Spatiotemporal Features for Effective Facial Expression Recognition

Hatice Çınar Akakın and Bülent Sankur

Bogazici University, Electrical & Electronics Engineering Department,
Bebek, Istanbul
{hatice.cinar,bulent.sankur}@boun.edu.tr
http://www.ee.boun.edu.tr

Abstract. We consider two novel representations and feature extraction schemes for automatic recognition of emotion related facial expressions. In one scheme facial landmark points are tracked over successive video frames using an effective detector and tracker to extract landmark trajectories. Features are extracted from landmark trajectories using Independent Component Analysis (ICA) method. In the alternative scheme, the evolution of the emotion expression on the face is captured by stacking normalized and aligned faces into a spatiotemporal face cube. Emotion descriptors are then 3D Discrete Cosine Transform (DCT) features from this prism or DCT & ICA features. Several classifier configurations are used and their performance determined in detecting the 6 basic emotions. Decision fusion applied to classifiers improved the recognition performance of best classifier by 9 percentage points. The proposed method was evaluated user independently on the Cohn-Kanade facial expression database and a state-of-the-art 95.34 % recognition performance is achieved.

Keywords: Facial expression analysis, spatiotemporal features, face prism.

1 Introduction

The human face is a rich source of nonverbal information. Indeed, not only it is the source of identity information but it also provides clues to understand social feelings and can be instrumental in revealing mental states via social signals. Facial expressions form a significant part of human social interaction. Automatic understanding of emotions from face images is instrumental in the design of affective human computer interfaces. Next generation human-computer interfaces will be empowered with the capability to recognize and to respond to nonverbal communication clues [1–4]. Dynamic approaches to detect emotions use spatiotemporal information extracted from the image sequences [5–10]. that consists of the changes in the landmark configuration as well changes in the textures appearance of the faces.

In this study, we consider two types of data representation for emotion analysis, the first one being facial landmark trajectories, the second one being the

K.N. Kutulakos (Ed.): ECCV 2010 Workshops, Part I, LNCS 6553, pp. 207–218, 2012.

evolution of the face texture patches. Discriminative features are extracted from these two face representations for the automatic facial expression recognition. Based on these features we develop a novel algorithm for automatic classification of emotional expressions in the face.

The paper is organized as follows. In the next section we briefly review related works. Section 3 describes the data representation types and extracted features. Section 4 presents the experimental results of the proposed algorithms. Finally, conclusions are drawn in Section 5.

2 Related Work

There are a number of different paradigms to recognize automatically the emotion corresponding a facial expression [3, 5, 7, 11–13]. A well established approach attempts to identify Action Units (AUs) [7, 8, 13] based on Facial Action Coding System (FACS) [14]. The facial behavior is decomposed and analyzed in terms of 46 action units (AUs), each of which is anatomically related to the individual facial muscles. Various combinations of these distinctive AUs are capable of representing a sufficiently large number of facial expressions. Another relevant paradigm is to model the appearance changes of the whole face or selected subregions by extracting discriminative features. For example, Tian [15] investigated the effects of different image resolutions for facial expression recognition by using geometric features and appearance-based features. Geometric features were extracted by tracking facial features which represented the shape and location of facial components e.g., mouth, eyes, brows, nose etc. Features were obtained by Gabor filtering applied on the difference images between neutral and expression faces. Similar to Tian's study [15], Bartlett et al. [11] also used Gabor filters for appearance based feature extraction from the still images. They obtained their best recognition results by selecting a subset of Gabor filters using AdaBoost and then training Support Vector Machines on the outputs of the filters selected by AdaBoost.

In addition to Gabor filters, which are robust to illumination changes and detect face edges on multiple scales and with different orientations, Local Binary Patterns (LBP) [16], Volumetric Local Binary Patterns (VLBP) [6] and Haar-like features [10] were also used for facial expression recognition and promising results were obtained.

