  
liming.chen@ec-lyon.fr

<sup>3</sup> University of Sfax, Research Group on Intelligent Machines,  
Sfax Engineering School, POB 3038 Sfax, Tunisia  
chokri.benamar@enis.rnu.tn

**Abstract.** Facial expression recognition is to determine the emotional state of the face regardless of its identity. Deriving an effective facial representation from original face images is a vital step for successful facial expression recognition. This paper presents a biological vision-based facial description, called Perceived Facial Images “PFI” applied to facial expression recognition. For the classification step, Scale Invariant Feature Transform “SIFT” is used to extract a local feature in images. Then, a matching computation is processed between a testing image and all train images for recognizing facial expression. To evaluate, the proposed approach is tested on the GEMEP FERA 2011 database and the Cohn-Kanade Facial Expression database. To compare, the developed algorithm achieves better experimental results than the other approaches in the literature.

**Keywords:** Facial Expression Recognition, Feature extraction, Perceived Facial Images “PFI”, SIFT, matching.

## 1 Introduction

Artificial recognition of facial expression has attracted a lot of attention in the last few years. Due to the potential applications of human-computer interface, multimedia, and surveillance, automatic facial expression recognition has become a hot topic in the communities of computer vision and pattern recognition. Six emotions are universally associated with distinct facial expressions: happiness, sadness, surprise, fear, anger and disgust [1][2]. Thus, most of the works focus on classifying an input facial image or sequence into one of these basic expressions.

Facial expression recognition is based on two steps which are the feature extraction and the classification [3]. In general, there are three common approaches

to extract facial feature: geometric feature based methods, appearance based methods and hybrid methods [4][5]. Different techniques have been proposed to classify facial expressions [6][7].

The biological vision-based facial description, namely Perceived Facial Images “PFI”, aims at increasing the robustness of face recognition system to the changes of facial expressions [8]. Thus, this paper proposes to use the “PFI” which are an appearance based feature, applied to facial expression recognition. The proposed “PFIs” simulate the response of complex neurons to gradient information within a certain neighborhood and possess properties of being highly distinctive as well as robust to illumination and geometric transformations [9][10]. Based on this intermediate facial description, SIFT-based local feature matching is utilized to compute similarity scores between test and train face images.

Firstly, the Perceived Facial Images “PFI” is introduced for facial expression recognition application. Then, the following “SIFT” based local feature matching process is shown. Finally, the experimental results achieved on the GEMEP FERA 2011 database and the Cohn-Kanade facial expression database are presented, analyzed and compared with the literature.

## 2 Biological Vision-Inspired Facial Description

Extracting an efficient representation of the face from images is an important step for successful facial expression recognition. The present paper suggests making use of a biological vision-based facial description namely Perceived Facial Images “PFI”, which was initially applied to 3D face recognition [9][11].

“PFI” aims at giving a visual representation simulating the human visual perception. The “PFIs” were inspired by the study of Edelman et al.[12], who proposed a representation of complex neurons in primary visual cortex. These complex neurons respond to a gradient at a particular orientation and spatial frequency, but the location of the gradient is allowed to shift over a small receptive field rather than being precisely localized.

### 2.1 Representation of the Complex Neuron Response

The proposed representation “PFI” aims at simulating the response of complex neurons, based on a convolution of gradients in specific directions in a given circular neighborhood. The precise radius value of the circular area needs to be experimentally fixed. Specifically, given an input image  $I$ , a certain number of gradient maps  $L_1, L_2, \dots, L_o$ , one for each quantized direction “ $o$ ”, are first computed. They are defined as:

$$L_o = \left(\frac{\partial I}{\partial o}\right)^+ \quad (1)$$

The “+” sign means that only positive values are kept to preserve the polarity of the intensity changes.

