

A Dynamic Approach for Detecting Naturalistic Affective States from Facial Videos during HCI

Hamed Monkaresi, M.S. Hussain, and Rafael A. Calvo

School of Electrical and Information Engineering, University of Sydney, Australia
{hamed.monkaresi, sazzad.hussain, rafael.calvo}@sydney.edu.au

Abstract. Significant progress has been made in automatic facial expression analysis using facial images and videos. The recognition reliability of most current approaches is still poor in naturalistic expressions compared to acted ones. Most of these methods use a static image of each expression that captures the characteristic image at the apex. However, according to psychologists, analyzing a sequence of images in a dynamic manner produces more accurate and robust recognition of facial affect expressions. In this paper, a new dynamic model is proposed for detecting naturalistic affect expressions. The Local Binary Pattern in Three Orthogonal Planes (LBP-TOP) is considered for modeling appearance and motion of facial features. The International Affective Picture System (IAPS) collection was used as stimulus for triggering naturalistic affective states. The dynamic approach produced an improvement of 16% for valence classification and 22% for arousal classification over previous studies.

Keywords: Affective computing, dynamic texture, facial expression recognition.

1 Introduction

It is generally accepted that computing is part of the fabric of our everyday living. We are being increasingly surrounded by cameras, recording our expressions during everyday interactions, whether on a mobile phone, a tablet or a PC. In order to offer the most effective interactions between human and machine, these devices with cameras will need the capacity to perceive and understand human expressions of emotion in realistic scenarios. Facial expression is one of the most cogent, naturally preeminent means for human beings to express emotions, comprehension, agreement or disagreement, and intentions, which regulate interactions with the environment and the people in it [1]. Particularly for the recognition of affective states, humans rely heavily on analyzing facial expressions [2]. While humans routinely extract much of this information automatically in real life situations, the systematic classification and extraction of facial expression information in the laboratory settings has proven to be very challenging. Facial Expression Recognition (FER) systems have been proposed for addressing these challenges [3], [4].

Until not long ago, the most commonly used labels in affect detection were the six basic emotions (fear, sadness, happiness, anger, disgust, and surprise), proposed by

Ekman [5], who suggested that these emotions are universally displayed and recognized from facial expressions. He also proposed the Facial Action Coding System (FACS) [6] which is a widely used method for describing facial muscle actions and corresponding expressions.

One of the inherent difficulties with the FACS coding scheme is that it requires a highly trained human expert to manually score each frame of a video. The development of a system that automatically detects the action units (AUs) is a challenging task because the coding system was originally created for static pictures rather than changing expressions over time. Although there has been remarkable progress in this area, the reliability of current automatic AU detection systems do not match the accuracy of humans [7]. This problem is more challenging in spontaneous expression recognition [8].

Two main analytic approaches in FER research considered static images and sequence of images. Most of the research on this area has been based on static images [9] or individual frames of an image sequence [10] and some research efforts toward using temporal models for facial expression recognition [11]. Psychological studies have suggested that facial motion is fundamental clue to the recognition of facial expressions. In addition, Bassili [12] demonstrated that humans are better at recognizing expressions from dynamic images compared to static ones. Few systems attempt to recognize fine-grained changes in the facial expression using dynamic information to analyze facial expression. In the simplest case, the change over consecutive frames or the change with respect to a neutral frame is used to determine the underlying expression [13].

On the other hand, most of the existing datasets focused on posed emotion rather than spontaneous ones [3]. Posed expressions are typically exaggerated and their dynamics are generally much stronger than in spontaneous day-to-day facial expressions, which make them a natural place to start training expression recognition systems. While FER systems should ideally consider spontaneous natural emotion for affect analysis, such approaches have rarely been explored [14].