In Zhao's study [6] the face was treated as a 3D volumetric data, and motion and appearance were jointly modeled by VLBP. Finally SVM classifiers were trained using the extracted VLBP features. Yang et al. [10] used dynamic Haar-like features extracted from manually aligned faces to capture the temporal characteristics of facial expressions, and used them to train a classifier via Adaboost. Shan [16] studied facial representation based on LBP features for facial expression recognition. They examined different machine learning methods, including template matching, SVM, LDA, and the linear programming technique on LBP features. They obtained their best results with Boosted-LBP by learning the most discriminative LBP features with AdaBoost, and the recognition performance of different classifiers were improved by using the Boosted-LBP features.

Zhou and et. al proposed an unsupervised temporal segmentation and clustering algorithm, Aligned Cluster Analysis (ACA) for dynamic facial event analysis, and a multi-subject correspondence algorithm for matching expressions [17]. Facial landmarks are tracked via person specific AAMs and shape and appearance features are extracted from upper and lower half of face. The experiments are conducted on naturally occuring RU-FACS [18] and Cohn-Kanade databases [19].

3 Data Representation and Feature Types

3.1 Facial Expression Database

In principle, facial expressions should be analyzed on spontaneously collected face videos [3, 20]. However, most of the spontaneous facial expression databases are not open for public use [18, 21] and manual labeling of spontaneous and naturally displayed facial emotions is difficult and error prone. Hence, we had to evaluate our facial expression recognition method on the well known and widely used CohnKanade database [19]. The database consists of 100 university students ranging in age from 18 to 30 years who enacted among many others, the six prototypical emotions, i.e., anger, disgust, fear, joy, sadness, and surprise. Image sequences showing the evolution of the emotion from neutral state to target level were digitized into 640 x 480 pixel arrays with 8-bit precision for gray scale values. The video rate is 30 frames per second.

The relevant part of the database for our work has 322 image sequences involving 92 subjects portraying the six basic emotions. The distribution of videos into the 6 emotion classes is not uniform and is given Tables 3 and 4. To guarantee that the reported results are practically person independent, first, we partitioned the dataset into ten groups of roughly equal number of subjects and sequences. Then, we applied ten fold cross validation testing where each fold included novel subjects.

3.2 Data Representation

We used two feature extraction schemes: *i*) Landmark trajectories; *ii*) Spatiotemporal cubes of face sequences. In order to process face videos in either feature scheme, we used both spatial normalization and temporal normalization. For spatial normalization we calculate the distance between eye inner corners (eye inner corner distance - EID), scale the face to a standard size and crop an m x n (typically 64 x 48 and 64 x 32) region of interest, as shown in Figure 1. The temporal normalization is realized by resampling trajectories to a fixed number of frames.

Spatiotemporal Prism of Face Sequences. Facial landmarks tracked across video frames enable us to align the faces and size normalize them. The most informative region of the face is cropped and these cropped patches are stacked

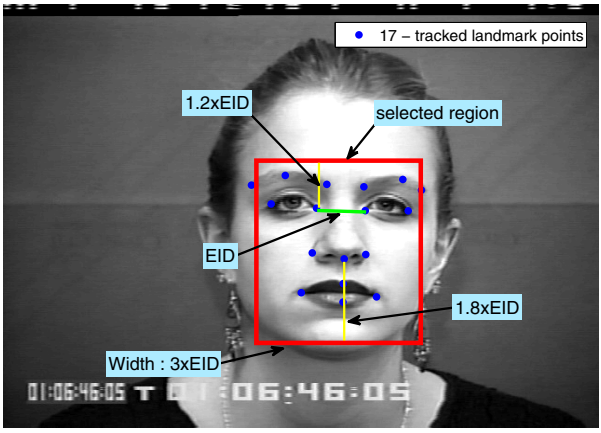

Fig. 1. Facial landmarks to be detected and the cropped face region. The cropped area is dimensioned according to the EID : (Distance between eye inner corners)

in time to form a 3D face prism. The formation of this prism is illustrated in Figure 2 and we denote this array of mxnxT voxels by V. We apply the resample function to the pixels of aligned face pixels to obtain a T-long pixel trajectory, independent of the actual duration of the emotion sequence.

Landmark Trajectories. Our automatic landmark detection algorithm marks a number (F) of facial landmark points tracks them over successive video frames [22] (Figure 3). The detection algorithm consists of block DCT-features trained with SVM classifiers and a graph incorporating statistical face configuration; the tracking algorithm uses adaptive templates, Kalman predictor and subspace regularization.