Each gradient map describes gradient norms of the input image in a direction “ $o$ ” at every pixel location. Then, the response of complex neurons is simulated by convolving its gradient maps with a Gaussian kernel  $G$ . The standard deviation of  $G$  is proportional to the radius of the given neighborhood area,  $R$ , as in (2).

$$\rho_o^R = G_R * L_o \tag{2}$$

The purpose of the convolution with Gaussian kernels is to allow the gradients to shift within a neighborhood without abrupt changes. At a given pixel location  $(x, y)$ , we collect all the values of the convolved gradient maps at that location and form the vector  $\rho^R(x, y)$  thus having a response value of complex neurons for each orientation “ $o$ ”.

$$\rho^R = [\rho_1^R(x, y), \rho_2^R(x, y), \dots, \rho_o^R(x, y)] \tag{3}$$

This vector,  $\rho^R(x, y)$ , is further normalized to unit norm vector, which is called the subsequent response vector and denoted by  $\underline{\rho}^R$ .

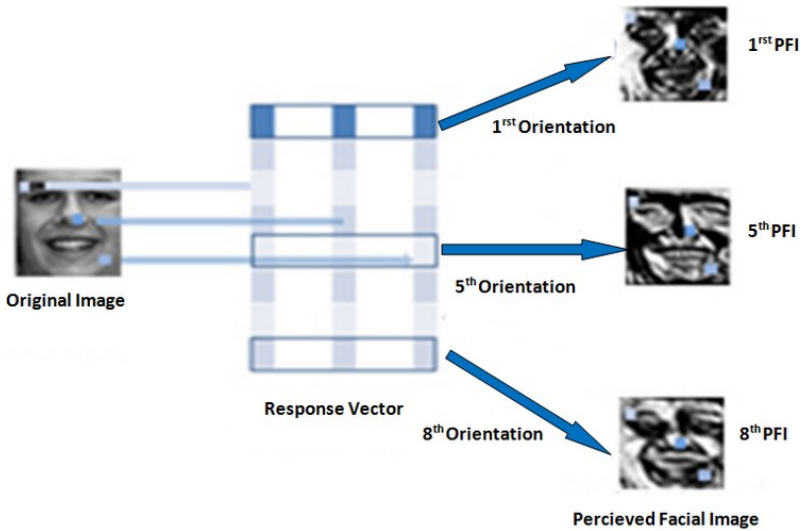


Fig. 1. An illustration of the “PFI” in 8 orientations

### 2.2 Facial Description by Response Vectors

Now a facial image can be represented by its perceived values of complex neurons according to the response vectors. Specifically, given a facial image  $I$ , Perceived Facial Images  $J_o$  are generated using complex neurons for each orientation “ $o$ ” defined as in (4).

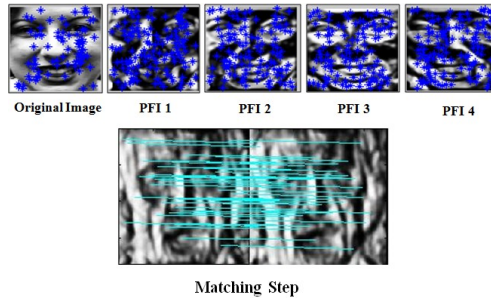
$$J_o(x, y) = \underline{\rho_o^R}(x, y) \quad (4)$$

Fig. 1 illustrates the process applied to a facial image. In this work, eight “PFIs” for eight quantized directions are computed.

### 3 SIFT Based Local Feature Matching

Scale Invariant Feature Transform “SIFT” [13] algorithm has been proposed for extracting a local descriptor of images features. The SIFT features are invariant to image scaling, rotation, and occlusion to a certain extent in illumination. The SIFT technique works by first detecting a number of keypoints in the given image and then computing local image descriptor of these keypoints [13]. For each detected keypoint, a feature vector with a dimension of 128 is extracted.

In this study, the SIFT descriptor is applied on each “PFI” separately. PFIs highlight local shape and texture characteristics of smooth facial images, many more SIFT-based keypoints can be detected for the following matching step than those in the original facial image. Fig.2 illustrates the Keypoint detection on the original facial expression image along with the “PFI” in eight directions.



**Fig. 2.** The upper row lists the SIFT keypoint detection for the original image and its first four “PFI”. The bottom one displays an example of the features matched in faces.

