Research Article ISSN: 2319-507X
S Bala Brahmam, IJPRET, 2013; Volume 2 (5): 1-20 IJPRET

INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

DESIGN OF EMBEDDED LINUX SYSTEM FOR FACIAL EXPRESSION DETECTION

S. BALA BRAHMAM[1], MRS. NIRMALA DEVI[2]

1. Student, M. Tech (Embedded Systems), Aurora technological and research institute, Uppal, HYD.
2. Associate Professor, Aurora technological and research institute, Uppal, HYD

Accepted Date: 16/11/2013 ; Published Date: 01/01/2014

Abstract: Facial expression recognition has been an active topic in electronics and computer science for over two decades, in particular facial action coding system action unit (AU) detection and classification of a number of discrete emotion states from facial expressive imagery using HAAR like features. Standardization and comparability have received some attention; for instance, there exist a number of commonly used facial expression databases. Here the system uses OpenCV library for capturing video stream. From that stream expression of the face is detected. This paper presents an embedded system of the first such challenge in automatic recognition of facial expressions. It details the challenge data, evaluation protocol, and the results attained in two sub-challenges: AU detection and classification of facial expression imagery in terms of a number of discrete emotion categories.

Keywords: Computers, Recognition, Facial Expression

Corresponding Author: MR. S. BALA BRAHMAM

Access Online On:

www.ijpret.com

How to Cite This Article:

S Bala Brahmam, IJPRET, 2013; Volume 2 (5): 1-20

PAPER-QR CODE

INTRODUCTION

COMPUTERS and other powerful electronic devices sur- round us in ever increasing numbers, with their ease of use continuously being improved by user-friendly interfaces. Yet, to completely remove all interaction barriers, the next- generation computing (a.k.a. pervasive computing, ambient intelligence, and human computing) will need to develop human-centered user interfaces that respond readily to naturally occurring multimodal human communication [39]. An important functionality of these interfaces will be the capacity to perceive and understand the user's cognitive appraisals, action tendencies, and social intentions that are usually associated with emotional experience. Because facial behavior is believed to be an important source of such emotional and interpersonal information [2], automatic analysis of facial expressions is crucial to human–computer interaction.

Even though automatic classification of face imagery in terms of the six basic emotion categories is considered largely solved, reports on novel approaches are published even to date (e.g., [28], [35], [49], and [52]). While, in truth, such systems directly to a small number of discrete affective states such as the six basic Most facial expression analysis systems developed so far adhere independent from any interpretation attempt, leaving the often impossible to measure, sign judgment is agnostic, often impossible to measure, sign judgment is agnostic, While message judgment is all about interpretation, with the judgment approach and as a facial movement that lowers and can be judged as possibly caused by "anger" in a message pulls the eyebrows closer together in a sign- judgment approach. ground truth being a hidden state that is often impossible to measure, sign judgment is agnostic, inference about the conveyed message to higher order decision making. to the message-judgment approach. They attempt to recognize a small set of prototypic emotional facial expressions said to relate emotions proposed by Ekman [15], [41], [57], [66].recognize prototypical facial expressions but not actually recognizing emotions, for brevity, we will refer to this process as "emotion recognition."

The first survey of the field was published in 1992 [44] and has been followed up by several others [20], [40], [68]. However, the question remains as to whether the approaches proposed to date actually deliver what they promise. To help answer that question, I felt that it was time to take stock, in an objective manner, of how far the field has progressed. Researchers often do report on the accuracy of the proposed approaches using a number of popular publicly available facial expression databases (e.g., the Cohn–Kanade database [26], the MMI Facial Expression database [42], [58], or the JAFFE database [33]). A periodical challenge in facial expression recognition would allow this comparison in a fair manner. It would clarify how far the field has come and would allow us to identify new goals, challenges, and targets.

Two main streams in the current research on automatic analysis of facial expressions consider facial affect (emotion) inference from facial expressions and facial muscle action de- tection [38], [41], [57], [66]. These streams stem directly from the two major approaches to facial expression measurement in psychological research [10]: message and sign judgment. The aim of the former is to infer what underlies a displayed facial expression, such as affect or personality, while the aim of the latter is to describe the outward "surface" of the shown behavior, such as facial movement or facial component shape.

Thus, a frown can be judged as possibly caused by "anger" in a message-judgment approach and as a facial movement that lowers and pulls the eyebrows closer together in a sign-judgment approach. While message judgment is all about inter- pretation, with the ground truth being a hidden state that is often impossible to measure, sign judgment is agnostic, independent from any interpretation attempt, leaving the inference about the conveyed message to higher order decision making. Most facial expression analysis systems developed so far adhere to the message-judgment approach.