Once the landmark coordinates are extracted from all frames of the video shot, the landmark coordinate data is organized as a $2F$xT matrix P. Each row of the P matrix represents the time sequence of the x or y coordinates of one of the F landmarks. In order to obtain landmark coordinates independent from the initial position of the head, the first column is subtracted from all other columns of P, so that we only consider the relative landmark displacements with respect to the first frame. This presupposes that the landmark estimates in the first frame of the sequence are reliable. Since duration of emotion videos is variable depending on the emotion type and upon the actor, we normalized the length of the landmark trajectories by using the "resample" function of Matlab so that all sequence data corresponding to some expression video shot had length. The resulting spatiotemporal data matrix P which can also be regarded as a 2D intensity image, where rows correspond to landmark coordinates and columns correspond to normalized time index. In our study we chose T as 16. Notice that the frame lengths of the image sequences in the Cohn − Kanade database vary from 9 to 47 frames. In our work we used 17 landmarks as illustrated in Figure

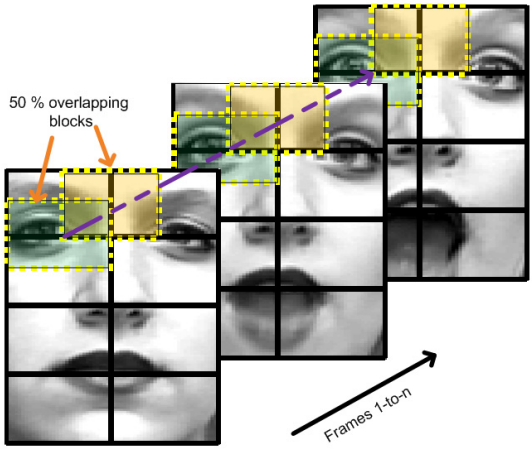

Fig. 2. Illustration of a spatiotemporal prism over a three-frame instance. There is a total of 21 blocks of size 16x16 extracted from the video shot. The blocks overlap by 50%.

3 which resulted in $F = 34$ coordinates. At this stage we have two types of representation of videos: trajectory matrix P_{34xT} and spatiotemporal face prism V_{mxnxT}. Either representation enables the use of subspace projections for feature extraction and classification. In this work we experimented with Discrete Cosine Transform (DCT) due to its good energy compaction property and which can serve the purpose of summarizing and capturing the video content. Other model-based transforms such as Gabor and data-driven transforms such as PCA, ICA and NMF will be explored in a future work.

3.3 Features from Spatiotemporal Face Prism

We consider both global and local 3D DCT transform of the V voxel at two different resolutions in order to investigate the performance of these two forms (local and global features) on expression classification. Martinez recently studied the effect of local and global view-based algorithms on expression variant face matching task [23].

The details of global and local feature extraction and classification methods are as follows:

Global 3D DCT transform: In this case, 3D DCT is applied to the whole 3D face prism V of sizes 64 x 48 x 16 and 64 x 32 x 16. This results in 3-D DCT arrays containing, respectively 49152 and 32768 coefficients. We have selected low-frequency DCT coefficients using 3D zigzag order (excluding the DC term) and and ordered them as a feature vector. Using 3D zigzag scan order we selected low order DCT coefficients with i, j, k indices such

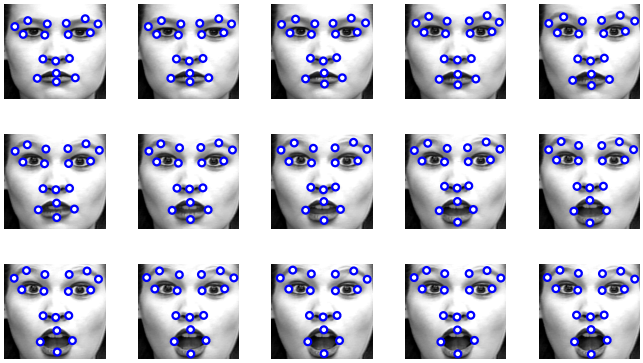

Fig. 3. Illustration of the 17 tracked facial landmarks on a sample image sequence

that 1 ¡ i +j+ k ¡= c. We determined c as 13 which roughly corresponds to 280 DCT coefficients of the face prism data. Notice that the selected DCT coefficients include only 0.5% and 0.8% of all coefficients.