SIFT is applied in each image in the training and the test set. Then, each keypoint descriptor extracted from the test image is matched to the set of descriptors extracted from every image in the training set. The best match for each descriptor is found by identifying its nearest neighbor in the set of keypoint descriptors from the training image. To discard keypoints whose descriptors do not have any good match to a training image, a subsequent threshold “ $t$ ” is used, which rejects matches that are too ambiguous. If the distance ratio between the closet neighbor and the second-closest neighbor is below some threshold “ $t$ ”, the match is kept; otherwise the match is rejected (Fig. 2). In this study, “ $t$ ” is empirically set at 0.6 as in [13]. The image in the database with the largest number of matching points is considered as the matched facial expression.

## 4 Facial Expression Recognition Results

The proposed approach is experimentally evaluated using two databases. The first part of this section is devoted to the obtained results using the GEMEP FERA 2011 database and the second part for the Cohn-Kanade Database.

### 4.1 Results Using the GEMEP FERA 2011

**Database.** The first database used in this study is a partition of the GEMEP corpus [14], developed by the Geneva Emotion Research Group at the University of Geneva. The challenge is divided in an action unit detection sub-challenge and an emotion detection sub-challenge. In this work, the emotion detection is considered, which calls for systems to attain the highest possible classification rate for the detection of five discrete emotions: anger, fear, joy, relief, and sadness.

The data set is divided into a training set and a test set. A total of 289 portrayals were selected (155 for training and 134 for testing). The test data is divided in three different partitions. The first is the partition which the test subjects are not present in the training data (Person Independent partition “PI”). The second partition consists of videos of subjects that are also part of the training set (Person Specific partition “PS”). The third partition is simply the entire (“overall”) test set [14].

In this work, the same pre-processing using in [14] is applied. In fact, the OpenCV face detector is used to extract the face location in each frame.

**Experimental Results.** The emotion detection can detect five discrete emotion classes. Each video has a single emotion label  $e \in E$ , where  $E = \{\text{Anger, Fear, Joy, Relief, Sadness}\}$ . The classification rate is first obtained per emotion, and then the average over all 5 emotions is computed. The classification rate for emotions is calculated based on a per-video prediction. It is calculated per emotion as the fraction of the number of videos correctly classified as that emotion divided by the total number of videos of that emotion in the test set. To decide the label of a test video of  $n$  frames, we find the emotion with the largest number of frames classified.

The “PFI” in eight directions is applied for each frame of a video train and test data. Then, the “SIFT” is used to extract a feature vector for each frame of a video train and test database. For each orientation, the classification rates are computed in each partition of test set.

On the first hand, for the “PS” test set the obtained results are the same in each direction. The classification rate is equal to 100%. After that, fusion which is the sum of the obtained results of the matching step for each orientation of the “PFI”, is computed. Thus, we obtained the same results using fusion (100%).

On the second hand, the performance is computed for “PI” and “overall” test set for each direction. The classification rates by the proposed method for the emotion detection are shown in Table 1.

**Table 1.** Classification Rates (%) on “PI” and “Overall” test set

		Anger	Fear	Joy	Relief	Sadness	Average
“PFI 1”	PI	64.2	33.3	95	50	46.6	60
	Overall	81.48	60	96.77	69.23	68	76.11
“PFI 2”	PI	57.1	46.6	100	50	13.3	56.25
	Overall	77.77	68	100	69.23	48	73.88
“PFI 3”	PI	50	33.3	95	43.7	13.3	50
	Overall	74.07	60	96.77	65.38	48	70.14
“PFI 4”	PI	50	46.6	95	43.7	40	57.5
	Overall	74.07	68	96.77	65.38	64	74.62
“PFI 5”	PI	42.8	26.6	100	37.5	33.3	51.25
	Overall	71.4	56	100	68.7	60	74
“PFI 6”	PI	50	40	95	50	20	53.75
	Overall	74.07	64	96.77	69.23	52	72.38
“PFI 7”	PI	35.7	33.3	100	50	40	55
	Overall	66.66	60	100	69.23	64	73.13
“PFI 8”	PI	50	26.6	100	62.5	33.3	57.5
	Overall	74.07	56	100	76.92	60	74.62

**Table 2.** Confusion Matrix for “PI” test set using Fusion

	Anger	Fear	Joy	Relief	Sadness
Anger	7	1	0	1	5
Fear	0	4	0	3	2
Joy	2	7	20	1	0
Relief	2	3	0	8	1
Sadness	3	0	0	3	7

Comparing the obtained results for each direction, the best recognition performance (Average=76.11%) is obtained by using the “PFI” on direction “1”. After that, the fusion is computed. Table 2 illustrates the confusion matrix for the “PI” test set using the fusion.