They attempt to recognize a small set of prototypic emotional facial expressions said to relate directly to a small number of discrete affective states such as the six basic emotions proposed by Ekman [15], [41], [57], [66]. Even though automatic classification of face imagery in terms of the six basic emotion categories is considered largely solved, reports on novel approaches are published even to date (e.g., [28], [35], [49], and [52]). While, in truth, such systems recognize prototypical facial expressions but not actually rec- ognizing emotions, for brevity, we will refer to this process as "emotion recognition."

Here HAAR cascaded algorithms are used for detecting the facial expressions. Haar-like features are digital image features used in object recognition. They owe their name to their intuitive similarity with Haar wavelets and were used in the first real-time face detection. As AUs are independent of any interpretation, they can be used as the basis for any higher order decision making process, including recognition of basic emotions [18], cognitive states like (dis)agreement and puzzlement [11], psychological states like pain [13], and socio-cultural signals like emblems (i.e., culture-specific interactive signals like wink, coded as left or right AU46), regulators (i.e., conversational mediators like exchange of a look, coded by AUs for eye position), and illustrators (i.e., cues accompanying speech like raised eyebrows, coded as AU1+AU2) [17]. Hence, AUs are extremely suitable to be used as midlevel parameters in an automatic facial behavior analysis system as they reduce the dimension- ality of the problem [60] (thousands of anatomically possible facial expressions [17] can be represented as combinations of 32 AUs).

In terms of feature representation, the majority of the au- tomatic facial expression recognition literature can be divided into three ways: those that use appearance-based features (e.g., [7], [24], and [35]), those that use geometric-feature-based approaches (e.g., [28] and [57]), and those that use both (e.g., [3] and [54]). Both appearance- and geometric-feature-based approaches have their own advantages and disadvantages, and we expect that systems that use both for this challenge will result in the highest accuracy.

Another way that existing systems can be classified is in the way they make use of temporal information. Some systems only use the temporal dynamics information encoded directly in the utilized features (e.g., [24] and [67]), others only employ machine learning techniques to model time (e.g., [50] and [56]), while others employ both (e.g., [57]). Currently, it is unknown what approach could guarantee the best performance.

An overview of a recent literature in the field is provided in Section II. In Section III I'm going to discuss about working of HAAR like features. In Section IV, described the challenge protocol for both the AU detection and emotion recognition sub-challenges. A detailed analysis of the results attained in this challenge is given in Section V. I conclude this paper with a discussion of the challenge and its results in Section VI.

II. OVERVIEW OF EXISTING WORKS

Below, I present a short overview of the main streams of automatic recognition of prototypical facial expressions associ- ated with discrete emotional states and of automatic detection of FACS AUs. For detailed surveys, we refer the reader to [41] and [66].

A. Emotion Recognition

Emotion recognition approaches can be divided into two groups based on the type of features used, either appearance- based features or geometry-based features. Appearance features describe the texture of the face caused by expression, such as wrinkles and furrows. Geometric features describe the shape of the face and its components such as the mouth or the eyebrows.

Within the appearance-based techniques, the theory of non- negative matrix factorization (NMF) has recently led to a num- ber of promising works. A technique called graph-preserving sparse NMF (GSNMF) was introduced by Zhi et al. [68] and applied to the problem of six-basic-emotion recognition. The GSNMF is an occlusion-robust dimensionality reduction tech- nique that can be employed either in a supervised or unsuper- vised manner. It transforms high-dimensional facial expression images into a locality-preserving subspace with sparse repre- sentation. On the Cohn–Kanade database, it attains a 94.3% recognition rate. On occluded images, it scored between 91.4% and 94%, depending on the area of the face that was occluded. Another recent NMF technique is nonlinear nonnegative component analysis, a novel method for data representation and classification proposed by Zafeiriou and Petrou [65]. Based on NMF and kernel theory, the method allows any positive definite kernel to be used and assures stable convergence of the optimization problem. On the Cohn–Kanade database, they attained an average 83.5% recognition rate over the six basic emotions.

Other appearance features that have been successfully em- ployed for emotion recognition are the local binary pattern (LBP) operator [49], [67], local Gabor binary patterns (LGBPs) [35], local phase quantization (LPQ) and histogram of oriented gradients [14], and Haar filters [31].

Most geometric-feature-based approaches use active appear- ance models (AAMs) or derivatives of this technique to track a dense set of facial points (typically 50–60). The locations of these points are then used to infer the shape of facial features such as the mouth or the eyebrows and thus to classify the facial expression. A recent example of an AAM-based technique is that of Asthana et al., who compare different AAM fitting algorithms and evaluate their performance on the Cohn–Kanade database, reporting a 93% classification accuracy [4]. Another example of a system that uses geometric features to detect emotions is that by Sebe et al. [47]. Piecewise Bézier volume deformation tracking was used after manually locating a number of facial points. They experimented with a large number of machine learning techniques.

Surprisingly, the best result was attained with a simple k-nearest neighbor technique that attained a 93% classification rate on the Cohn–Kanade database.

