ISSN 2464-4617 (print) WSCG 2016 - 24th Conference on Computer Graphics, Visualization and Computer Vision 2016
ISSN 2464-4625 (CD-ROM)

Facial expression recognition using salient facial patches

Hazar Mliki
MIRACL-ENET'COM
University of Sfax
Tunisia (3018), Sfax
mliki.hazar@gmail.com

Mohamed Hammami
MIRACL-FSS
University of Sfax
Tunisia (3018), Sfax
mohamed.hammami@fss.rnu.tn

ABSTRACT

This paper proposes a novel facial expression recognition method composed of two main steps: offline step and online step. The offline step selects the most salient facial patches using mutual information technique. The online step relies on the already selected patches to identify the facial expression using an SVM classifier. In both steps, the LBP operator was used to extract facial expressions features. Through an extensive experiments on the JAFFE and KANADE databases, we have shown that our method, thanks to the salient selected patches, has the advantage of being much faster with a significant gain in recognition performance.

Keywords

machine vision, facial expression recognition, Mutual Information, LBP

1 INTRODUCTION

In recent years, there has been an increasing interest on facial expressions recognition as it is one of the most important cues to our emotional state [VTG$^+$15]. In fact, by analysing the emotional state of one person, we can easily extract information about its mood, feeling and personality. Therefore, facial expressions recognition has been involved in many computer vision applications, like surveillance systems, human-machine interaction, gaming and remote monitoring of patients [SGA09]. Although the continued research interest on facial expressions topic, recognizing facial expression with a high accuracy remains a challenging task due to the variation of facial expressions across human culture and to the context-dependent variation even for the same person.

Developing an efficient facial representation from face images is a key step to succeed facial expression recognition task. Actually, facial expression recognition includes two main stages: the facial feature extraction and the classification strategy. Facial feature extraction consists of deriving features which maximize between class variations whereas minimize within class variation of facial expressions. Hence, facial expressions recognition performance depends heavily on the choice of features used by the classifier. Relying on the way how facial features are extracted for classification, previous methods for facial expression recognition can be classified into two main approaches: geometric approach and global approach.

Geometric approach is based on the shape and locations of facial components such as the mouth, the nose, the eyes and the eyebrows. Then, the different distances between feature points and the relative sizes of the major face components are computed to form a feature vector. For instance, in [LBA99b] [GD03], the authors applied a geometric position of 34 manually selected points and a set of Gabor wavelet coefficients at these points. Some other authors [Ham06] compute relative distances to encode the geometric distance variations. Other [SJD08] used the geometric feature extracted by Active Appearance Model to perform facial expression recognition. Geometry approach is more robust to scale, size, head orientation variation. However, it requires reliable facial feature detection, which is a challenging task. Thus, most of the above cited methods, mainly [LBA99b] [GD03], require a manual selection of facial points which is not suitable for the autonomy aspect of the method. Moreover, facial features are unable to encode facial texture change such as wrinkles and furrows which are important for facial expression modeling.
In contrast, global approach encodes the appearance texture of the whole face which includes wrinkles, bulges and furrows. In this context, image filters are applied to the whole face so as to extract facial appearance variation which usually generates a high-dimensional feature vector. Accordingly, some

subspace learning methods such as principal component analysis (PCA) [DC99] and its independent form (PCI) [DGG06] are frequently performed to build new low subspace representation of the original face image. Then, matching is performed within the new subspace.

To sum up, we notice that geometric methods provide good perceptive justification for facial expression recognition. However, they depend on the accurate detection of facial features and require space costs for computation. Nonetheless, global methods inspect the appearance face variations which make them powerful to extract the discriminative information. Taking all this into account, we introduce in this paper a new method for facial expressions recognition which belongs to the global approach.

The remaining parts of the paper are organized as follows: in section 2, we introduce the proposed method, then we discuss the experiments in section 3 and conclude this paper in section 4.