Based on these results, the classification rate on “PI” test set using simple fusion is equal to 57.5%. But the obtained rates using the “PFI 1” are the best. Also, noting that the obtained results on facial expression “Fear” are quite a bit worse. But, the obtained results on facial expression “Joy” are more convincing.

**Comparison with the Literature.** GEMEP FERA database are used for facial expression recognition for many researches. Table 3 gives performance comparisons between the proposed method and the existing systems using GEMEP FERA database. The obtained rates using the “PFI 1” are considered on the Table 3.

**Table 3.** Comparison with the Literature using GEMEP FERA database

Systems	Recognition rate (%)		
	PI	PS	Overall
Dhall et al.[15]	64.9	83.8	73.4
Dahmane et Meunier[16]	57.9	87	70
Littlewort et al. [17]	71.4	83.7	76.1
Meng et al.[18]	60.9	83.7	70
Srivastava et al.[19]	63.6	73	67.2
Tariq et al.[20]	65.5	100	79.8
Yang et Bhanu [21]	75.2	96.2	83.8
Baseline [14]	44	73	56
Proposed methods: “PFI” + SIFT	60	100	76.11

Based on the Table 3, the obtained classification rates on the “PI” test set is not the best compared to other methods. This is explained by the difficulty of a facial expression recognition of a subject who has not used in the training set. Therefore, must be used high database formed by several images. But, the proposed approach gives the best rate on the “PS” test set comparing with the other methods.

#### 4.2 Results Using the Cohn-Kanade Facial Expression Database

**Database.** The second database used in this study is the Cohn-Kanade facial expression database [22]. This database consists of 97 students. For each subject from the Cohn-Kanade database, six expressive faces (happiness, anger, sadness, surprise, fear, and disgust) have been selected. The pre-processing steps described before are applied on all the images of the database.

**Table 4.** Classification Rates (%) for Emotion Recognition using the Cohn-Kanade Database

	Anger	Disgust	Fears	Happy	Sadness	Surprise	Average
“PFI 1”	99.34	96.04	98.38	98.87	98.40	98.50	98.25
“PFI 2”	100	94.99	97.98	98.70	98.86	97.95	98.08
“PFI 3”	99.56	95.78	96.76	98.70	97.95	96.83	97.59
“PFI 4”	99.34	94.46	97.77	99.02	97.95	97.20	97.62
“PFI 5”	99.78	95.51	96.56	98.86	98.63	97.95	97.88
“PFI 6”	99.78	94.46	97.57	99.02	98.86	97.95	97.94
“PFI 7”	99.56	95.25	98.79	99.19	98.40	98.50	98.28
“PFI 8”	99.34	96.83	98.79	99.02	98.63	98.50	98.51

**Experimental Results.** The proposed approach is repeated using the Cohn-Kanade database which is divided in the train set and test set. Eight “PFIs” are computed for each image separately. Then, SIFT is proposed to extract the features of images. For each direction of “PFI”, the classification rates are computed on each facial expression and are presented in the Table 4.

Comparing the obtained rate for each orientation, the best result is obtained using the “PFI” on direction “8”. In addition, the confusion matrix is calculated using the fusion and is illustrated on the Table 5.