In our previous study [15], we implemented a geometric-based system to detect valence and arousal through head movement and changes in skin color. Reasonable accuracy was achieved for user-independent analysis. In this study, we have used the same dataset that is used in [15] to compare the performance of our new proposed model. Here, we propose a dynamic texture-based model, which considers the appearance and motion of facial displays simultaneously. The appearance and the motion of facial objects like eyes, eyebrows and mouth during a certain period of time can be valuable sources for detecting spontaneous affective states. These dynamic facial features have been extracted using local binary patterns in three orthogonal planes (XY, XT and YT). The XY plane provides information related to appearance and the XT and YT planes contain the information related to the motion of facial objects. Another novelty of this work is automatic alignment of facial objects. In a similar work [16], the position of eyes had been used for alignment which were set manually. According to the nature of our dataset, which contains rigid head motions, we have implemented an automatic facial object tracker for extracting the facial features.

2 Related Research

Two main approaches have been followed in the area of facial expression analysis; geometric-based and appearance-based approaches. Geometric features include shapes and positions of face components, and the location of fixed facial points [9] such as the corners of the eyes, eyebrows, etc. In most cases, the position of these components and fixed points are detected in the first frame, and motion of these objects are tracked throughout the sequence. A geometric approach that attempts to automatically detect temporal segments of AUs was used by Pantic et al. [17], [18].

Appearance-based methods analyze the deformations of the face skin in both static and dynamic space for recognizing facial expression. Dynamic texture-based method can be seen as a generalization of appearance-based approaches. FER systems which use appearance-based features have been proposed in [10]. Several researchers have used Gabor wavelet coefficients as features [19]. Bartlett et al. [10] have tried different methods, such as explicit feature measurement, Independent Component Analysis (ICA), and Gabor wavelets. Finally they reported that Gabor wavelets provide the best results [20]. Other techniques explored in this field include optical flow [21] and Active Appearance Models [22].

Both geometric-based and appearance-based approaches have advantages and disadvantages. Geometric-based methods mostly rely on the motion of a number of points, and ignore much of the information related to skin texture changes. In contrast, appearance-based methods may be more susceptible to changes in illumination, rigid head motions and differences between individuals [3]. Tian et al. [23] used a combination of geometric-based and appearance-based features (Gabor wavelets) for recognizing facial AUs. They claimed that the former features outperform the latter ones, yet using both yields the best result.

2.1 Dynamic Texture Approach

Dynamic Texture (DT) recognition is an emerging new method of appearance-based activity recognition. Dynamic or temporal textures refer to textures with motion. A DT can be defined as a "spatially repetitive, time-varying visual pattern that forms an image sequence with certain temporal stationarity" [24]. Typical examples of DTs are smoke, fire, sea waves, and talking faces. Many existing approaches for recognition of DTs are based on optical flow [25]. A different approach can be seen in [26], where the system identification techniques are used to learn generative models.

Facial expression recognition can also be defined as one of the most suitable application for the DT recognition techniques. Zhao and Pietikäinen [16] used Volume Local Binary Patterns (VLBPs) for facial expression recognition. VLBPs are a temporal extension of local binary patterns often used in 2-dimensional texture analysis for recognizing facial expressions. In the study [16], the face was divided into overlapping blocks and the extracted features in each block were concatenated into a single feature vector for classification (with SVM). The approach showed promising results, although only the six prototypic emotions were recognized and no temporal segmentation was performed. They normalized the face using the eye position in the

first frame, but they ignore any rigid head movement that may occur during the sequence. In addition, they used fixed overlapping blocks distributed evenly over the face instead of focusing on specific regions of the face such as mouth, eyes and eyebrows, which include valuable information about facial expressions. The second attempt at using DT in facial expression analysis was performed by Koelstra et al. [27]. They estimated non-rigid motion between consecutive frames by applying either non-rigid registration using Free-form Deformations (FFDs) or Motion History Images (MHIs). For each AU, a quad-tree decomposition was defined to identify face regions related to the AU. In these regions, orientation histogram feature descriptors were extracted. Finally, a combined Gentle-Boost classifier and a Hidden Markov Model (HMM) were used to classify the sequence in terms of AUs and their temporal segments. However, increasing evidence suggests that deliberate behavior differs in visual appearance, audio profile, and timing from spontaneously occurring behavior.

