Improved Performance in Facial Expression Recognition Using 32 Geometric Features

Giuseppe Palestra[1(✉)], Adriana Pettinicchio[2], Marco Del Coco[2], Pierluigi Carcagnì[2], Marco Leo[2], and Cosimo Distante[2]

[1] Department of Computer Science, University of Bari, Bari, Italy
giuseppe.palestra@gmail.com
[2] National Institute of Optics, National Research Council, Arnesano, LE, Italy

Abstract. Automatic facial expression recognition is one of the most interesting problem as it impacts on important applications in human-computer interaction area. Many applications in this field require real-time performance but not all the approach are suitable to satisfy this requirement. Geometrical features are usually the most light in terms of computational load but sometimes they exploits a huge number of features and do not cover all the possible geometrical aspect. In order to face up this problem, we propose an automatic pipeline for facial expression recognition that exploits a new set of 32 geometric facial features from a single face side covering a wide set of geometrical peculiarities. As a results, the proposed approach showed a facial expression recognition accuracy of 95,46% with a six-class expression set and an accuracy of 94,24% with a seven-class expression set.

Keywords: Facial expression recognition · Human-computer interaction · Geometric features · Random forest

1 Introduction

In communication and interaction between people, facial expression become essential for immediate transmission of emotion and social intentions. During the last decades, the field of facial expression has received growing interest from research community because of the rapidly development in computer technologies.

An automatic system for Facial Expression Recognition (FER) should be able to recognize the six basic face expression defined in by Facial Action Coding System (FACS) developed by Ekman and Friesen [2], that probably represents the most spread study on facial activity.

This kind of system usually consists of three basic modules as presented in many surveys such as [1]: Face Detection, Features Extraction, Facial Expression Classification.

Face Detection is a two-class problem related to the presence or not of the face in the shown image [17]. Several recent works, in this field, make use of the well known Viola-Jones face detector [3], a quite generic and widely spread

V. Murino and E. Puppo (Eds.): ICIAP 2015, Part II, LNCS 9280, pp. 518–528, 2015.
DOI: 10.1007/978-3-319-23234-8_48

algorithm that minimizes computational time while achieving high detection accuracy.

Feature Extraction represents the most important phase in the procedural pipeline as it is able to influence the whole accuracy of the system because of the fact that facial features allowed us to identify and describe the different parts of the face, such facial contour, eyebrows, eyes, nose and mouth .

Basically, there are two types of features: appearance and geometrical features. Appearance features concern skin features, fold and wrinkles whereas geometrical features are obtained from landmarks of the face (eyes, eyebrows, mouth, nose, cheeks, lips and chin).

Appearance methods are exploited in many works in recent years. For instance a Local Binary Pattern descriptor is used by Zhao et al. in [10] and Zhao, G. et. al. [11]. In [12] Jabid et al. chose to work with local directional pattern features (LDPA). Anyway appearance based methods involves a huge amount of data and could be too complex for a real time features extraction. In order to reduce the features extraction complexity some works have been oriented to the use of hybrid models defined from geometrical and appearance features, such as in [16]. Even so, noise issue introduced by the combination of the two approaches could lead to a non satisfying recognition accuracy.

On the other hand geometrical methods are characterized by low computational complexity and good accuracy. In [9] the authors propose a set of 125 geometrical features to perform the classification on facial expression. Geometrical features in association with Neural Network on an image sequence are exploited by [14]. Active Shape Model (ASM) was used to define many approaches to facial expression recognition: in Bevilacqua et al. [5], polygonal features were used, in particular, the areas of five polygons was calculated whereas in Loconsole et al. [4], linear and eccentricity features using facial landmarks and differential features were considered, the latter were defined by subtraction neutral expression features from one the six primary expression features.

Anyway, to the best of our knowledge, a combination of linear, eccentricity, polygonal and slope features was not yet used.