**Table 5.** Confusion Matrix (%) using the Fusion on the Cohn-Kanade database

	Anger	Disgust	Fear	Happy	Sadness	Surprise
Anger	99.78	1.05	0.41	0	0.23	0.56
Disgust	0	95.78	0.20	0.33	0	0
Fear	0.22	0.80	97.98	0.48	0.91	0.19
Happy	0	1.05	0.81	99.19	0	0.37
Sadness	0	1.32	0.60	0	98.63	0.37
Surprise	0	0	0	0	0.23	98.50

It can be noted that the classification rate is equal to 98.31% using the fusion. To conclude, The obtained rate using the “PFI 8” is better than the fusion. However, the obtained rates on facial expression “Disgust” are very low compared to other. But, the obtained rates on facial expression “Anger” are more convincing. Also, the obtained recognition rate using the Cohn-Kanade database is better than the rate using the GEMP FERA database. This is explained by the effectiveness of using a big database formed by an enough number of images.

**Comparison with the Literature.** In general, it is difficult to fairly compare the obtained results for facial expression recognition with other related works. Because applying the same protocol is difficult in the classification step. Table 6 shows performance comparisons between the proposed method and the existing systems on facial expression recognition using the Cohn-Kanade database.

**Table 6.** Comparison with the Literature using Cohn-Kanade Database

Systems	Recognition rate
IMG [23]	89.35 %
LBP [23]	96.57 %
Gabor [23]	95.38 %
Zernike Moment [24]	73.20 %
Radial encoding of local Gabor features [25]	91.51%
Pyramid histogram of orientated gradient [26]	96.33%
Differential optical flow [27]	95.45%
Proposed methods: “PFI” + SIFT	98.31 %



It is obvious that the suggested approach, applying Perceived Facial Images “PFI” and SIFT, leads to the best classification rate (98.31%), which is higher than the other methods.

## 5 Conclusion

In this paper, a biological vision-based facial description, namely Perceived Facial Images “PFI” has been applied in facial expression recognition system. The proposed “PFI” simulate the response of complex neurons to gradient information within a pre-defined neighborhood and possess properties of being highly distinctive as well as robust to illumination and geometric transformations.

Based on this intermediate facial description, SIFT-based local feature matching is then used to compute similarity scores between train set and test set. The contribution to the literature is that the SIFT technique is tested on a image representation “PFI” by applying to facial expression recognition. The proposed approach is evaluated on the GEMEP FERA 2011 database and the Cohn-Kanade facial expression database. Promising experimental results are achieved which prove its effectiveness for facial expression recognition.

## References

1. Ekman, P.: Facial expressions of emotion: an old controversy and new findings. *Philos. Trans. R. Soc.* 335, 63–69 (1992)
2. Meftah, I.T., Thanh, N.L., Ben Amar, C.: Sharing Emotional Information Using A Three Layer Model. In: *Sixth International Conference on Internet and Web Applications and Services*, St. Maarten, The Netherlands Antilles, pp. 130–135 (2011)
3. Jemai, O., Zaied, M., Ben Amar, C., Alimi, M.A.: Pyramidal Hybrid Approach: Wavelet Network with OLS Algorithm Based-Image Classification. *International Journal of Wavelets, Multiresolution and Information Processing* 9(1), 111–130 (2011)
4. Zaied, M., Said, S., Jemai, O., Ben Amar, C.: A novel approach for face recognition based on fast learning algorithm and wavelet network theory. *Inter. Journal of Wavelets, Multiresolution and Information Processing* 9, 923–945 (2011)
5. Mejdoub, M., Ben Amar, C.: Classification improvement of local feature vectors over the KNN algorithm. *International Journal “Multimedia Tools and Applications”*, 1–22 (2011) ISSN 1380-7501
6. Meftah, I.T., Thanh, N.L., Ban Amar, C.: Multimodal Recognition of Emotions using a Formal Computational Model. In: *International Conference on Signal Image Technology & Internet Systems*, Naples, Italy, pp. 541–546 (2012)
7. Meftah, I.T., Thanh, N.L., Ben Amar, C.: Emotion Recognition using KNN Classification for User Modeling and Sharing of Affect States. In: *International Conference on Neural Information Processing*, Doha, Qatar, pp. 234–242 (2012)
8. Boughrara, H., Chen, L., Ben Amar, C., Chtourou, M.: Face Recognition under varying Facial Expression Based on Perceived Facial Images And Local Feature Matching. In: *Inter. Conference on Information Technology and e-Service*, Tunisia (2012)