3 Methodology

3.1 LBP-TOP

The LBP proposed by Ojala et al. [28] is one of the powerful methods for texture description. The LBP operator labels the neighborhood region of each pixel of an image by thresholding the pixels with the central value. Considering P neighborhood-pixels, 2^P different micro-patterns could be addressed. By calculating LBP for all pixels in an image and calculating the distribution of each pattern, a specific histogram could be extracted for each image. This LBP histogram is a powerful indicator of the image, which has been proved to be successful in different pattern recognition applications. By defining different radius (R) and number of neighboring points (P), several types of LBP could be extracted. The best values for R and P depends on the application and general characteristics of image sets. An example of circular 8 neighboring pixels with $R=1$ is presented in Figure 1.

The LBP operator was originally designed for static images. Recently, Zhao and Pietikäinen [16] proposed an extended version of LBP to describe dynamic textures. Usually, a video sequence is thought of as a stack of XY planes in axis T, but it is easy to ignore that a video sequence can also be seen as a stack of XT planes in axis Y and YT planes in axis X, respectively. They divided each video sequence into three orthogonal sets of 2-dimensional planes. The LBP could be computed for each set of planes separately. The Local Binary Pattern in Three Orthogonal Planes (LBP-TOP) [16] descriptor for each video clip is calculated by concatenating three LBP histograms. Figure 2, shows the LBP-TOP procedure. In such a representation, DT is encoded by the XY-LBP, XT-LBP, and YT-LBP, whereas the appearance and motion in three directions of DT are considered, incorporating spatial domain information (XY-LBP) and two spatial temporal co-occurrence statistics (XT-LBP and YT-LBP).

With this approach, the number of bins is only 3×2^P , which makes the extension to many neighboring points easier and also reduces the computational complexity compare with volume-based LBP (VLBP) method [16].

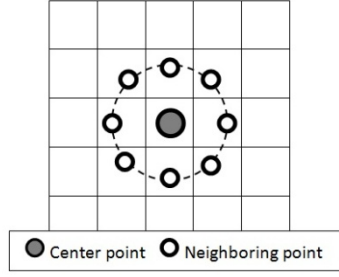


Fig. 1. An example of LBP operator ($P=8$, $R=1$)

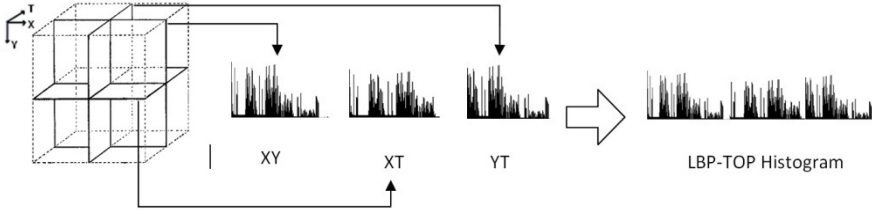


Fig. 2. LBP-TOP procedure: Extracting LBP features from each set of planes and concatenating to a single histogram [16]

Like the LBP representation, the radius in axes X , Y , and T and the number of neighboring points in the XY , XT , and YT planes can also be different, which can be marked as R_X , R_Y , and R_T , P_{XY} , P_{XT} , and P_{YT} . The corresponding feature is denoted as $LBP-TOP_{P_{XY};P_{XT};P_{YT};R_X;R_Y;R_T}$.