Sung and Kim [52] used AAMs to track facial points in 3-D videos. They introduce Stereo Active Appearance Models (STAAM), which improves the fitting and tracking of standard AAMs by using multiple cameras to model the 3-D shape and rigid motion parameters. A layered generalized discriminant analysis classifier, which is based on linear discriminant analy- sis, is then used to combine the 3-D shape and registered 2-D appearance. Unfortunately, although the approach appears to be promising, it was evaluated for only three expressions, and no results on a benchmark database (such as the Cohn–Kanade or MMI Facial Expression databases) were presented.

Current challenges in automatic discrete emotion recognition that remain to be addressed are dealing with out-of-plane head rotation, spontaneous expressions, and recognizing mixtures of emotions. Out-of-plane rotation and mixtures of emotions are two problems that are likely to coincide when moving to spontaneous real-world data. While some progress has been made in dealing with occlusions and tracking facial points in imagery of unseen subjects (e.g., [45] and [68]), these two elements remain a challenge as well.

B. AU Detection

AU detection approaches can be divided into a number of ways. Just as for emotion recognition, it is possible to divide them into systems that employ appearance-based features, geometric features, or both. Another way of dividing them is how they deal with the temporal dynamics of facial expressions: Frames in a video can either be treated as being independent of each other (this includes methods that target static images) or a sequence of frames that can be treated by a model that explicitly encompasses the expression's temporal dynamics.

A recently proposed class of appearance-based features that have been used extensively for face analysis is dense local appearance descriptors. First, a particular appearance descrip- tor is computed for every pixel in the face. To reduce the dimensionality of the problem and the sensitivity to alignment of the face, the descriptor responses are then summarized by histograms in predefined subregions of the face. For AUs, this approach was followed by Jiang et al., using LBP and LPQ [24]. Another successful appearance descriptor is the Gabor wavelet filter. Littlewort et al. [31] select the best set of Gabor filters using GentleBoost and train support vector machines (SVMs) to classify AU activation. Some measure of AU inten- sity is provided by evaluating for a test instance the distance to the separating hyperplane provided by the trained SVM. Haar- like features were used in an AdaBoost classifier by Whitehill and Omlin [62].

An example of an appearance-based approach that explicitly models a facial expression's temporal dynamics is that of Koelstra et al. [27]. In their work, they propose a method that detects AUs and their temporal phase onset, apex, and offset using free-form deformations and motion history images as appearance descriptors and hidden Markov models as machine learning technique.

In the geometric feature category, Valstar and Pantic [59] automatically detect 20 facial points and use a facial point tracker based on particle filtering with factorized likelihoods to track this sparse set of facial points. From the tracked points, both static and dynamic features are computed, such as the distances between pairs of points or the velocity of a facial point. With this approach, they are able to detect both AU activation and the temporal phase onset, apex, and offset.

Simon et al. [50] use both geometric and appearance-based features and include modeling of some of the temporal dynam- ics of AUs in a proposed method using segment-based SVMs. Facial features are first tracked using a person-specific AAM so that the face can be registered before extracting SIFT features. Principal component analysis (PCA) is applied to reduce the dimensionality of this descriptor. The proposed segment-based SVM method combines the output of static SVMs for multiple frames and uses structured-output learning to learn the begin- ning and end time of each AU. The system was evaluated for eight AUs on the M3 database (previously called RU-FACS), attaining an average of 83.75% area under the ROC curve.

When facing real-world data, researchers have to face prob- lems such as very large data sizes or low AU frequencies of occurrence. In their work, Zhu et al. focus on the automatic selection of an optimal training set using bidirectional boot- strapping from a data set with exactly such properties [69]. The features used are identical to those described and used by Simon et al. [50]. The proposed dynamic cascades with bidi- rectional bootstrapping apply GentleBoost to perform feature selection and training instance selection in a unified framework. On the M3 database, the system attained an average 79.5% are under the ROC curve for 13 AUs. For an overview of more recent work by researchers at CMU, see [29].

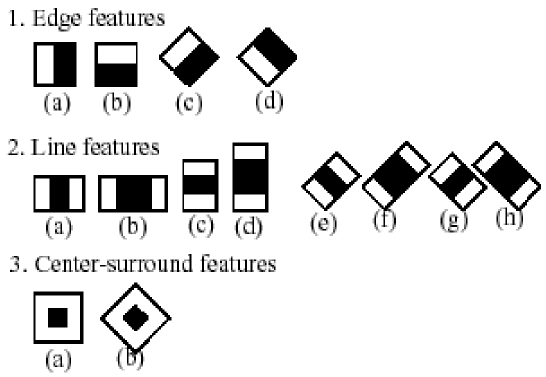
Current challenges in AU detection include handling of out- of-plane head rotations and occlusion, two conditions that occur frequently in real-world data of spontaneous expressions. Because AUs are more localized in the face than expressions of discrete emotions, the problem of occlusion is much bigger for AUs than for emotions. Likewise, out-of-plane head rotations can cause self-occlusions of parts of the face that display some AUs, making the problems caused by out-of-plane head poses harder than it is for emotions. Another issue of moving to data of spontaneous expressions is that the co-occurrences between AUs become much harder to model, compared to the limited number of co-occurrence patterns in databases of posed expressions such as the Cohn–Kanade database.