2 PROPOSED WORK

The proposed method is based on psychological studies [Mag07] which show that some facial muscles are responsible of facial expressions appearance. These facial muscles are mainly located around some facial features such as the mouth, the nose and the eyes. The proposed method aims to define automatically the salient facial patches responsible of the local facial appearance variations. The proposed method is composed of two steps: offline step and online step. Both steps are performed after locating the face region using Viola and Jones Face detector [VJ04]. The offline step selects the most salient facial patches using mutual information technique. The online step relies on the already selected patches to identify the facial expression using an SVM classifier. In both steps, the extraction of the feature vector is carried out by LBP operator. The choice of such an operator is motivated by the fact that the face can be perceived as a combination of micro-models (patches). Such process allows managing the local variations of the face mainly due to illumination variation. Figure 1 describes the proposed method.

Our main contributions are:

- Automatic selection of the most salient face patches including the most discriminant descriptors to recognize facial expressions. Unlike the existing works which used manual and unpresice regions selection methods [FJJ09] [ST08] [LP12], we introduced a new algorithm based on Mutual Information technique to select automatically the descriptive patches. The identification of such patches reduces the complexity of the proposed method and thus accelerates the recognition process.
- Genericity of the selected facial patches. In fact, these patches are independent from the face images database and the used descriptor.

2.1 The off-line Step

This step seeks to select the active salient patches which are responsible of the facial expression deformation and appearance. Thus, we computed the facial feature vector using LBP operator. Then, we adapted the mutual information technique to select the most discriminant patches.

To extract facial expression features, we used the texture information by applying the LBP operator [OPH96]. We choose this operator thanks to its simplicity of computation which allows analysing images in real time as well as its invariance to rotation and illuminations variations. The LBP features are fast derived in a single scan through the raw image, whilst still including enough facial information in a compact representation.
After detecting the face region, we converted it to a grayscale image and applied an elliptical mask to get rid of hair, neck and all the noise that can appear jointly with the face. Thereafter, for a 64×64 pixels face region [LFCY06], we divided it into 64 patches each one is sized of 8×8 pixels. Finally, we coded each patch with an LBP histogram of 256 bins.
The choice of the number of patches is discussed in the experimental section. Figure 2 shows the process of feature vector extraction.

The selection of the optimal actives patches is the key point in our solution as it defines the quality and the performance of our method. The assumption here is that some patches may be insignificant, correlated or irrelevant and consequently, it would be interesting to remove them from the recognition process.
We have adapted the mutual information technique to select patches involving the most discriminant information for facial expressions recognition task. The mutual information (also called cross-entropy or gain-information) is a method of features selection widely used to measure the stochastic dependence of two discrete and random features [Soo00]. The mutual information between two variables x and y is defined based on their joint probabilistic distribution $p(x,y)$ and the respective marginal probabilities $p(x)$ and $p(y)$ as follow:

$$I(X,Y) = \int_{\Omega_Y}\int_{\Omega_X} p(x,y) log_2 (\frac{p(x,y)}{p(x)p(y)}) dxdy \qquad (1)$$