In this paper we propose an innovative and automatic pipeline aided to recognize facial expression from a static image exploiting the combination of features derived by geometrical information. In particular the joint use of linear eccentricity, polygonal and angular features has been adopted in order to improve the recognizing accuracy respect the most recent works. As a first step the face is detected and 20 facial landmarks are automatically extracted by means of the automatic facial keypoints extractor proposed in [7]. Successively the detected landmarks are processed in order to construct a set of 32 geometrical facial features based on linear, eccentricity, polygonal and angular information. Successively the set of features is given as input to the classifier that supplies the predicted facial expression.

Moreover we also focused attention on the classification stem by analyzing the performance with three different methods usually adopted for this task. More specifically performances with Support Vector Machines (SVM's), Random Forest and k-Nearest Neighbours (k-NN) have been tested. All the experiments were performed on the Extended Cohn-Kanade (CK+) facial expression database [8].

The rest of the paper is organized as follows: Section 2 deeply describes the proposed methodology; Section 3 is aimed to build a suitable testing dataset and to find the best configuration for the proposed pipeline; Section 4 reports the experimental results of the presented approach on both static images and video sequences and, finally, Section 5 gives conclusions and introduces future works.

2 Proposed Metodology

This section is aimed to present a detailed description of the pipeline proposed for FER classification and reported in Figure 1. As a first step an algorithm oriented to the *face detection* is applied and then a *facial landmarks extraction* module is applied. Once the facial landmarks are available they are given as input to the *geometrical feature computation* step that computes 32 geometrical features; finally the computed features are analyzed by a *classification* module in order to get a prediction of the facial expression. In the followings each step is further detailed.

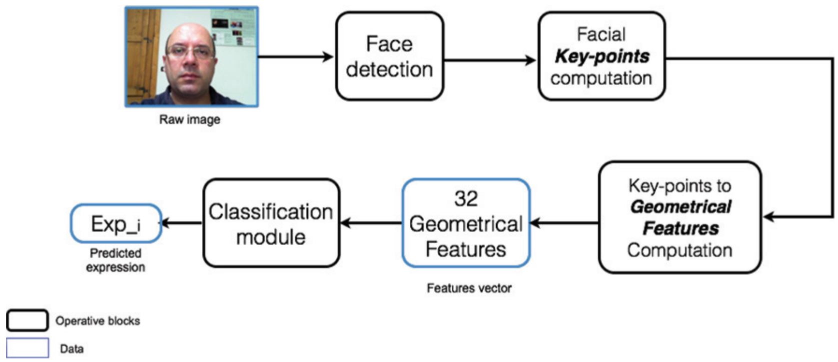


Fig. 1. Proposed pipeline: the face is detected and then facial key-points are extracted and used to compute the 32 geometrical features. Finally the vector of features is sent to the classification module.

Face detection and facial landmarks extraction are the first steps. Face detection is achieved by a Viola-Jones based face detector [3]. The detected face is then processed with a facial landmarks extractor exploiting the STaked Active Shape Model (STASM) approach. STASM uses Active Shape Model for locating 77 facial landmarks with a simplified form of SIFT descriptors and it operates with Multivariate Adaptive Regression Splines (MARS) for descriptor matching. This modified ASM is fast and it has been proved to perform better than existing techniques for automatic face landmarking on frontal faces [7].

32 geometrical features are then computed. More specifically, considering the study of Ekman et al. [2], an accurate observation of facial expression and landmark points extracted from STASM, a set of 32 features, useful to recognize facial expressions, has been defined.

In order to reduce the whole computational costs a face symmetry with respect the vertical axis passing through the center of the face has been assumed and 20 STASM landmarks have been used. More specifically, for the upper part of the face, that involves features related to eyebrows, eyes and cheeks, just one side has been considered (the left side). On the other hand, for the lower part of the face, features related to nose and mouth have been chosen involving both face sides.