Aside from AU detection, the detection of an AU's tempo- ral phase transitions (onset, apex, and offset), as well as its intensity, is a partially unsolved problem. Being able to predict these variables would allow researchers to detect more complex higher level behavior such as deception, cognitive states like (dis)agreement and puzzlement, or psychological states like pain [11], [13]

III HAAR like features

HAAR CASCADE CLASSIFIERS

The core basis for Haar classifier object detection is the Haar-like features. These features, rather than using the intensity values of a pixel, use the change in contrast values between adjacent rectangular groups of pixels. The contrast variances between the pixel groups are used to determine relative light and dark areas. Two or three adjacent groups with a relative contrast variance form a Haar-like feature. Haar-like features, as shown in Figure 1 are used to detect an image [8]. Haar features can easily be scaled by increasing or decreasing the size of the pixel group being examined. This allows features to be used to detect objects of various sizes.



Integral Image

The simple rectangular features of an image are calculated using an intermediate representation of an image, called the integral image. The integral image is an array containing the sums of the pixels' intensity values located directly to the left of a pixel and directly above the pixel at location (x,y) inclusive. So if A[x,y] is the original image and AI[x,y] is the integral image then the integral image is computed as shown in equation 1.

$$AI[x,y] = \sum_{x' \le x, y' \le y} A(x',y')$$

Lienhart and Maydt, require another intermediate representation called the rotated integral image or rotated sum auxiliary image [5]. The rotated integral image is calculated by finding the sum of the pixels' intensity values that are located at a forty five degree angle to the left and above for the x value and below for the y value. So if A[x,y] is the original image and AR[x,y] is the rotated integral image then the integral image is computed as shown in equation 2.

$$AR[x,y] = \sum_{x' \le x, x' \le x - |y - y'|} A(x',y')$$

It only takes two passes to compute both integral image arrays, one for each array. Using the appropriate integral image and taking the difference between six to eight array elements forming two or three connected rectangles, a feature of any scale can be computed. Thus calculating a feature is extremely fast and efficient. It also means calculating features of various sizes requires the same effort as a feature of only two or three pixels. The detection of various sizes of the same object requires the same amount of effort and time as objects of similar sizes since scaling requires no additional effort.

Classifiers Cascaded

Although calculating a feature is extremely efficient and fast, calculating all 180,000 features contained within a 24 × 24 sub-image is impractical [Viola 2001, Wilson 2005]. Fortunately, only a tiny fraction of those features are needed to determine if a sub-image potentially contains the desired object. In order to eliminate as many sub-images as possible, only a few of the features that define an object are used when analyzing sub-images. The goal is to eliminate a substantial amount, around 50%, of the sub-images that do not contain the object. This process continues, increasing the number of features used to analyze the sub-image at each stage. The cascading of the classifiers allows only the sub-images with the highest probability to be analyzed for all Haar-features that distinguish an object. It also allows one to vary the accuracy of a classifier. One can increase both the false alarm rate and positive hit rate by decreasing the number of stages. The inverse of this is also true. Viola and Jones were able to achieve a 95% accuracy rate for the detection of a human face using only 200 simple features. Using a 2 GHz computer, a Haar classifier cascade could detect human faces at a rate of at least five frames per second.

Training Classifiers For Facial Features

Detecting human facial features, such as the mouth, eyes, and nose require that Haar classifier cascades first be trained. In order to train the classifiers, this gentle AdaBoost algorithm and Haar feature algorithms must be implemented. Fortunately, Intel developed an open source library devoted to easing the implementation of computer vision related programs called Open Computer Vision Library (OpenCV). The OpenCV library is designed to be used in conjunction with applications that pertain to the field of HCI, robotics, biometrics, image processing, and other areas where visualization is important and includes an implementation of Haar classifier detection and training. To train the classifiers, two set of images are needed. One set contains an image or scene that does not contain the object, in this case a facial feature, which is going to be detected. This set of images is referred to as the negative images. The other set of images, the positive images, contain one or more instances of the object.

The location of the objects within the positive images is specified by: image name, the upper left pixel and the height, and width of the object [1]. For training facial features 5,000 negative images with at least a mega-pixel resolution were used for training. These images consisted of everyday objects, like paperclips, and of natural scenery, like photographs of forests and mountains.