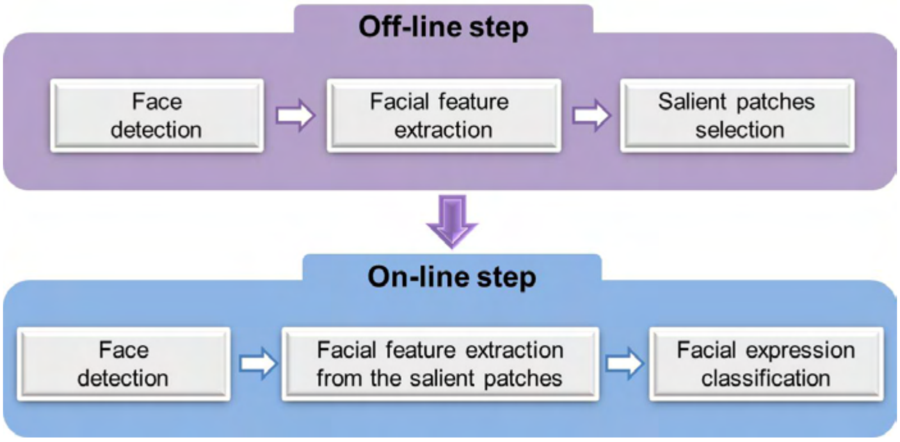

Figure 1: The proposed method for facial expression recognition

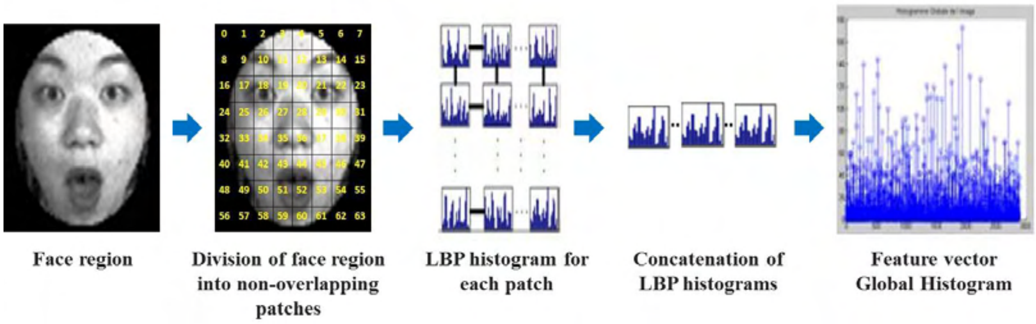

Figure 2: The process of feature vector extraction

Where Ω_X and Ω_Y are respectively the sample space of X and Y. Regarding $p(x)$, $p(y)$, and $p(x,y)$, they are respectively the probability density functions of X, Y, and (X,Y). In the pattern recognition applications, we expect a feature set that can remove the uncertainty of the class variable as much as possible. This can be achieved by finding a feature set $S_m = X_1, X_2, \ldots, X_m$ which jointly have the largest dependency on the target class c. This large dependency defines the Max-Dependency which has the following form in Eq.(2)

$$max \quad D(S_m, c) \tag{2}$$

Despite the theoretical value of Max-Dependency, it is often hard to get an accurate estimation for the multivariate density $p(x_1, \ldots, x_m)$ and $p(x_1, \ldots, x_m, c)$, because of the high-dimensional space. The high-dimensional space is due to the number of samples which is often insufficient and the multivariate density estimation which involves computing the inverse of the high-dimensional covariance matrix that is usually an ill-posed problem [PLD05]. So as the Max-Dependency criterion is hard to implement, an alternative is to select features based on maximal relevance criterion.
Actually, the max-Relevance creterion aims to select features that approximate with the mean value of all mutual information values between the individual features x_i and a class c. In fact, it searches features satisfying Eq.(3) which approximate $D(S_m, C)$ in Eq.(2) with the mean value of all mutual information values between the individual features x_i and the class c.

$$max \quad D(S_m, C), \quad D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i, c) \tag{3}$$

Our goal is to adapt the mutual information technique to select the most relevant patches. Thus, we calculated the relevance score of each facial feature using the criterion of maximum relevance. Based on the relevance score of each facial feature, we calculated the relevance score of each patch by summing up the relevance score of the patch features averaged by the number of features. This average score presents a measure of the patch pertinence. Finally, we sorted the relevance of patches relaying on their relevance score. An overview of Mutual Information adapted algorithm for regions selection is detailed below.

2.2 The on-line step

After the determination of the salient patches, this step is dedicated to the online-identification of facial expressions.