3.2 Feature Extraction

An LBP description computed over the whole facial expression sequence encodes only the occurrences of the micro-patterns without any indication about their locations. One solution for this effect is dividing an image to several blocks. The LBP histograms for each block are calculated and concatenated to a single histogram. If the image divides into N blocks, the number of bins in final histogram would be $N \times 3 \times 2^P$. In our proposed model, three fixed blocks correspond to left-eye, right-eye and mouth were considered for each image sequence. So, for each video clip, $3 \times 3 \times 2^P$ features were extracted. These regions were detected and extracted automatically using an extended boosted cascade classifier [29]. Then, deformation of eyes and mouth were monitored during expression of an affective state. In order to have the same size of blocks in each image, the detected objects were resized to fixed sizes. In this study, the size was same for all video clips (10 seconds).

Three variations of LBP-TOP method were used for feature extraction which were different in the radius and the number of neighbor points ($LBP-TOP_{8,8,8,1,1,1}$,

LBP-TOP_{6,6,6,2,2,2} and LBP-TOP_{8,8,8,3,3,3}). The classification results for each set of features are shown in the section 5. A total of 2304 features were extracted by LBP-TOP_{8,8,8,1,1,1} and LBP-TOP_{8,8,8,3,3,3} while 576 features were extracted by LBP-TOP_{6,6,6,2,2,2} operator.

Feature selection techniques were applied prior to classification to reduce dimensionality, which automatically removed unnecessary features. To serve the purpose, the correlation based feature selection (CFS) method was used for choosing the best subset of features. This technique evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them [30].

3.3 Classification

All videos were synchronized with time stamps corresponding to the presented IAPS images and their normative ratings. The features extracted from videos for each 10-second windows were labeled using the normative ratings and self-reports (three levels of valence and arousal). A Matlab based computational framework as part of the Sien-to framework [31], was used for feature selection and classification. The feature selection was implemented in Matlab using the DMML¹ wrapper for Weka. The classification was performed in Matlab using MatlabArsenal², a wrapper for the classifiers in Weka. Three machine learning algorithms; k-nearest neighbor ($k=1$), linear support vector machine (SVM) and J48 decision tree were selected for classification. Then, a vote classifier with the average probability rule for combining the classifiers was applied [32]. The training and testing for both types of classes (valence and arousal) was performed separately with a 10-fold cross validation. The classification accuracy was used as the overall classification performance metric. In addition, the F-measure (from precision and recall) was calculated as an indication of how well each affective state was classified. For the classification scores of precision (P) and recall (R), the F-measure ($F1$) is calculated by; $F1=2((P \times R)/(P+R))$.

4 Experiment

4.1 Participants and Materials

The participants were 20 undergraduate/postgraduate engineering students at the University of Sydney. The participants' age ranged from 18 to 30 years and there were 8 males and 12 females. Due to major eye occlusion caused by eyeglasses from 7 subjects, results are presented for 13 subjects. This study was approved by the University of Sydney's Human Ethics Research Committee prior to data collection. The participants signed an informed consent prior to the experiment. The experiment, which took approximately an hour was conducted indoors with a varying amount of ambient sunlight entering through windows in combination with normal artificial fluorescent

¹ DMML: featureselection.asu.edu/software.php

² MatlabArsenal: cs.siu.edu/~qcheng/featureselection/index.html

light. Participants were asked to sit in front of a computer and interact normally while their video was recorded by an ordinary webcam (Logitech Webcam Pro 9000). All videos were recorded in color (24-bit RGB with 3 channels, 8 bits/channel) at 15 frames per seconds (fps) with pixel resolution of 640×480 pixels and saved in AVI format.

4.2 Procedure

The participants viewed emotionally charged photos from the IAPS collection. A total of 90 images (three blocks of 30 images each) for 10 seconds each were presented, followed by 6 seconds pauses between the images. The images were selected so that the IAPS valence and arousal scores for the stimuli spanned a 3×3 valence/arousal space (IAPS normative ratings). Participants also self-reported their emotions by clicking radio buttons on the appropriate location of 3×3 valence/arousal grid after viewing each image. In this paper, results are presented using the normative ratings instead of self-reports. Therefore, the computational model was trained and tested using a balanced class distribution, which could be suitable for evaluating accuracies of classification without applying any up or down sampling techniques. The normative ratings are useful because they are standardized scientifically for assessing basic and applied problems in psychology [33]. Moreover, many people do not know how to recognize, express and label/scale their own feelings, therefore self reports sometimes can be unreliable [34]. However, self-reports provide important information and should not be ignored; therefore the collected self ratings will be used as an extension of this work in future studies.