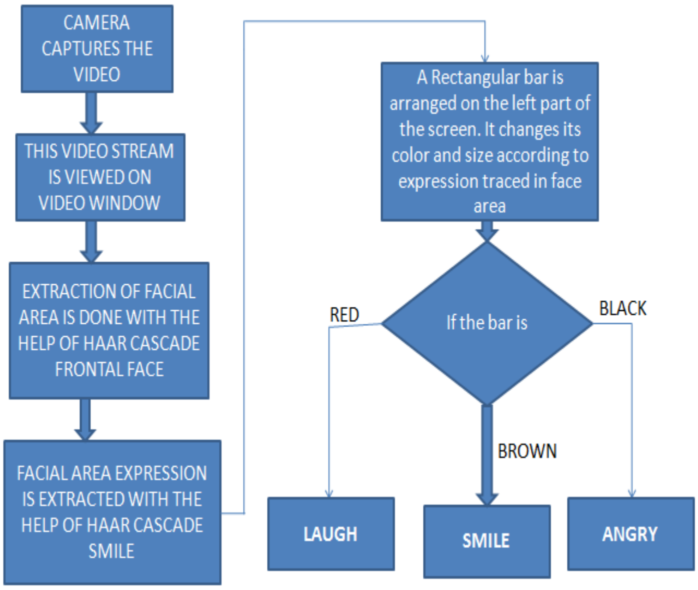
In order to produce the most robust facial feature detection possible, the original positive set of images needs to be representative of the variance between different people, including, race, gender, and age. A good source for these images is National Institute of Standards and Technology's (NIST) Facial Recognition Technology (FERET) database. This database contains over 10,000 images of over 1,000 people under different lighting conditions, poses, and angles. In training each facial feature, 1,500 images were used. These images were taken at angles ranging from zero to forty five degrees from a frontal view. This provides the needed variance required to allow detection if the head is turned slightly

Three separate classifiers were trained, one for the eyes, one for the nose, and one for the mouth. Once the classifiers were trained, they were used to detect the facial features within another set of images from the FERET database. The accuracy of the classifier was then computed as shown in Table 1. With the exception of the mouth classifier, the classifiers have a high rate of detection. However, as implied by [3], the false positive rate is also quite high.

Facial Feature	Positive Hit Rate	Negative Hit Rate
Eyes	93%	23%
Nose	100%	29%
Mouth	67%	28%

Regionalized Detection

Since it is not possible to reduce the false positive rate of the classifier without reducing the positive hit rate, a method besides modifying the classifier training attribute is needed to increase accuracy. The method proposed to is to limit the region of the image that is analyzed for the facial features. By reducing the area analyzed, accuracy will increase since less area exists to produce false positives. It also increases efficiency since fewer features need to be computed and the area of the integral images is smaller. In order to regionalize the image, one must first determine the likely area where a facial feature might exist. The simplest method is to perform facial detection on the image first. The area containing the face will also contain facial features. However, the facial feature cascades often detect other facial features as illustrated in Figure 4. The best method to eliminate extra feature detection is to further regionalize the area for facial feature detection. It can be assumed that the eyes will be located near the top of the head, the nose will be located in the center area and the mouth will be located near the bottom. The upper 5/8 of the face is analyzed for the eyes. This area eliminates all other facial features while still allowing a wide variance in the tilt angle. The center of the face, an area that is 5/8 by 5/8 of the face, was used to for detection of the nose. This area eliminates all but the upper lip of the mouth and lower eyelid. The lower half of the facial image was used to detect the mouth. Since the facial detector used sometimes eliminates the lower lip the facial image was extended by an eighth for mouth detection only.



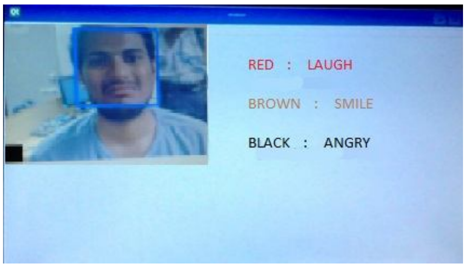
IV Results

The challenge is divided into two sub-challenges. The goal of the AU detection sub-challenge is to identify in every frame of a video whether an AU was present or not (i.e., it is a multiple-label binary classification problem at frame level). The goal of the emotion recognition sub-challenge is to recognize which emotion was depicted in that video. The first step in facial feature detection is detecting the face. This requires analyzing the entire image. The second step is using the isolated face(s) to detect each feature. The result is shown in Figure 5. Since each the portion of the image used to detect a feature is much smaller than that of the whole image, detection of all three facial features takes less time on average than detecting the face itself. Using a 1.2GHz AMD processor to analyze a 320 by 240 image, a frame rate of 3 frames per second was achieved. Since a frame rate of 5 frames per second was achieved in facial detection only by using a much faster processor, regionalization provides a tremendous increase in efficiency in facial feature detection.

Regionalization also greatly increased the accuracy of the detection. All false positives were eliminated, giving a detection rate of around 95% for the eyes and nose. The mouth detection has a lower rate due to the minimum size required for detection. By changing the height and width parameter to more accurately represent the dimensions of the mouth and retraining the classifier the accuracy should increase the accuracy to that of the other feature.