Block-based (local) 3D DCT transform: Here we consider sub-prisms of the spatio-temporal V matrix. A sub-prism consists of a 16x16 block on the face plane as was illustrated in Figure 2, and of the total temporal length T. Thus in fact sub-prisms become with this choice of dimensions 16x16x16 cubes, and each such cube is subjected to DCT. Notice that since the face blocks overlap by 50% the face is covered by B=7x5=35 blocks for the 64x48 sized crop and by B=7x3=21 blocks for the 64 x 32 sized crop. The 16^3 DCT coefficients are zigzag scanned and the first 280 DCT coefficients are selected from each transform cube. Finally the selected DCT coefficients of all cubes are concatenated into a single vector to serve as a feature vector q with dimension 280 x B.

The outcome of the DCT transform (global or local) is a set of transform vectors, one from each emotion sequence. In the case of global transform, the vectors of selected DCT coefficients forms the feature vector itself that is , the 280-dimensional global 3D DCT coefficients. In the case of local transforms, the DCT coefficient dimensionality is excessive (Bx280, i.e., 9800 = 280x35 or 5880 = 280x21), hence must be reduced. In order to use subspace methods, we organized the ensemble of vectors into a data matrix, Q having Bx280 rows and r columns equal to the set of training videos. We obtain feature vectors for each emotion video by using Independent Component Analysis [24] algorithm. The data matrix is decomposed as

$$Q = W.F \tag{1}$$

W is the set ICA basis vectors and F their independent mixing coefficients. Prior to ICA, the row dimension of the Q matrix was reduced from 9800 (or 5880, respectively) to 200 via PCA, and this dimension was experimentally shown to

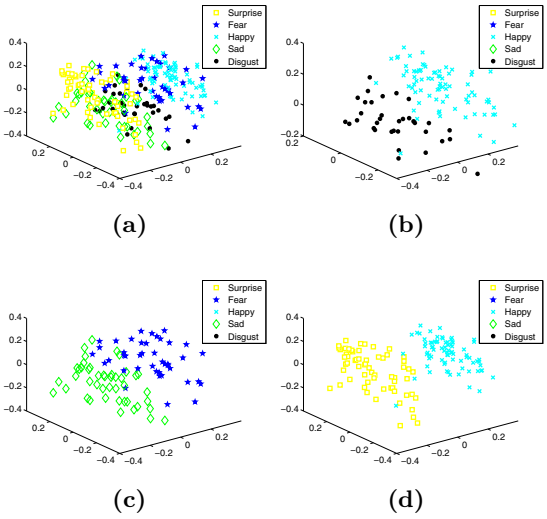

Fig. 4. (a)Illustration of global DCT coefficients (dimension is reduced with MDS). Each color represents a different expression class. (b)Disgust and happy classes, (c) Sad and fear classes, (d) Surprised and happy classes.

yield the best results. Notice that, 200 ICA features is 0.4% and 0.6% of the number of calculated DCT coefficients.

When a test vector arrives, its feature vector q is extracted and it is demixed via the pseudoinverse of W, W^+, and the distance of the resulting mixing vector $f = W^+q$ is measured with respect to the training set of mixing coefficient vectors in F. Figure 4 illustrates the discriminative property of global DCT coefficients. One can see that, disgust-happy, sad-fear and surprised-happy expression pairs can be discriminated by looking at the figures.

3.4 Feature Extraction from Landmark Trajectory Matrix

We apply ICA [24] to the landmark trajectory matrix P in order to extract discriminative and sparse features for different expressions. We vectorize each trajectory matrix, P, to form a column vector q of length 34x16 (544) by by lexicographical ordering of the matrix elements. The resulting $Q_P = [q_1 q_2 \dots q_r]$ 544xr matrix represents the ensemble of training data, where r is again the number of training expressions. Prior to ICA algorithm, the columns of Q_P are reduced to dimension 20 via PCA algorithm.