Algorithm 1 Relevent patches selection using Max-Relevance criterion

Variables:
N = Number of features
M = Number of patches
NF_{Patch} = Number of features per patch
FeatRel[] = Table of relevance per feature
PatchRel[] = Table of relevance per Patch
PatchRelSort[] = Sorted table of patches relevance
Sum = Sum of features relevance score per patch
1-Compute the relevance score for every feature.
for i = 1, ..., N
FeatRel[i] = MaxRel($Feat_i$)
End_For
2-Compute the relevance score for every patch.
for k = 1, ..., M
for j = 1, ..., NF_{Patch}
Sum = Sum + FeatRel[j + ($NF_{Patch} \times$ k)]
End_For
PatchRel[k] = Sum / NF_{Patch}
End_For
3-Sort the RegRel table according to the patch relevance score.
PatchRelSort[] = sort(PatchRel);

Unlike the existing work [SG08] [SGM05] [FC15] which extract the feature vector from the whole image, we applied the LBP operator only the most discriminating patches to compute the LBP histogram. Thereafter, we concatenated the different LBP histograms to a single LBP histogram describing the overall appearance of the displayed expression as well as the spatial relationships between the selected patches. This LBP histogram involves information about the local distribution of the salient patches, such as the edges, the spots and the flat areas, to statistically describe the facial expression. Figure 3 describes the extraction of the feature vector from the relevant patches.

The generated LBP histogram provides a description of facial expression in three levels: the histogram labels involve information on a pixel-level, the summed labels of each patch describe the information on a region-level, and the concatenated histograms of each patch includes a description of the observed facial expression on a global-level.

In our work, we used the seven common classes of facial expressions: the neutral expression and the Ekman basic six expressions [Ekm72] : Neutral, Happiness, Fear, Disgust, Anger, Sadness and Surprise (cf. figure 4)

To build the facial expressions classifier, we processed with the SVM classifier [Vap98] as it allows a non-linear classification and it is independent from the size of the data space. Moreover, the robustness of the SVM classifier has already been proven in several studies of facial expressions recognition [BLFM03] [LBF$^+$04]. As the SVM classifier takes binary decisions, a multi-class classification is performed by a cascading of binary classifiers with a scenario of vote. Thus, we described each face with a feature vector describing the preselected salient patches. Finally, the SVM classifier is applied to find out the optimal separation plan between facial expressions classes, and hence identify the corresponding facial expression class.

3 EXPERIMENTAL STUDY

Before presenting the results of the proposed method, we briefly describe the corpus and the used validation techniques.

3.1 Description of the corpus

The evaluation of the proposed method for facial expression recognition was performed on two databases:

- The JAFFE database (The Japanese Female Facial Expression) [LBA99a]: is widely used in the facial expressions research community. It is composed of 213 images of 10 Japanese women displaying seven facial expressions: the six basic expressions and the neutral one. Each subject has two to four examples for each facial expression.
- The KANADE database [KCT00]: is composed of 486 video sequences of people displaying 23 facial expressions within the six basic facial expressions. Each sequence begins by a neutral expression and finish with the maximum intensity of the expression. For fair comparison between KANADE and JAFFE databases, we selected from the KANADE database the first image (neutral expression) and the last three images (with the maximum intensity of the expression) of 10 people chosen randomly. Moreover, we selected the six basic facial expressions and the neutral one.

3.2 Techniques of validation

As a measure of validation, we used the Correct Classification Rate (CCR) of an expression defined as follow:

$$CCR = \frac{\text{Number of samples correctly classified as expression (E)}}{\text{Number of total samples with the expression (E)}} \tag{4}$$

The CCR is computed using the K-cross validation, with K = 10. Therefore, we segmented both of the image databases (JAFFE, KANADE) to 10 sets, and each time we use 9 sets for learning and keep the remaining set (not learned) for the test. We calculate the CCR for each test set and then we averaged these rates.