O/P: Laugh

O/P: Smile



O/P: Angry

ACKNOWLEDGEMENT:

I would like to thank my HOD of ECE department Associate Professor Mrs. Latha, Internal project guide and Associate Professor Mrs. Nirmala Devi and Project Coordinator Assistant Professor Ms. V. Sudha for their great support.

REFERENCES

1. T. Ahonen, A. Hadid, and M. Pietikäinene, “Face description with local binary patterns: Application to face recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

2. N. Ambady and R. Rosenthal, “Thin slices of expressive behavior as pre-dictors of interpersonal consequences: A meta-analysis,” Psychol. Bull., vol. 11, no. 2, pp. 256–274, Mar. 1992.

3. A. B. Ashraf, S. Lucey, J. F. Cohn, T. Chen, Z. Ambadar, K. M. Prkachin, and P. E. Solomon, “The painful face—Pain expression recognition using active appearance models,” Image Vis. Comput., vol. 27, no. 12, pp. 1788–1796, Nov. 2009.

4. A. Asthana, J. Saragih, M. Wagner, and R. Goecke, “Evaulating aam fitting methods for facial expression recognition,” in Proc. Int. Conf. Affective Comput. Intell. Interact., 2009, pp. 1–8.

5. T. Baltrusaitis, D. McDuff, N. Banda, M. Mahmoud, R. El Kaliouby, P. Robinson, and R. Picard, “Real-time inference of mental states from facial expressions and upper body gestures,” in Proc. IEEE Int. Conf. Autom. Face Gesture Anal., 2011, pp. 909–914.

6. T. Bänziger and K. R. Scherer, “Introducing the geneva multimodal emo- tion portrayal (gemep) corpus,” in Blueprint for Affective Computing: A Sourcebook, K. R. Scherer, T. Bänziger, and E. B. Roesch, Eds. Oxford, U.K.: Oxford Univ. Press, 2010, ser. Series in Affective Science, ch. 6.1, pp. 271–294.

7. M. Bartlett, G. Littlewort-Ford, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, “Fully automatic facial action recognition in spontaneous behaviour,” in Proc. IEEE Int. Conf. Autom. Face Gesture Recog., 2006, pp. 223–230.

8. S. W. Chew, P. Lucey, S. Lucey, J. Saragih, J. Cohn, and S. Sridharan, “Person-independent facial expression detection using constrained local models,” in Proc. IEEE Int. Conf. Autom. Face Gesture Anal., 2011, pp. 915–920.

9. J. Cohn and P. Ekman, “Measuring facial action by manual coding, facial emg, and automatic facial image analysis,” in Handbook of Nonverbal Behavior Research Methods in the Affective Sciences, R. R. K. S. J. A. Harrigan, Ed. New York: Oxford Univ. Press, 2005, pp. 9–64.

10. J. F. Cohn, "Foundations of human computing: Facial expression and emotion," in Proc. ACM Int.l Conf. Multimodal Interf., 2006, vol. 1, pp. 610–616.

11. D. Cunningham, M. Kleiner, C. Wallraven, and H. Bülthoff, "The compo- nents of conversational facial expressions," in Proc. ACM Int. Symp. Appl. Perception Graph. Vis., 2004, pp. 143–149.

12. M. Dahmane and J. Meunier, "Emotion recognition using dynamic grid- based hog features," in Proc. IEEE Int. Conf. Autom. Face Gesture Anal., 2011, pp. 884–888.

13. A. C. de C. Williams, "Facial expression of pain: An evolutionary ac- count," Behav. Brain Sci., vol. 25, no. 4, pp. 439–488, Aug. 2002.

14. A. Dhall, A. Asthana, R. Goecke, and T. Gedeon, "Emotion recognition using phog and lpq features," in Proc. IEEE Int. Conf. Autom. Face Gesture Anal., 2011, pp. 878–883.

15. P. Ekman, Face of Man: Universal Expression in a New Guinea Village. New York: Garland, 1982.

16. P. Ekman, "An argument for basic emotions," Cog. Emotion, vol. 6, no. 3/4, pp. 169–200, May 1992.

17. P. Ekman and W. Friesen, "The repertoire of nonverbal behavioral categories—Origins, usage and coding," Semiotica, vol. 1, no. 1, pp. 49–98, 1969.

18. P. Ekman, W. V. Friesen, and J. C. Hager, FACS Manual. Salt Lake City, UT: A Human Face, May 2002.

19. R. El Kaliouby and P. Robinson, "Generalization of a vision-based computational model of mind-reading," in Proc. 1st Int. Conf. Affective Comput. Intell. Interact., 2005, pp. 582–589.

20. B. Fasel and J. Luettin, "Automatic facial expression analysis: A survey," Pattern Recognit., vol. 36, no. 1, pp. 259–275, Jan. 2003.