5 Results

The classification for detecting 3-degrees of valence (positive, neutral and negative) and arousal (low, medium, high) are presented in two user models: user-dependent models and combined model. For each set of model, the extracted features from the three methods (LBP-TOP_{8,8,8,1,1,1}, LBP-TOP_{6,6,6,2,2,2} and LBP-TOP_{8,8,8,3,3,3}) are evaluated and discussed in the following subsections.

5.1 User-Dependent Models

The subsequent analysis focuses on developing user-dependent models. Distinct models were developed and validated for each participant. Figure 3 presents the mean and standard deviation of classification accuracies assessing the overall performance of discriminating 3-degrees of valence and arousal using three proposed methods (LBP-TOP_{8,8,8,1,1,1}, LBP-TOP_{6,6,6,2,2,2} and LBP-TOP_{8,8,8,3,3,3}). The last set of bars, indicate the best result achieved from our recent method for detecting valence and arousal levels [15]. The new results indicate that the classifier was more accurate in discriminating three levels of valence and arousal for all three dynamic texture-based methods compared to [15]. The accuracy of valence detection was improved by 16% using

LBP-TOP_{8,8,8,3,3,3} features. The 22% improvement was also achieved by LBP-TOP_{8,8,8,3,3,3} features for arousal detection. This finding demonstrate the power of the proposed dynamic texture method in detecting user's affective states compare with the geometric-based method in [15].

According to the results, the number of neighboring points influence the performance of affect detection. LBP-TOP_{8,8,8,1,1,1} and LBP-TOP_{8,8,8,3,3,3} produced almost the same results whereas LBP-TOP_{6,6,6,2,2,2} failed to improve the performance of the classification. This might indicate the importance of the considering number of neighborhood points compare with the radius in the circular LBP-TOP operator. However, considering more neighboring points would increase the computational complexity.

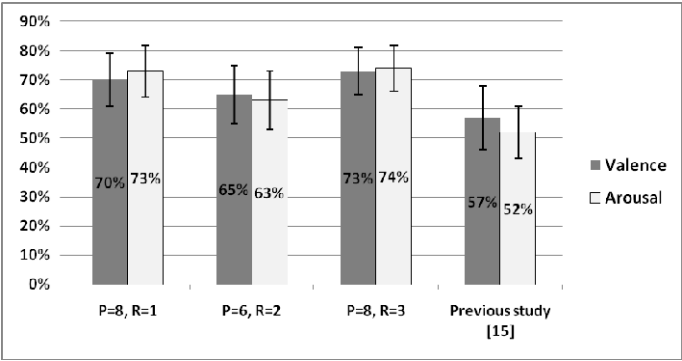


Fig. 3. The mean and standard deviation of the classification accuracy for detecting 3 levels of valence and arousal (User-dependent models)

Next we investigate the classification accuracy of the individual degrees of valence and arousal. Table 1 shows the mean and standard deviation of *F1* as per valence and arousal category across the 13 subjects for the vote classifier.