9. Huang, D., Ben Soltana, W., Ardabilian, M., Wang, Y.H., Chen, L.: Textured 3D face recognition using biological vision-based facial representation and optimized weighted sum fusion. In: Workshop on Biometrics. Colorado Springs, CO (2011)
10. Said, S., Ben Amor, B., Ben Amar, C., Daoudi, M.: Fast and Efficient 3D Face Recognition using Wavelet Networks. In: International Conference on Image Processing, Egypt, pp. 4153–4156 (2009)
11. Ben Soltana, W., Ardabilian, M., Chen, L., Ben Amar, C.: A mixture of Gated Experts Optimized using Simulated Annealing for 3D Face Recognition. In: International Conference on Image Processing, Art. N 6116304, Belgium, pp. 3037–3040 (2011)
12. Edelman, S., Intrator, N., Poggio, T.: Complex cells and object recognition (1997) (unpublished manuscript), <http://kybele.psych.cornell.edu/edelman/archive.html>
13. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 60, 91–110 (2004)
14. Valstar, M.F., Jiang, B., Mehu, M., Pantic, M., Scherer, K.: The First Facial Expression Recognition and Analysis Challenge. In: Conference on Automatic Face and Gesture Recognition, California (2011)
15. Dhall, A., Asthana, A., Goecke, R., Gedeon, T.: Emotion recognition using phog and lpq features. In: Int'l Conf. Automatic Face and Gesture Analysis, pp. 878–883 (2011)
16. Dahmane, M., Meunier, J.: Emotion recognition using dynamic gridbased hog features. In: Int'l Conf. Automatic Face and Gesture Analysis, pp. 884–888 (2011)
17. Littlewort, G., Whitehill, J., Wu, T., Butko, N., Ruvolo, P., Movellan, J., Bartlett, M.: The motion in emotion—a cert based approach to the fera emotion challenge. In: Int'l Conf. Automatic Face and Gesture Analysis, pp. 897–902 (2011)
18. Meng, H., Romera-Paredes, B., Berthouze, N.: Emotion recognition by two view svm 2k classifier on dynamic facial expression features. In: Int'l Conf. Automatic Face and Gesture Analysis, pp. 854–859 (2011)
19. Srivastava, R., Roy, S., Yan, S., Sim, T.: Accumulated motion images for facial expression recognition in videos. In: Int'l Conf. Automatic Face and Gesture Analysis, pp. 903–908 (2011)
20. Tariq, U., Lin, K.-H., Li, Z., Zhou, X., Wang, Z., Le, V., Huang, T., Lv, X., Han, T.: Emotion recognition from an ensemble of features. In: Int'l Conf. Automatic Face and Gesture Analysis, pp. 872–877 (2011)
21. Yang, S., Bhanu, B.: Facial expression recognition using emotion avatar image. In: Int'l Conf. Automatic Face and Gesture Analysis, pp. 866–871 (2011)
22. Kanade, T., Cohn, J., Tian, Y.L.: Comprehensive database for facial expression analysis. In: Int'l Conference on Automatic Face and Gesture Recognition (2000)
23. Xiao, R., Junzhao, Q., Zhang, D., Shi, P.: Facial expression recognition on multiple manifolds. *Pattern Recognition* 44, 107–116 (2011)
24. Mehdi, S., Lajevardi, Z., Hussain, M.: Higher order orthogonal moments for invariant facial expression Recognition. *Digital Signal Processing* 20, 1771–1779 (2010)
25. Gu, W., Xiang, C., Venkatesh, Y.V., Huang, D., Lin, H.: Facial expression recognition using radial encoding of local Gabor features and classifier synthesis. *Pattern Recognition* 45, 80–91 (2012)
26. Li, Z.S., Imai, J., Kaneko, M.: Facial expression recognition using facial component based bag of words and PHOG descriptors. *Information and Media Technologies* 5(3), 1003–1009 (2010)
27. Sanchez, A., Ruiz, J.V., Moreno, A.B., Montemayor, A.S., Hernandez, J., Pantrigo, J.J.: Differential optical flow applied to automatic facial expression recognition. *Neurocomputing* 74, 1272–1282 (2011)