21. T. Gehrig and H. Ekenel, "Facial action unit detection using kernel partial least squares," in Proc. IEEE ICCV , 2011, pp. 2092–2099.

22. H. Gunes and M. Pantic, "Automatic, dimensional and continuous emo- tion recognition," Int. J. Synthetic Emotions, vol. 1, no. 1, pp. 68–99, Jan./Jun. 2010.

23. H. Gunes, B. Schuller, M. Pantic, and R. Cowie, "Emotion representation, analysis and synthesis in continuous space: A survey," in Proc. IEEE Int. Conf. Autom. Face Gesture Recog., 2011, pp. 827–834.

24. B. Jiang, M. Valstar, and M. Pantic, "Action unit detection using sparse appearance descriptors in space-time video volumes," in Proc. IEEE Int. Conf. Autom. Face Gesture Recog., 2011, pp. 314–321.

25. T. Kanade, "Picture processing system by computer complex and recognition of human faces," Ph.D. dissertation, Tokyo Univ., Tokyo, Japan, 1973.

26. T. Kanade, J. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in Proc. 4th IEEE Int. Conf. Autom. Face Gesture Recog., 2000, pp. 46–53.

27. S. Koelstra, M. Pantic, and I. Patras, "A dynamic texture based ap- proach to recognition of facial actions and their temporal models," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 11, pp. 1940–1954, Nov. 2010.

28. I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," IEEE Trans. Image Process., vol. 16, no. 1, pp. 172–187, Jan. 2007.

29. F. D. la Torre and J. F. Cohn, "Facial expression analysis," in Guide to Visual Analysis of Humans: Looking at People, V. K. T. B. Moeslund, A. Hilton, and L. Sigal, Eds. New York: Springer-Verlag, 2011.

30. G. Littlewort, J. Whitehill, T. Wu, N. Butko, P. Ruvolo, J. Movellan, and M. Bartlett, "The motion in emotion-a cert based approach to the fera emotion challenge," in Proc. IEEE Int. Conf. Autom. Face Gesture Anal., 2011, pp. 897–902.

31. G. Littlewort, J. Whitehill, T. Wu, I. Fasel, M. Frank, J. Movellan, and M. Bartlett, "The computer expression recognition toolbox (cert)," in Proc. IEEE Int. Conf. Autom. Face Gesture Recog., 2011, pp. 298–305.

32. C. Liu, J. Yuen, and A. Torralba, "Sift flow: Dense correspondence across scenes and its applications," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 5, pp. 978–994, May 2010.

33. M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in Proc. 3rd IEEE Int. Conf. Autom. Face Gesture Recog., 1998, pp. 200–205.

34. H. Meng, B. Romera-Paredes, and N. Berthouze, "Emotion recognition by two view SVM_2k classifier on dynamic facial expression features," in Proc. IEEE Int. Conf. Autom. Face Gesture Anal., 2011, pp. 854–859.

35. S. Moore and R. Bowden, "Local binary patterns for multi-view facial expression recognition," Comput. Vis. Image Understand., vol. 115, no. 4, pp. 541–558, Apr. 2011.

36. T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on featured distribution," Pattern Recognit., vol. 29, no. 1, pp. 51–59, Jan. 1996.

37. T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 7, pp. 971–987, Jul. 2002.

38. M. Pantic and M. Bartlett, "Machine analysis of facial expressions," in Face Recognition, K. Delac and M. Grgic, Eds. Vienna, Austria: I-Tech Educ. Publ., 2007, pp. 377–416.

39. M. Pantic, A. Nijholt, A. Pentland, and T. Huang, "Human-centred intelli- gent human-computer interaction (hci2): How far are we from attaining it?" Int. J. Auton. Adapt. Commun. Syst., vol. 1, no. 2, pp. 168–187, Aug. 2008.

40. M. Pantic and L. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 12, pp. 1424–1445, Dec. 2000.

41. M. Pantic and L. J. M. Rothkrantz, "Toward an affect-sensitive multimodal human-computer interaction," Proc. IEEE, vol. 91, no. 9, pp. 1370–1390, Sep. 2003.

42. M. Pantic, M. F. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in Proc. Int. Conf. Multimedia Expo., 2005, pp. 317–321.

43. A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet, "Simplemkl," J. Mach. Learn. Res., vol. 9, pp. 2491–2521, 2008.

44. A. Samal and P. A. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: A survey," Pattern Recognit., vol. 25, no. 1, pp. 65–77, Jan. 1992.

45. J. Saragih, S. Lucey, and J. Cohn, "Face alignment through subspace constrained mean-shifts," in Proc. Int. Conf. Comput. Vis., 2009, pp. 1034–1041.