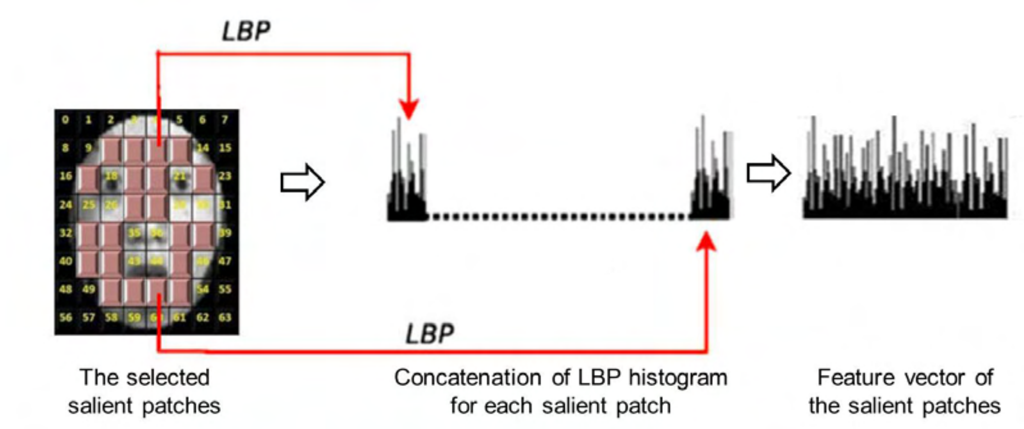

Figure 3: Feature vector extraction from the salient patches

Figure 4: The six basic expressions, from left to right : Anger, Disgust, Fear, Happiness, Sadness and Surprise.

3.3 Results of the proposed method

The experiments described in this section are justified by three reasons: (1) validate the choice of the number of division of the face image into patches , (2) validate the convenience of selecting the discriminant patches, and finally (3) compare the performance of our method with the most known works in literature.

3.3.1 First series of experiments

Through this experiment, we determine the number of the most appropriate division. Therefore, we tested different number of face divisions. Table 3.3.1 presents this experimental study.

The obtained results show that dividing the face image into 8*8 or 9*9 patches leads to the same CCR (93.89%). We opted for 8*8 divisions since it has the smallest dimension feature vector.

3.3.2 Second series of experiments

To select the most salient patches for facial expression recognition, we examined the evolution of the CCR through the number of the selected patches. Figure 5 shows this evolution.

Based on this assessment, we perceive how the CCR increases rapidly with the patches having the highest relevance score. In fact, we achieved the best CCR (93.89%) using only 21 patches. These patches are mainly located around the areas of the mouth, the eyes, the eyebrows and the nose (cf. figure 6) which validates the psychologists studies [Mag07].

To validate the relevance of the selected patches, we examined their independency from the database and the used descriptor. Therefore, we first applied our method of salient patches selection on a second images database: The KANADE database. The selected patches are shown in Figure 7 (b).

As shown in Figure 7, our method of patches selection produced 25 patches. Among the 25 selected patches, 21 are the same as those selected in the JAFFE database (cf. Figure 7 (a)). These results show an important overlap between the selected patches in JAFFE and KANADE databases. This proves the independency of the selected patches from the database and hence the genericity of our facial expression recognition method.

Besides, to study the independency of the selected patches from the used descriptor, we applied our method of salient patches selection on JAFFE database using the DWT (Discrete Wavelet Transform) descriptor. The choice of DWT operator rely on its several advantages mainly its simplicity of computation which allows analyzing real-time images as well as its invariance to illumination variations. Such an operator has been widely exploited in the context of facial expression recognition [ZZG04] [CW02] [MS00]. In fact, the DWT analyzes the image in different resolution levels using a low-pass and a high-pass filters. By applying

Number of patches	5×5	6×6	7×7	8×8	9×9
Feature vector size	12800	18432	25088	32768	41472
CCR	88.73%	90.61%	92.95%	93.89%	93.89%