The proposed features are shown in the Figure 2) and detailed in the followings. *Linear features* are defined by the Euclidean distance between 2 points. More precisely 15 linear features are used in the proposed approach:

- 3 for left eyebrow,
- 2 for left eye
- 1 for cheeks;
- 1 for nose
- 8 for mouth.

Polygonal features are determined by the area of irregular polygons constructed on three or more facial landmark points. This area is computed by the Gauss equation:

$$A = \frac{1}{2} \cdot \left| \sum_{i=1}^{n} (x_i \cdot y_{i+1} - x_{i+1} \cdot y_i) \right| \tag{1}$$

where x_i and y_i are the Cartesian coordinates of the i-th facial landmark point and n represents the number of sides of the polygon. In this case 3 polygonal features have been defined:

- 1 for the left eye;
- 1 between corners of left eye and left corner of mouth;
- 1 for mouth.

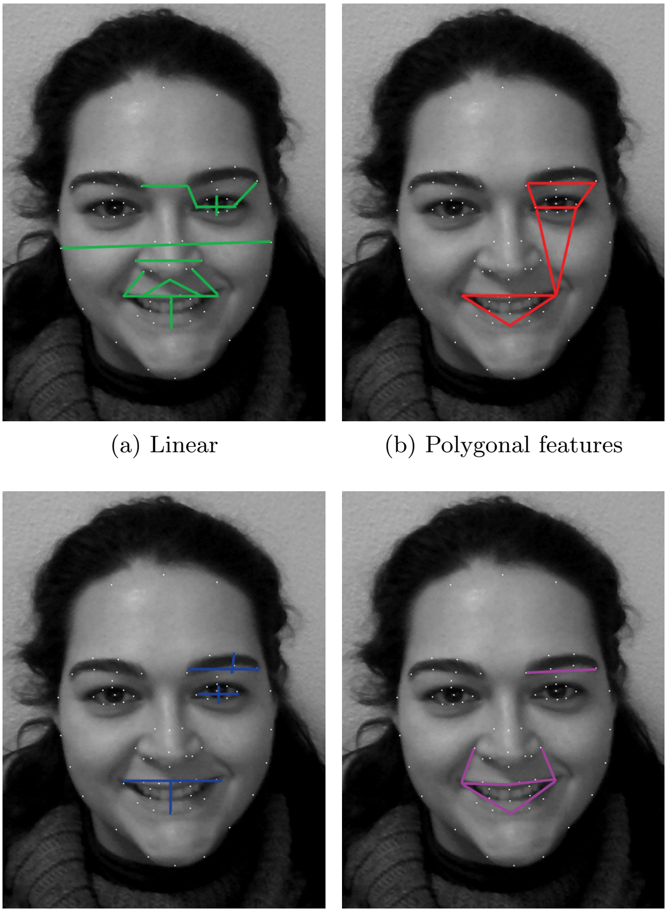
Elliptical features are defined by the major and minor ellipse axes ratio. In particular, 7 ellipses among landmarks point are chosen.

- 1 for left eyebrow;
- 3 for the left eye: eye, upper and lower eyelids;
- 3 for mouth: upper and lower lips.

Slope features are defined by the slope (m) of the line joining two facial points a and b .

$$m = \frac{y_b - y_a}{{}_{x}a - y_a} \tag{2}$$

More specifically the slope features are used in order to define 7 of the proposed features



(a) Linear (b) Polygonal features

(c) Elliptical features (d) Slope features

Fig. 2. Geometrical Facial Features of left side of the face. Features related to eyebrows, eye and cheeks, only one side involves just the left side of the face.

- 1 for left eyebrow;
- 6 for mouth corners.

Classification module is the last step that analyzes the features vector in order to get a prediction in terms of facial expression. In recent years a huge amount of possible classifiers has been proposed. However every classifier shows variable performance depending on the peculiarities of the input data. In this work a new set of features has been proposed introducing an uncertainty about the best way to perform the classification step. With the idea to give an answer to this issue three different classifiers have been tested: a bank of Support Vector Machines (SVMs), Random Forest (RF), and the k-NN.