Table 1. The mean and standard deviation of *F1* values for detecting 1-3 degrees of valence and arousal (user-dependent models)

		Valence			Arousal		
		Low	Medium	High	Low	Medium	High
LBP-TOP _{8,8,8,1,1,1}	Mean	0.76	0.62	0.72	0.75	0.66	0.76
	Std	0.10	0.11	0.15	0.11	0.11	0.10
LBP-TOP _{6,6,6,2,2,2}	Mean	0.72	0.57	0.65	0.66	0.53	0.70
	Std	0.14	0.09	0.17	0.16	0.12	0.10
LBP-TOP _{8,8,8,3,3,3}	Mean	0.79	0.64	0.76	0.76	0.68	0.78
	Std	0.10	0.11	0.12	0.09	0.10	0.11

The LBP-TOP_{8,8,8,3,3,3} had the highest accuracy for detecting all three degrees of valence and arousal. Among the 3-degrees of valence, low-valence was more predictable followed by high-valence and medium-valence respectively. The same pattern

was seen for arousal classification. The performance of all proposed methods for detecting high-arousal showed the best result whereas medium-arousal could not predict well enough using DT based methods.

5.2 Combined Model

Data from individual participants was first standardized (converted to z-scores) to address individual variations of features caused by differences in skin colors or environmental illumination. Then the standardized instances were combined to yield one large data set with 1154 instances. The dimensionality of the data was also reduced by selecting the best features prior to each classification task by using CFS. Figure 4 shows the classification results for 3-degrees of valence and arousal using three variations of LBP-TOP operators for feature extraction. The set of last bars, indicate the best result achieved for the combined model from [15]. The LBP-TOP methods showed better performance than our previous method [15]. The accuracy of three levels of valence detection was raised by increasing the radius (R) of circular LBP-TOP operator. The LBP-TOP_{8,8,8,3,3,3} showed the best performance in detecting valence and arousal. Accordingly for general model we can argue that besides increasing number of neighboring points (P), increasing the radius (R) of the LBP-TOP circular operator could improve the accuracy of affect detection. Achieving the lower accuracy compared to user-dependent models was expected, because there is a mixture of instances from different participants in combined model. This model could be extended to a general (user-independent) model.

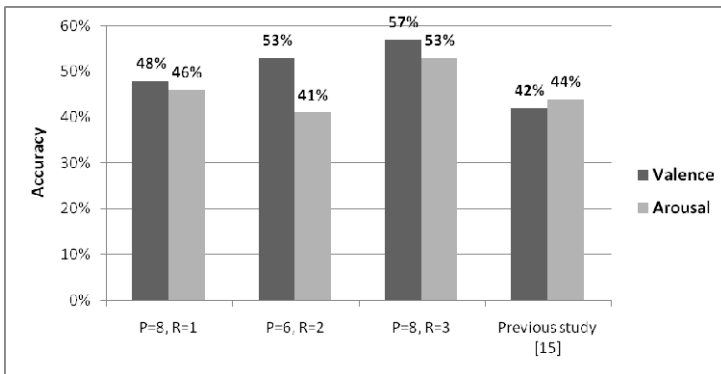


Fig. 4. The classification accuracy for detecting three levels of valence and arousal (Combined model)

Feature Selection Analysis. Analyzing the selected features using CFS gives us a better insight about the important features for detecting affect. Therefore, we investigate the selected features from the combined model. The selected features were divided into two groups; appearance-based features, which were selected from XY planes and motion-based features, which were selected from XT and YT planes.

Figure 5 gives the proportion of features that contribute from each group (appearance-based and motion-based) for valence and arousal. Almost the same proportion of appearance-based features and motion-based features were observed for valence in all three methods. As for arousal, appearance-based features had very high contribution using LBP-TOP_{8,8,8,3,3,3} and motion-based features had very high contribution using LBP-TOP_{8,8,8,1,1,1}. Despite the differences in valence and arousal, this analysis reflects that both appearance-based and motion-based features are essential for detecting affective states in naturalistic HCI.