Table 1: CCR based on the number of the patches

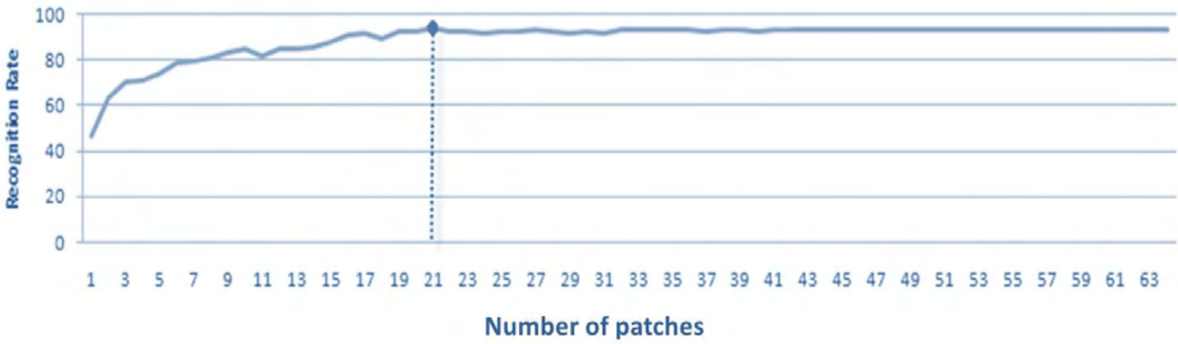

Figure 5: Evolution of the CCR through the number of the selected patches

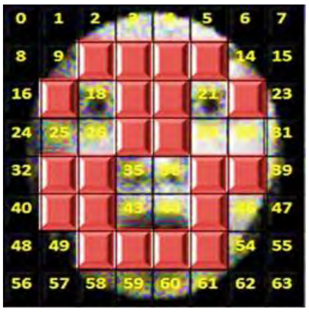

Figure 6: The selected patches

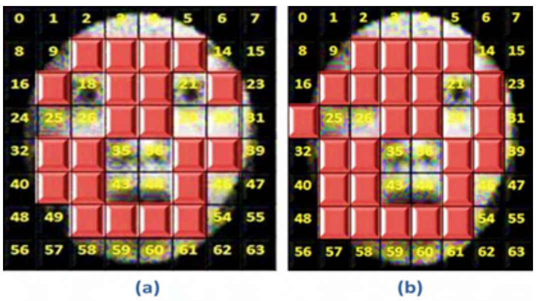

Figure 7: The selected patches from the JAFFE database (a) and the KANADE database (b)

the DWT operator on JAFFE database, 25 patches were selected (Figure 8 (b)).

According to Figure 8, we find out that among the 25 selected patches, 21 are the same as those selected with the LBP operator (cf. Figure 8 (a)). This overlap between the selected patches shows the independency of the selected patches from the used descriptor and thus the genericity of the proposed recognition method.

In order to attest the contribution of selecting the discriminating patches in the proposed method, we compared the facial expression recognition performance with those without selection and with selection. This comparison concerns not only the recognition rate, but also the size of the feature vector and the time execution. Table 3.3.2 shows this assessment.

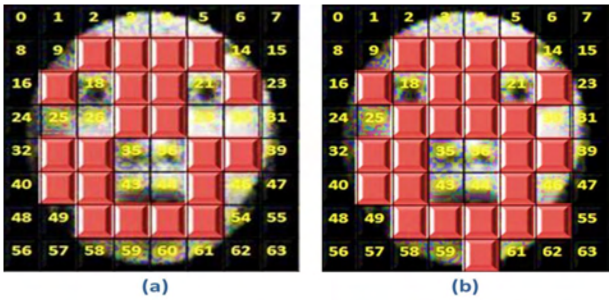

Figure 8: The selected patches using the LBP operator (a) and the DWT operator (b)

Relying on the obtained results, three conclusions are drawn. The first is the contribution of selecting discriminative patches in terms of performance: a gain of 0.47% in facial expression recognition rate. The second is the contribution in terms of space memory: a gain of more than 3 times in the size of the feature vector. The third is the contribution in terms of speed: a gain in time execution of almost 5 time, which is very important for real-time applications.