3.5 Classifiers for Emotion Videos and Their Fusion

Whatever the extracted features, we used a modified nearest neighbor (MNN) classifier is used. The MNN classifier with one neighbor consists of the sum of the

Table 1. Data types and extracted features (P:landmark trajectory matrix, V: spatiotemporal face prism)

Data Type	Spatiotemporal Features (Dimension)	Feature Subspace	Reduced Dimension
Landmark Trajectories	normalized P matrix (34x16)	ICA	(20 x 1)
V matrix (64x48x16)	local block based 3D DCT (280 x 35 blocks)	ICA	(200 x 1)
V matrix (64x48x16)	global 3D DCT of V	-	(280 x 1)
V matrix (64x32x16)	local block based 3D DCT (280 x 21 blocks)	ICA	(200 x 1)
V matrix (64x32x16)	global 3D DCT of V	-	(280 x 1)

minimum and median distances. In other words, distances of the feature vector of a test video are computed to all training vectors in each expression class, and quite stable results are obtained by summing the minimum and the median of these distances. Finally, the test expression is assigned the label of the minimum distance class.

Since we considered more than one feature set and several classifier parameter settings, we experimented with fusion schemes and found them quite advantageous. We observed that decision level fusion improves the classification performance significantly. In this study, the decision level fusion is implemented by summing the similarity scores of different classifiers. We used five different feature sets as summarized in Table 1. In the similarity score fusion, the class with highest summed similarity score is chosen as the winner. We applied unit-norm normalization to the similarity scores of the classifiers.

4 Experimental Results

We tested our facial expression recognition algorithm on the Cohn−Kanade facial expression database [19]. Experiments are conducted using leave-one-group-out cross-validation testing scheme. The recognition results reported in this study are computed as the average of the 10-fold testing. We designated five classifiers to take roles in fusion scheme. All these classifiers (Table 2)use MNN classification, but otherwise differ in the data representation and the features extracted:

Individual performances of the classifiers over expression categories are represented in Table 3. It is observed that global 3D DCT coefficients surpass the performance of other features and classifiers. Table 4 summarizes the confusion matrix obtained by fusing the similarity scores of the five classifiers with different feature and data types. Table 4 demonstrates that the score fusion of the five classifiers (Table 2) increases the overall facial expression recognition performance to 95.34 % where as the best individual classifier achieves only 86.3% recognition accuracy. Table 5 compares the recognition performance of

Table 2. Designed classifiers and their features

Classifier - I:	20 ICA coefficients of landmark trajectory matrice P
Classifier - II:	280 global 3D DCT coefficients of $V_{64x32x16}$
Classifier - III:	280 global 3D DCT coefficients of $V_{64x48x16}$
Classifier - IV:	200 ICA features extracted from 3D DCT coefficients of face blocks of $V_{64x32x16}$ (280 x 21 DCT coefficients reduced to 200 ICA)
Classifier - V:	200 ICA features extracted from 3D DCT coefficients of face blocks of $V_{64x48x16}$ (280 x 35 DCT coefficients reduced to 200 ICA)

Table 3. Comparative results of individual classifiers

Class	Surprise	Happiness	Sadness	Fear	Disgust	Anger	Total
# of sequences	70	83	49	45	40	35	322
Classifier - I	84.3	80.7	87.8	46.7	87.5	77.14	78.3
Classifier - II	94.3	100	85.7	51.1	87.5	77.14	85.7
Classifier - III	95.7	97.6	81.6	57.8	90	80	86.3
Classifier - IV	85.7	80.7	83.7	80	92.5	82.9	83.9
Classifier - V	87.1	78.3	79.6	84.4	90	85.7	83.5

our fusion approach and those of alternative methods in the literature, all being conducted on Cohn−Kanade facial expression database. Although we cannot make direct and fair comparisons among these alternative approaches due to different experimental setups and preprocessing conditions (e.g., manual or automatic registration and alignment of face images), nevertheless the scores gives an overall picture of relative performances. Presently our approach gives the second best reported overall recognition results on the Cohn-Kanade facial expression database. Notice that our proposed expression recognition system achieves slightly better recognition rates for surprise, happiness, sadness and disgust expressions when compared with Zhao's [6] recognition rates with the same expressions. Note that In Zhou's study [17] 112 sequences from 30 randomly selected subjects of Cohn-Kanade database is used for classification of 5 expression classes (happy, sad, fear angry and surprise), and they achieved 95.9% classification accuracy.