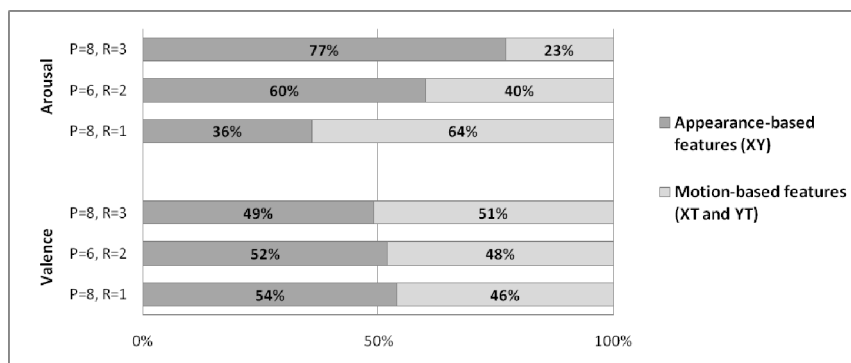


Fig. 5. The percentage of selected features in valence and arousal analysis

6 Conclusion

This study investigated dynamic features of facial expressions for the detection of affective states during naturalistic interactions. Both the user-independent model and a combined model have been evaluated for classifying 3-degrees of valence and arousal.

Results indicate that the best method for extracting dynamic features (LBP-TOP_{8,8,8,3,3,3}) can achieve 73% and 74% accuracies for detecting degrees of valence and arousal respectively with the user-dependent model. As for the combined model, 57% and 53% accuracies were achieved valence and arousal respectively. For both models, the dynamic features have shown improvement over our previous study using chromatic and movement features [15].

In this study the correlation based feature selection algorithm and a vote classifier was used. In future work, other machine learning techniques along with other video based or multimodal features can be investigated to improve the accuracy.

References

1. Ekman, P., Rosenberg, E.L.: What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS). Oxford University Press, USA (2005)

2. Picard, R.W.: *Affective Computing*. The MIT Press (2000)
3. Zeng, Z., Pantic, M., Roisman, G.I., Huang, T.S.: A survey of affect recognition methods: audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 39–58 (2009)
4. Calvo, R.A., D'Mello, S.: Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. *IEEE Transaction on Affective Computing* 1, 18–37 (2010)
5. Ekman, P., Friesen, W.V.: Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology* 17, 124–129 (1971)
6. Ekman, P., Friesen, W.V.: *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. Consulting Psychologists Press, Palo Alto (1978)
7. D'Mello, S., Graesser, A.: Multimodal semi-automated affect detection from conversational cues, gross body language, and facial features. *User Modeling and User-Adapted Interaction* 20, 147–187 (2010)
8. Hoque, M.E., el Kaliouby, R., Picard, R.W.: When Human Coders (and Machines) Disagree on the Meaning of Facial Affect in Spontaneous Videos. In: Ruttkay, Z., Kipp, M., Nijholt, A., Vilhjálmsón, H.H. (eds.) *IVA 2009. LNCS*, vol. 5773, pp. 337–343. Springer, Heidelberg (2009)
9. Pantic, M., Rothkrantz, L.J.M.: Facial action recognition for facial expression analysis from static face images. *Systems, Man, and Cybernetics* 34, 1449–1461 (2004)
10. Bartlett, M.S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., Movellan, J.: Fully Automatic Facial Action Recognition in Spontaneous Behavior. In: *7th International Conference on Automatic Face and Gesture Recognition (FGR 2006)*, pp. 223–230. IEEE (2006)
11. Cohen, I., Sebe, N., Garg, A., Chen, L.S., Huang, T.S.: Facial expression recognition from video sequences: temporal and static modeling. *Computer Vision and Image Understanding* 91, 160–187 (2003)
12. Bassili, J.N.: Emotion recognition: the role of facial movement and the relative importance of upper and lower areas of the face. *Journal of Personality and Social Psychology* 37, 2049–2058 (1979)
13. Michel, P., El Kaliouby, R.: Real time facial expression recognition in video using support vector machines. In: *Proceedings of the 5th international conference on Multimodal Interfaces - ICMI 2003*, p. 258 (2003)
14. Hoque, M.E., McDuff, D.J., Picard, R.W.: Exploring Temporal Patterns in Classifying Frustrated and Delighted Smiles. *IEEE Transaction on Affective Computing* (2012)
15. Monkaresi, H., Calvo, R.A., Sazzad Hussain, M.: Automatic natural expression recognition using head movement and skin color features. In: Tortora, G., Levialdi, S., Tucci, M. (eds.) *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI 2012)*, pp. 657–660. ACM, New York (2012)
16. Zhao, G., Pietikäinen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Trans. Pattern Analysis and Machine* 29, 915–928 (2007)
17. Pantic, M., Patras, I.: Dynamics of Facial Expression: Recognition of Facial Actions and Their Temporal Segments. *IEEE Trans. Systems, Man, and Cybernetics* 36, 433–449 (2006)
18. Valstar, M.F., Pantic, M.: Combined Support Vector Machines and Hidden Markov Models for Modeling Facial Action Temporal Dynamics. In: Lew, M., Sebe, N., Huang, T.S., Bakker, E.M. (eds.) *HCI 2007. LNCS*, vol. 4796, pp. 118–127. Springer, Heidelberg (2007)