3.3.3 *Third series of experiments*

This series of experiments aims to compare the proposed method performance with the most known works in the literature [SO04] [ZZ11] [LBA99b] [ZLSA98]. For fair comparison, we selected the methods which performed their experiments on JAFFE database with a 10-cross-validation evaluation technique. Table 3.3.3 shows this comparative study.

	Without selection	With selection
Number of patches	64	21
Feature vector size	16384	5376
Time execution per image (ms)	19 ms	04 ms
CCR	93.42 %	93.89 %

Table 2: The contribution of patches selection in terms of CCR and time execution

Methods	[LBA99b]	[ZLSA98]	[SO04]	[ZZ11]	The proposed method
CCR	92.00%	90.10%	69.40%	81.59%	93. 89%

Table 3: Comparative study between the proposed method and some previous works on JAFFE database

From Table 3.3.3, the proposed method affords the best recognition rate (93.89%), whereas the highest rate recorded by the studied methods is 92.00%.
Besides the satisfied results in terms of the recognition rate and the required memory space, we have shown through this series of experiments that our method has the advantage of being much faster with a significant gain in the execution time.

4 CONCLUSION

This paper introduces a new method for facial expression recognition using the most discriminant facial patches. These patches were selected automatically using the mutual information technique. Facial feature extraction was performed using the LBP operator applied only on the preselected facial patches. The experimental study showed the improvement while using only salient patches. In fact, we succeed not only to improve facial expression recognition performance but also to speed up the recognition task which is a very important gain for real time applications.
As future work, we intend to experiment our method with more different facial expression databases. Furthermore, we aim to include the temporal information of facial expressions which may provides more accurate classification results.

5 REFERENCES

[BLFM03] M. S. Bartlett, G. Littlewort, I. Fasel, and R. Movellan. Real time face detection and facial expression recognition : Development and application to human computer interaction. In *CVPR Workshop on CVPR for HCI*, 2003.

[CW02] J. T. Chien and C. C. Wu. Discriminant wavelet faces and nearest feature classifiers for face recognition. *Journal of IEEE Trans. Patt. Anal. Mach. Intelligence*, pages 1644–1649, 2002.

[DC99] M. Dailey and G. Cottrell. Pca gabor for expression recognition. *The Journal of UCSD Computer Science and Engineering Technical Report*, 1999.

[DGG06] K. Delac, M. Grgic, and S. Grgic. Independent comparative study of pca ica and lda on the feret data set. *The Journal of Wiley Periodicals*, pages 1121–1137, 2006.

[Ekm72] P. Ekman. Universals and cultural differences in facial expressions of emotions. In *The Nebraska Symposium of Motivation*, 1972.

[FC15] Y. Fang and L. Chang. Multi-instance feature learning based on sparse representation for facial expression recognition. *Journal of MultiMedia Modeling*, pages 224–233, 2015.

[FJJ09] L. W. Feng, L. S. Juan, and W. Y. Jiang. Automatic facial expression recognition based on local binary patterns of local areas. In *WASE International Conference on Information Engineering*, 2009.

[GD03] G. D. Guo and C. R. Dyer. Simultaneous feature selection and classifier training via linear programming: A case study for face expression recognition. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 346–352, 2003.

[Ham06] Z. Hammal. *Segmentation des traits du visage, analyse et reconnaissance d'expressions faciales par le Modele de Croyance Transferable*. PhD thesis, Universite de Joseph Fourier de Grenoble, 2006.

[KCT00] T. Kanade, J. Cohn, and Y. L. Tian. Comprehensive database for facial expression analysis. In *International Conference on Automatic Face and Gesture Recognition*, pages 46–53, 2000.

[LBA99a] M. Lyons, J. Budynek, and S. Akamatsu. Automatic classification of single facial images. *Journal of IEEE Transaction Pat-*

tern Analysis Machine Intell, pages 1357–1362, 1999.