3 Experimental Setup

In this section the selection of a suitable set of images to test and optimize the proposed pipeline is initially carried-out. Then the best classification algorithm among the considered ones is experimentally defined.

All the evaluation tests have been conducted on the Extended Cohn-Kanade (CK+) data set, a facial image database of 123 individuals of different gender, race and age [8]. 8-bit grayscale image sequences from neutral to peak expression were digitized into 640 by 490 pixels of resolution. More specifically just 97 individuals have been considered, as suggested in [10]. This reduction was related both to the lack of emotional labels for several subjects that could affect the classification phase, severe lighting conditions of images and to the necessity to balance the data set. This way 2 different subsets have been built by the selection of static images among the 593 sequences of the 97 subjects. The first one was a 6 expression subset with the following distribution among different expressions: anger (86), disgust (95), fear (98), happiness (86), sadness (81) and surprise (101) with a total number of images of 547. A second subset with 7 expression has been built by adding 87 images with neutral expression.

Once a suitable testing dataset has been selected the next step is the selection of the best classification strategy for the proposed geometrical features set. As previously highlighted three different classification methods, Support Vector Machines, k-Nearest Neighbors (with k=1 to obtain high accuracy) and Random Forest have been compared. More specifically a k-fold validation on both the built subsets (6 and 7 expression) has been used with a k=10 as already found in previous studies. The use of both the subsets is a key point that allows the results to be more general and then to avoid all possible ambiguities related to the dataset constraints.

The experimental results, summarized in Table 1, show that in both the experiments, Random Forest experienced the best classification performed. Moreover these results confirmed the ability of the proposed solution to fully recognize facial expressions from a small set of geometric features.

Going into details of the 6 expressions results, the recognition rates were: 82,41%, 74,76% and 95,46% using respectively Multi-SVM, k-Nearest Neighbors and Random Forest. Whereas, taking into account 7 expression classification Multi-SVM reached 70,94%, whereas k-Nearest Neighbors achieved 71,96% and Random Forest reached 94,24%.

4 Experimental Results

Once the best pipeline settings have been found the next step is to perform a more accurate experimental session in order to deeply investigate the potentially of the proposed approach. To this aim the confusion tables are presented and discussed in subsection 4.1 and then the performance obtained are compared with those of the the most relevant works presented in the literature and illustrated in subsection 4.2. Finally the behavior on video sequences is highlighted and discussed.

Table 1. Table of results for six-class and seven-class expression recognition using Multi-SVM, Random Forest and k-NN on one side (left) of the face (Unit: %).

Expression	Six-class Expression			Seven-class Expression		
	Multi-SVM	Random Forest	k-NN	Multi-SVM	Random Forest	k-NN
Anger	44.71	88.24	68.24	71.76	87.06	61.18
Disgust	88.42	93.68	81.05	80.23	96.84	78.95
Fear	77.55	97.96	74.49	58.02	96.94	72.45
Happiness	98.84	95.35	81.40	78.95	97.67	82.56
Sadness	93.83	97.53	69.14	77.55	90.12	60.49
Surprise	91.09	100.00	74.26	73.27	99.01	74.26
Neutral	-	-	-	56.82	92.05	73.86
Accuracy	**82.41**	**95.46**	**74.76**	**70.94**	**94.24**	**71.96**

4.1 Discussion on Confusion Tables

This subsection compares and discusses the performance of the proposed approach choosing features from the left or the right side of the face. As described in Section 2, some of the 32 proposed geometrical features exploit the face symmetry and they are computed on the left side of the face in order to reduce the computational complexity. With the aim to prove the symmetry hypothesis, the same set of features, developed on the right side of the face, has been considered. All experiments have been performed on both the 6 and 7 expression subsets with the Random Forest classifier and the previously mentioned k-fold validation system with k=10. Tables 2 and 3 show confusion matrices, respectively for the 6 and 7 expressions subsets, summarizing both the results obtained with the use of the features from left or right side of the face. Regarding the 7 expressions subset, the proposed approach achieved 95,46% on left side and 94,36% on right side, while with the 6 expressions subset, it reached 94,24% on left side, 93,32% on right side. Despite a difference less than 1,5%, the proposed approach consistently achieved high recognition rates when applied to the left side of the face as well as if applied to the right one. Going into details, this deeper investigation highlights an ambiguity between anger, disgusted and sad expressions. This seems quite reasonable, for all these types of expressions, where strict lips and low position of eyebrows are very similar in location and shape.