19. Guo, G., Dyer, C.R.: Learning from examples in the small sample case: face expression recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B* 35, 477–488 (2005)
20. Littlewort, G., Bartlett, M.S., Fasel, I., Susskind, J., Movellan, J.: Dynamics of facial expression extracted automatically from video. *Image and Vision Computing* 24, 615–625 (2006)
21. Anderson, K., McOwan, P.W.: A real-time automated system for the recognition of human facial expressions. *IEEE Transactions on Systems, Man, and Cybernetics, Part B* 36, 96–105 (2006)
22. Lucey, S., Ashraf, A.B., Cohn, J.F.: Investigating Spontaneous Facial Action Recognition through AAM Representations of the Face. In: Delac, K., Grgic, M. (eds.) *Face Recognition*, pp. 275–286. I-Tech Education and Publishing (2007)
23. Tian, Y., Kanade, T., Cohn, J.F.: Evaluation of Gabor-Wavelet-Based Facial Action Unit Recognition in Image Sequences of Increasing Complexity. In: *Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 229–234 (2002)
24. Chetverikov, D., Renaud, P.: A brief survey of dynamic texture description and recognition. In: *Proc. Conf. Computer Recognition Systems*, pp. 17–26 (2005)
25. Polana, R., Nelson, R.: Temporal Texture and Activity Recognition. In: Shah, M., Jain, R. (eds.) *Motion-Based Recognition*, pp. 87–115 (1997)
26. Saisan, P., Doretto, G., Wu, Y.: Dynamic texture recognition. In: *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 58–63 (2001)
27. Koelstra, S., Pantic, M., Patras, I.: A dynamic texture-based approach to recognition of facial actions and their temporal models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, 1940–1954 (2010)
28. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition* 29, 51–59 (1996)
29. Castrillón, M., Déniz, O., Guerra, C., Hernández, M.: ENCARA2: Real-time detection of multiple faces at different resolutions in video streams. *Journal of Visual Communication and Image Representation* 18, 130–140 (2007)
30. Hall, M.A.: Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning. In: *ICML 2000 Proceedings of the Seventeenth International Conference on Machine Learning*, pp. 359–366 (2000)
31. Calvo, R.A., Hussain, M. S., Aghaei Pour, P., Alzoubi, O.: Siento: An Experimental Platform for Behavior and Psychophysiology in HCI. In: D’Mello, S., Graesser, A., Schuller, B., Martin, J.-C. (eds.) *ACII 2011, Part II. LNCS*, vol. 6975, pp. 307–308. Springer, Heidelberg (2011)
32. Kuncheva, L.I.: *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons (2004)
33. Lang, P., Bradley, M.: *International affective picture system (IAPS): Technical manual and affective ratings*. Psychology (1997)
34. Picard, R.: Affective computing: challenges. *International Journal of Human-Computer Studies* 59, 55–64 (2003)