46. B. Schuller, M. Valstar, F. Eyben, G. McKeown, R. Cowie, and M. Pantic, “Avec 2011 the first international audio/visual emotion challenge,” in Proc. ACII, 2011, pp. 415–424.

47. N. Sebe, M. Lew, I. Cohen, T. Gevers, and T. Huang, “Authentic facial expression analysis,” Image Vis. Comput., vol. 25, no. 12, pp. 1856–1863, Dec. 2007.

48. T. Senechal, V. Rapp, L. Prevost, H. Salam, R. Seguier, and K. Bailly, “Combining lgbp histograms with aam coefficients in the ulti-kernel svm framework to detect facial action units,” in Proc. IEEE Int. Conf. Autom. Face Gesture Anal., 2011, pp. 860–865.

49. C. Shan, S. Gong, and P. McOwan, “Facial expression recognition based on local binary patterns: A comprehensive study,” Image Vis. Comput. J., vol. 27, no. 6, pp. 803–816, 2009.

50. T. Simon, M. H. Nguyen, F. De La Torre, and J. Cohn, “Action unit detection with segment-based SVMS,” in Proc. IEEE Conf. CVPR, Jun. 2010, pp. 2737–2744.

51. R. Srivastava, S. Roy, S. Yan, and T. Sim, “Accumulated motion images for facial expression recognition in videos,” in Proc. IEEE Int. Conf. Autom. Face Gesture Anal., 2011, pp. 903–908.

52. J. Sung and D. Kim, “Real-time facial expression recognition using staam and layered gda classifier,” Image Vis. Comput., vol. 27, no. 9, pp. 1313–1325, Aug. 2009.

53. U. Tariq, K.-H. Lin, Z. Li, X. Zhou, Z. Wang, V. Le, T. Huang, X. Lv, and T. Han, “Emotion recognition from an ensemble of features,” in Proc. IEEE Int. Conf. Autom. Face Gesture Anal., 2011, pp. 872–877.

54. Y. Tian, T. Kanade, and J. Cohn, “Recognizing action units for facial expression analysis,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 2, pp. 97–115, Feb. 2001.

55. Y. L. Tian, T. Kanade, and J. F. Cohn, Handbook of Face Recognition. New York: Springer-Verlag, 2005.

56. Y. Tong, J. Chen, and Q. Ji, “A unified probabilistic framework for spontaneous facial action modeling and understanding,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 2, pp. 258–273, Dec. 2010.

57. M. Valstar andM. Pantic, “Combined support vector machines and hidden Markov models for modeling facial action temporal dynamics,” in Proc. ICCV-HCI, 2007, pp. 118–127.

58. M. Valstar and M. Pantic, "Induced disgust, happiness and surprise: An addition to the mmi facial expression database," in Proc. 3rd Int. Workshop EMOTION, 2010, pp. 65–70, (Satellite of LREC): Corpora for Research on Emotion and Affect.

59. M. Valstar and M. Pantic, "Fully automatic recognition of the temporal phases of facial actions," IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 42, no. 1, pp. 28–43, Feb. 2012.

60. M. F. Valstar and M. Pantic, "Biologically vs. logic inspired encoding of facial actions and emotions in video," in Proc. IEEE Int. Conf. Multimedia Expo., 2006, pp. 325–328.

61. P. Viola and M. Jones, "Robust real-time object detection," Int. J. Comput. Vis., vol. 57, no. 2, pp. 137–154, 2002.

62. J. Whitehill and C. Omlin, "Haar features for facs au recognition," in Proc. Int'l. Conf. Automatic Face and Gesture Recognition, 2006, pp. 97–101.

63. T. Wu, N. Butko, P. Ruvolo, J. Whitehill, M. Bartlett, and J. Movellan, "Action unit recognition transfer across datasets," in Proc. IEEE Int. Conf. Autom. Face Gesture Anal., 2011, pp. 889–896.

64. S. Yang and B. Bhanu, "Facial expression recognition using emotion avatar image," in Proc. IEEE Int. Conf. Autom. Face Gesture Anal., 2011, pp. 866–871.

65. S. Zafeiriou and M. Petrou, "Nonlinear non-negative component analysis algorithms," IEEE Trans. Image Process., vol. 19, no. 4, pp. 1050–1066, Apr. 2010.

66. Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 1, pp. 39–58, Jan. 2009.

67. G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary pattern with an application to facial expressions," IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 6, pp. 915–928, Jun. 2007.

68. R. Zhi, M. Flierl, Q. Ruan, and W. Kleijn, "Graph-preserving sparse nonnegative matrix factorization with application to facial expression recognition," IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 41, no. 1, pp. 38–52, Feb. 2011.

69. Y. Zhu, F. De la Torre, J. Cohn, and Y. Zhang, “Dynamic cascades with bidirectional bootstrapping for action unit detection in spontaneous facial behavior,” IEEE Trans. Affective Comput., vol. 2, no. 2, pp. 79–91, Apr.–Jun. 2011.