Table 4. Confusion matrix of the 6 facial expression classes with decision fusion (average recognition rate 95.34)

	Surprise	Happiness	Sadness	Fear	Disgust	Anger
Surprise	100					
Happiness		97.6	1.2			1.2
Sadness			98			2
Fear	2.2	6.7		86.7	4.4	
Disgust				2.5	95	2.5
Anger		2.9	2.9		5.6	88.6

Table 5. Performance comparison with alternative methods from the literature. Since experimental conditions vary, albeit slightly, we report for each method the number of subjects (#P), the number of sequences (#S) and the number of classes analyzed (# C) (SI:Subject Independent; LOSO: leave-one-subject-out)

ref.	Zhao07 [6]	Shan09 [16]	Yang09 [10]	Tian04 [15]	Bartlett05 [11]	**Ours**
(P,S,C)	(97,374,6)	(96,320,7)	(96,300,6)	(97,375,6)	(90,313,7)	(92,322,6)
test	10-fold SI	10-fold SI	1-fold SI	10-fold	LOSO SI	10-fold SI
Surprise	98.65	92.5	99.8	–	–	**100**
Happiness	96.04	90.1	**99.1**	–	–	97.6
Sadness	95.89	61.2	**97.8**	–	–	**97.8**
Fear	**94.64**	70	91.6	–	–	86.7
Disgust	94.74	92.5	94.1	–	–	**95**
Anger	**97.88**	66.6	97.3	–	–	88.6
Neutral	–	95.2		–	–	–
Total	**96.26**	95.1(6 class) 92.6(7 class)	–	94	93.8	**95.34**

We observed that fear and anger expressions are the most difficult categories for the classification. One of the reason is that, when we check the annotations made by different research groups it is found that annotators are not always agree for the label of fear and anger videos. The same expression video is inferred as different emotional expressions from different observers. Therefore, interpretation of fear and anger expressions is not an easy task even for humans.

We have the following observations:

- Global DCT features can discriminate the happiness and surprise expressions better than local block based DCT features. This is evidenced by classifiers II and III in Table 2 which have significantly higher performance among the six. Their average on the two positive emotions is 97.5% and 96.65% respectively. We can conjecture that these expressions cause relatively holistic changes on the face.
- Local block-based DCT features, in contrast, are more effective for the classification of negative expressions such as fear and anger, the two which are most often confused. The average performance of classifiers V and VI on the four negative emotions are 84.8% and 84.9% respectively. Thus partitioning the face into 21 or 35 blocks enables more local information capture to discriminate subtle local appearance changes.
- We observe that spatiotemporal 3D DCT features provide better overall recognition performance than landmark coordinate features. In fact, the average score of the former is 84.9% while that of the landmarks is 78.3%. The only exception is that the landmark coordinate features recognize sadness class slightly better than the 3D DCT features.

– Decision fusion of the classifiers is beneficial in that overall recognition rate is improves by 9 percentage points vis--vis the best individual classifier, Classifier III. Apparently all individual classifiers have a net contribution to the decision fusion.

5 Conclusions and Future Work

In this study we have introduced two novel facial expression recognition algorithms. The first approach analyzes expressions using a spatiotemporal prism consisting of temporal juxtaposition of aligned faces. The second approach forms vectors from several landmark trajectories and projects them onto the ICA subspace. The method based on DCT coefficients of the spatiotemporal prism is found superior to that based on subspace projection of landmark trajectories. Finally fusion of a variety of classifiers at different parameter settings and/or operating on different data modalities improves the overall emotion recognition accuracy significantly.

The next step of our study will be to evaluate the performance of our proposed algorithm on real world situations where expressions are spontaneously and naturally occurring. Spontaneous emotional facial expressions include head motions and facial expressions. The analysis of expressions under head pose has only recently been addressed in the literature. We conjecture that facial landmark based scheme will more successful in tracking expressions head movements and while the spatiotemporal prism will still be instrumental in interpreting plastic deformations of the face due to expressions. One can also envision the use of boosting algorithms, e.g., Adaboost, for the selection of appropriate features from the plethora of given ones.