4.2 Comparison with State-of-the-Art Methods

In this section, the accuracy of the proposed system was compared against common feature-based approaches. Unfortunately, there exist many different evaluation protocols in literature that make the comparison very challenging. In order to perform a comparison as fair as possible, the competitors were chosen among those solutions that use an evaluation protocol based on the CK+ dataset with 6 expressions.

Considering geometrical methods, in [9] the authors showed a graph of the number of features versus recognition accuracy, for both training and testing

Table 2. Confusion matrix of six-class expression recognition using the left and the right side of the face (Unit: %).

Expression	Anger	Disgust	Fear	Happiness	Sadness	Surprise	Side
Anger	88.24	5.88	0.00	0.00	5.88	0.00	Left
	88.24	5.88	0.00	1.18	4.71	0.00	Right
Disgust	4.21	93.68	1.05	0.00	1.05	0.00	Left
	3.16	90.53	3.16	0.00	3.16	0.00	Right
Fear	0.00	0.00	97.96	1.02	0.00	1.02	Left
	0.00	0.00	95.92	3.06	0.00	1.02	Right
Happiness	0.00	0.00	4.65	95.35	0.00	0.00	Left
	0.00	0.00	2.33	97.67	0.00	0.00	Right
Sadness	1.23	1.23	0.00	0.00	97.53	0.00	Left
	4.94	1.23	1.23	0.00	92.59	0.00	Right
Surprise	0.00	0.00	0.00	0.00	0.00	100.00	Left
	0.00	0.00	0.00	0.00	0.00	100.00	Right

Table 3. Confusion matrix of seven-class expression recognition using the left and the right side of the face (Unit: %).

Expression	Anger	Disgust	Fear	Happiness	Neutral	Sadness	Surprise	Classifier
Anger	87.06	5.88	0.00	0.00	2.35	4.71	0.00	Left
	88.24	5.88	0.00	1.18	2.35	2.35	0.00	Right
Disgust	2.11	96.84	0.00	0.00	1.05	0.00	0.00	Left
	2.11	90.53	3.16	0.00	1.05	3.16	0.00	Right
Fear	0.00	0.00	96.94	2.04	0.00	0.00	1.02	Left
	0.00	0.00	96.94	2.04	1.02	0.00	0.00	Right
Happiness	0.00	0.00	2.33	97.67	0.00	0.00	0.00	Left
	0.00	0.00	3.49	96.51	0.00	0.00	0.00	Right
Neutral	3.41	0.00	0.00	0.00	92.05	4.55	0.00	Left
	3.41	1.14	0.00	0.00	90.91	4.55	0.00	Right
Sadness	1.23	3.70	0.00	0.00	4.94	90.12	0.00	Left
	6.17	3.70	0.00	0.00	0.00	90.12	0.00	Right
Surprise	0.00	0.00	0.99	0.00	0.00	0.00	99.01	Left
	0.00	0.00	0.00	0.00	0.00	0.00	100.00	Right

data. The highest classification accuracy of 95,17% was achieved with a minimum of 125 feature vectors, so three times the number of features we proposed in the present paper (only 32). In [14], a 81% recognition rate has been achieved for 4 emotions, using a geometric features extraction method and Neural Network classifications on sequence of images. In another work [15], authors use static images and geometrical methods as in our approach and obtain their 90,33% of accuracy. Youssif and Asker [16] reach a 93,5% six-classes recognition rate combining geometrical and appearance methods (an hybrid method).