[LBA99b] M. J. Lyons, J. Budynek, and S. Akamatsu. Automatic classification of single facial images. *The Journal of IEEE Trans. Pattern Analysis and Machine Intelligence*, pages 1357–1362, 1999.

[LBF+04] G. Littlewort, M. Bartlett, I. Fasel, J. Susskind, and J. Movellan. Dynamics of facial expression extracted automatically from video. In *IEEE Workshop on Face Processing in Video*, 2004.

[LFCY06] S. Liao, W. Fan, A. C. S. Chung, and D. Y. Yeung. Facial expression recognition using advanced local binary patterns, tsallis entropies and global appearance features. In *The International Conference on Image Processing*, pages 665–668, 2006.

[LP12] R. Londhe and V. Pawar. Facial expression recognition based on affine moment invariants. *The IJCSI International Journal of Computer Science Issues*, pages 388–392, 2012.

[Mag07] A. F. Magalhaes. *The Psychology of Emotions : The Allure of Human Face*. Pessoa Press, 2007.

[MS00] E. Morales and F. Y. Shih. Wavelet coefficients clustering using morphological operations and pruned quadtrees. *Journal of Pattern Recognition*, pages 1611–1620, 2000.

[OPH96] T. Ojala, M. Pietikainen, and D. Harwood. A comparative study of texture measures with classification based on feature distributions. *Journal of Pattern Recognition*, pages 51–59, 1996.

[PLD05] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency max-relevance and min-redundancy. *Journal of IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1226–1238, 2005.

[SG08] C. Shan and T. Gritti. Learning discriminative lbp-histogram bins for facial expression recognition. In *British Machine Vision Conference*, 2008.

[SGA09] B.L. Sheaffer, J.A. Golden, and P. Averett. Facial expression recognition deficits and faulty learning: Implications for theoretical models and clinical applications. *International Journal of Behavioral Consultation and Therapy*, pages 31–55, 2009.

[SGM05] C. Shan, S. Gong, and P. W. McOwan. Robust facial expression recognition using local binary patterns. In *IEEE International Conference on Image Processing*, pages 370–373, 2005.

[SJD08] P. Sungsoo, S. Jongju, and K. Daijin. Facial expression analysis with facial expression deformation. In *The in Proc. IAPR Int. Conf. Pattern Recog*, pages 1–4, 2008.

[SO04] Y. Shinohara and N. Otsu. Facial expression recognition using fisher weight maps. In *The IEEE Int Conf Automatic Face and Gesture Recognition*, pages 499–504, 2004.

[Soo00] E. Soofi. Principal information theoretic approaches. *Journal of the American Statistical Association*, pages 1349–1353, 2000.

[ST08] C. Shan and T.Gritti. Learning discriminative lbphistogram bins for facial expression recognition. In *British Machine Vision Conference*, 2008.

[Vap98] V. N. Vapnik. Statistical learning theory. *Wiley*, 1998.

[VJ04] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, pages 137–154, 2004.

[VTG+15] M. F. Valstar, T.Almaev, J. M. Girard, G. McKeown, M. Mehu, L. Yin, M. Pantic, and J. F. Cohn. Fera 2015 - second facial expression recognition and analysis challenge. In *the 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, pages 1–8, 2015.

[ZLSA98] Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu. Comparison between geometry-based and gabor wavelets-based facial expression recognition using multilayer perceptron. In *Automatic Face and Gesture Recognition*, pages 454–459, 1998.

[ZZ11] X. Zhao and S. Zhang. Facial expression recognition based on local binary patterns and kernel discriminant isomap. *Journal of Sensors*, pages 9573–9588, 2011.

[ZZG04] B. L. Zhang, H. Zhang, and S. Ge. Face recognition by applying wavelet subband representation and kernel associative memory. *Journal of IEEE Trans Neural Networks*, pages 166–177, 2004.