References

1. Vinciarelli, A., Pantic, M., Bourlard, H.: Social signal processing: Survey of an emerging domain. Image and Vision Computing 27, 1743–1759 (2009)
2. Gatica-Perez, D.: Automatic nonverbal analysis of social interaction in small groups: A review. Image and Vision Computing 27, 1775–1787 (2009)
3. Sebe, N., Lew, M., Sun, Y., Cohen, I., Gevers, T., Huang, T.: Authentic facial expression analysis. Image and Vision Computing 25, 1856–1863 (2007)
4. Zeng, Z., Pantic, M., Roisman, G., Huang, T.: A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence 31, 39–58 (2009)
5. Hupont, I., Cerezo, E., Baldassarri, S.: Facial emotional classifier for natural interaction, vol. 7, pp. 1–12 (2008)
6. Zhao, G., Pietikainen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, 915–928 (2007)
7. Zhang, Y., Ji, Q.: Active and dynamic information fusion for facial expression understanding from image sequences. IEEE Transactions on Pattern Analysis and Machine Intelligence 27, 699–714 (2005)

8. Wang, T., James Lien, J.J.: Facial expression recognition system based on rigid and non-rigid motion separation and 3D pose estimation. Pattern Recognition 42, 962–977 (2009)
9. Pantic, M., Patras, I.: Dynamics of facial expression: Recognition of facial actions and their temporal segments from face profile image sequences. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 36, 433–449 (2006)
10. Yang, P., Liu, Q., Metaxas, D.N.: Boosting encoded dynamic features for facial expression recognition. Pattern Recognition Letters 30, 132–139 (2009)
11. Bartlett, M.S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., Movellan, J.: Recognizing facial expression: Machine learning and application to spontaneous behavior. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 568–573 (2005)
12. Dornaika, F., Davoine, F.: Simultaneous facial action tracking and expression recognition in the presence of head motion. International Journal of Computer Vision 76, 257–281 (2008)
13. Tsalakanidou, F., Malassiotis, S.: Real-time 2D+3D facial action and expression recognition. Pattern Recognition 43, 1763–1775 (2010)
14. Ekman, P., Friesen, W.: Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto (1978)
15. Tian, Y.L.: Evaluation of face resolution for expression analysis. In: CVPR Workshop on Face and Video, p. 82 (2004)
16. Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on local binary patterns a comprehensive study. Image and Vision Computing 27 (2009)
17. Zhou, F., De la Torre, F., Cohn, J.: Unsupervised discovery of facial events. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (2010)
18. Bartlett, M.S., Littlewort, G., Frank, M.G., Lainscsek, C., Fasel, I.R., Movellan, J.R.: Automatic recognition of facial actions in spontaneous expressions. Journal of Multimedia 1, 22–35 (2006)
19. Kanade, T., Cohn, J., Tian, Y.L.: Comprehensive database for facial expression analysis. In: Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2000, pp. 46–53 (2000)
20. Zeng, Z., Fu, Y., Roisman, G.I., Wen, Z., Hu, Y., Huang, T.S.: Spontaneous emotional facial expression detection. Journal of Multimedia 1, 1–8 (2006)
21. Zeng, Z., Hu, Y., Fu, Y., Huang, T.S., Roisman, G.I., Wen, Z.: Audio-visual emotion recognition in adult attachment interview. In: Proceedings of the 8th International Conference on Multimodal Interfaces, pp. 139–145. ACM, New York (2006)
22. Akakin, H.C., Sankur, B.: Analysis of Head and Facial Gestures Using Facial Landmark Trajectories. In: Fierrez, J., Ortega-Garcia, J., Esposito, A., Drygajlo, A., Faundez-Zanuy, M. (eds.) BioID_MultiComm2009. LNCS, vol. 5707, pp. 105–113. Springer, Heidelberg (2009)
23. Martinez, A.M.: Matching expression variant faces. Vision Research, 1047–1060 (2003)
24. Oja, E.: Independent component analysis: algorithms and applications. Neural Networks 13, 411–430 (2000)