On the other hand, concerning appearance methods, recently, Zhao et al. [10] obtained 94,88% of recognition accuracy. Their selected static images of 96 subjects, using for each sequence, one neutral image and three peak faces, thus the data set results unbalanced on neutral emotion. Zhao, G. and Pietikäinen, M. [11] reached 96,26% of accuracy, using local binary patterns and SVM classifiers but it was only tested on manually aligned image sequences. In Jabid et al. [12], they achieved 93,69% of recognition accuracy, using local directional pattern features, which are similar to the LBP feature with SVM. The best average recognition accuracy of the different methods proposed by researchers is around 95%, on the Cohn-Kanade facial expression database, but one limitation of the existing facial expression recognition methods is that they attempt to recognize facial expression from sequence of images where facial expression evolves from a neutral state to a fully expressed state. In the current study, we randomly chose static images from the CK+ data set in order to avoid any possible calibration of the system. Table 4 reports the comparison with the State-of-the-Art demonstrating that the proposed approach gave the best average recognition rate and this represents an important advancement in the field of automatic recognition of facial expressions.

Table 4. Performance comparison with State-of-the-Art Methods (CK+ 6 expressions) (Unit: %).

Appearence/Hybrid			Geometrical			
[10]	[12]	[16]	[9]	[14]	[15]	Proposed
94.88	93.69	93.50	95.17	81.00	90.33	**95.46**

4.3 Tests on Video Sequences

This section aims at analyzing the behavior of the proposed pipeline when applied to image sequences. The experiments reported in previous sections were relative to the recognition of facial expressions in a static image containing a peak expression whereas, in common application contexts, the automatic systems have to perform FER by analyzing image sequences in which not all the images contain a clear expression, or where there are transitions between expressions. In order to obtain a quantitative accuracy evaluation of the capability to recognize the expressions embedded in image sequences, the proposed pipeline has been tested on the sequences of the 97 subjects selected from the CK+ dataset. Each CK+ sequences start with a neutral face expression that bring to peak expression and the number of frames, for a particular expression, changes for every subject. With the purpose to make the proposed pipeline suitable to work with video sequences, the following rule has been introduced: a sequence i was considered correctly recognized if in the first m_1 frames there exist at least n_1 images classified as containing the neutral expression and, at the same time, in the last m_2 frames there exist at least n_2 images classified as containing the expression which the sequence is labeled with. Moreover, a test with the purpose to test the classification just for the expressive face was carried out. In this case a sequence i was considered correctly recognized if in the last m_2 frames there exist at least n_2 images classified as containing the expression which the sequence is labeled with. In the experiments the following setting parameters were chosen: $m_1 = 3$,$n_1 = 1$, $m_2 = 5$, $n_2 = 4$ after a carefully experimental evaluation. Over the tested sequences, the percentage of correct classification was of 73.56% for six-class classification and 69.23% for seven-class classification that are very encouraging outcomes.

5 Conclusions and Future Works

In this work, a new pipeline for facial expression recognition has been proposed. More specifically, this study adopts a set of 32 geometrical features exploiting the face symmetry in order to save computational load and keep high accuracy performances. In order to optimize the potential of the proposed set, three different classification approaches have been tested leading to choose the Random Forest approach as the most suitable for this specific features vector. Experimental sessions have been performed on publicly available dataset, experiencing recognition robustness also in real world environments.

To give an idea of the complexity of the algorithm, the actual CPU time taken to process 1 image of the CK+ database was measured. The proposed approach, in the implemented version, recognizes the facial expression in about 1 second, working in R2014a Matlab environment and using an Intel Core i3 (1.8 GHz) with 4 GB of RAM. Future works will address the implementation in a intermediate-level language in order to speed-up the procedure. Where appropriate, processor supplementary instructions will also be used to achieve real-time processing. Furthermore, will be explored the proposed approach with other very challenging problems including more severe head pose variations and occlusions. Spontaneous facial expressions, common in many practical applications, will also be studied.

Acknowledgments. This work has been supported in part by Italian Ministry for Education, University and Research (MIUR) in the framework of "2007-2013 NOP for Research and Competitiveness" under Grants SARACEN with code PON04a3_00201.

References

1. Pantic, M., Rothkrantz, L.: Automatic analysis of facial expressions: The state of the art. IEEE Trans. Pattern Anal. Mach. Intell. **22**, 1424–1445 (2000)
2. Ekman, P., Friesen, W.: In Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto (1978)
3. Viola, P., Jones, M.: Robust real-time face detection. Int. J. Comput. Vision. **57**, 137–154 (2004)

4. Loconsole, C., Runa Miranda, C., Augusto, G., Frisoli, A., Orvalho, V.: Real-time emotion recognition: a novel method for geometrical facial features extraction. In: Proc. of 9th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, January 5–8, Lisbon, Portugal (2014)
5. Bevilacqua, V., D'Ambruoso, D., Mandolino, G., Suma, M.: A new tool to support diagnosis of neurological disorders by means of facial expressions. In: Proceedings of IEEE 2011 International Workshop on Medical Measurements and Applications Proceedings, May 30–31, Bari, Italy pp. 544–549 (2011)
6. Asthana, A., Saragih, J., Wagner, M., Goecke, R.: Evaluating AAM Fitting Methods for Facial Expression Recognition. In: Proc. of the Inter. Conf. on Affective Computing and Intelligent Interaction, September 10–12, Amsterdam, pp. 1–8 (2009)
7. Milborrow, S., Nicolls, F.: Active Shape Models with SIFT Descriptors and MARS. In: Proceedings of the 9th International Conference on Computer Vision Theory and Applications, January 5–8, Lisbon, Portugal (2014)
8. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z.: The Extended Cohn-Kanade Dataset (CK+): A Complete Dataset for Action Unit and Emotion-Specified Expression. In: Proc. of the 3rd IEEE Workshop on CVPR for Human Communication Behavior Analysis, June 13–18, San Francisco, CA, USA, pp. 94–101 (2010)
9. Ghimire, D., Lee, J.: Geometric Feature-Based Facial Expression Recognition in Image Sequences Using Multi-Class AdaBoost and Support Vector Machines. Sensors **13**, 7714–7734 (2013)
10. Zhao, X., Zhang, S.: Facial expression recognition based on local binary patterns and kernel discriminant isomap. Sensors **11**, 9573–9588 (2011)
11. Zhao, G., Pietikäinen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Trans. Pattern Anal. Mach. Intell. **29**, 915–928 (2007)
12. Jabid, T., Kabir, M.H., Chae, O.: Robust facial expression recognition based on local directional pattern. ETRI J **32**, 784–794 (2010)
13. Shan, C., Gong, S., McOwan, P.: Facial expression recognition based on local binary patterns: A comprehensive study. Image Vis. Comput. **27**, 803–816 (2009)
14. Rao, K.S., Saroj, V.K., Maity, S., Koolagudi, S.G.: Recognition of emotions from video using neural network models. Expert Systems with Applications **38**, 13181–13185 (2011)
15. Khanum, A., Mufti, M., Javed, M.Y., Shafiq, M.Z.: Fuzzy case-based reasoning for facial expression recognition. Fuzzy Sets and Systems **160**, 231–250 (2009)
16. Youssif, A.A., Asker, W.A.: Automatic facial expression recognition system based on geometric and appearance features. Computer and Information Science **4**, 115–124 (2011)
17. Martiriggiano, T., Leo, M., D'Orazio, T., Distante, A.: Face Recognition by Kernel Independent Component Analysis. In: Ali, M., Esposito, F. (eds.) IEA/AIE 2005. LNCS (LNAI), vol. 3533, pp. 55–58. Springer, Heidelberg (2005)