2017 2nd International Conference on Telecommunication and Networks (TEL-NET 2017)

An Efficient Algorithm for Facial Landmark Detection using Haar-Like Features coupled with Corner Detection following Anthropometric Constraints

Naveen Kumar Gondhi
DoCSE, SMVDU, Katra

Navleen Kour
DoCSE, SMVDU, Katra

Sidra Effendi
DoCSE, SMVDU, Katra

Keshav Kaushik
DoCSE, SMVDU, Katra

Abstract— **Our internal behavior, various situations we are going through, our feelings, all are reflected by face. Facial landmarking is an important step in facial emotions recognition. Landmarking is the process of detection, extraction and localization of fiducial points of face such as silent points on corners of eyes, nose tip, etc. In this paper, we have presented a new automatic facial landmarks detection system which does not require any complicated pre-processing and simple to implement. Our landmarking technique combines Viola-Jones detection algorithm for feature detection with Harris corner detection and then coarse to fine strategy is implemented using an efficient algorithm. Using the Haar like features reduces the cost of brute force search, also provides advantage of speed. Additional selection of sub-regions is also exploited using anthropometric constraints, to limit the search region. This further reduces false detection rate and improves accuracy significantly. A sub-algorithm named Iterative best fit algorithm is used find a land mark exploiting its commonality and geometric configuration and can be used in other contexts as well. The method is tested on JAFEE database and accuracy rate is 93%.**

Keywords—component; - Facial Landmark Detection; Viola-Jones algorithm; Harris Corner Detection; JAFEE Database.

I. INTRODUCTION

In this modern era of advanced research and technology, detection and extraction of facial landmarks are of utmost importance that has influenced various parts of our lives such as recognition of face and facial expressions, gesture understanding, face expression analysis [1], animation, face tracking [1,2], face animation [1,3], registration [1,4], video tracing[1]. Face landmarks are treated as the anchor points. The most commonly used landmarks are corners of eyes, nose, lips, mouth, eyebrow and various other facial parts. Landmarks such as eye corners or nose tip are known to be little affected by facial expressions and are easy to locate and are in fact referred to as fiducial points or primary landmarks [5]. The facial landmarks which are found with the help of the fiducial points are knows as ancillary points. Important applications of landmarking can be found in facial identity in digital photos, sign language interpretation, nursing homes [6], offices [7], etc.

Approaches present today for facial landmark detection can be grouped in two main classes: local and global methods [8]. The global methods [8] are more capable of detecting more landmarks than the local ones, which can mostly detect landmarks quickly. Almost all global methods use either ASM (Active Shape Models) [9] or AAM method (Active Appearance Models) [10]. In the ASM technique, the algorithm make searches in order to obtain the best match using a shape model while as in AAM, the main goal of the algorithm [11] is to obtain the best match with a combined model using texture and shape. In local methods [12], the algorithms are used to detect landmarks such as we can take the example of the corner of the eyes or the tip of the nose [13], without using information from other parts of the face. There are also some situations in which a combination of global and local method is used where classifiers are trained for different landmark with the Viola & Jones object detection approach. Over the last few years, there have been numerous face recognition techniques that has been developed and used for the purpose [14] of facial expressions recognition. The whole literature survey is in [15].

In this paper, we propose [16] a new facial landmark detection system to detect landmarks in human faces. The proposed system is a combination of local and global methods to get a mix of robustness and speed without any complicated pre-processing. It is a coarse to fine strategy which uses haar-like features and corner detection along with geometric constraints to get facial landmarks. The technique is simple and easy in implementation. The remainder of the paper is structured as follows: Section II describes the proposed method in detail with all the steps involved, then followed by Section III, in which we have presented the experiments and the results obtained. The last section describes some ideas about our future work and in this section we conclude our work properly.

II. PROPOSED WORK

The proposed method involves the system to detect a set of landmarks in frontal faces. This system is comprised of four parts. In the first part is the pre-processing step and second part has the methodology of detecting features, that is, eyes and mouth in the face. The third part detects the corner in the face and fourth part is the core of the system, where the results of two previous stages is combined and landmark is detected using proposed efficient algorithm called 'Landmarks' and the iterative-best –fit algorithm is also explained which chooses the landmark from a no. of possibilities . A detailed block diagram of the proposed method is depicted by figure 1

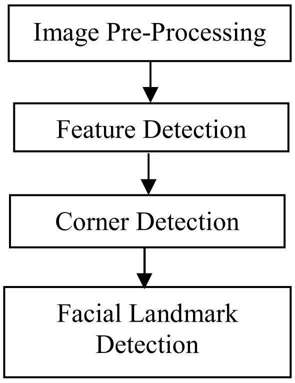


Fig.1. Steps of Facial Landmark Detection

A. Image Pre-Processing

Face images from different databases have diverse resolutions, backgrounds and are captured under varying illumination. Preprocessing module is used to make the images more comparable. In our system, complex pre-processing module is not required and consists of following components: image resizing, face detection and image cropping. The image is resized to get a square image. From this image face area is detected using the Viola-Jones method [17], which requires the face image to be roughly frontal (i.e., rotations up to are acceptable) [18], based on the Haar-like features and AdaBoost learning algorithm. We assumed that there is only one face per image. The Viola and Jones method is an object detection algorithm providing competitive object detection rate [17]. It is primarily designed for the problem of face detection. The features used by Viola and Jones are derived from pixels selected from rectangular areas imposed over the picture and show high sensitivity to the vertical and horizontal lines. AdaBoost is an adaptive learning algorithm that can be used in conjunction with many other learning algorithms to improve their performance. AdaBoost is adaptive in the sense that subsequent classier built iteratively are made to fix instances misclassified by previous classier. At each iteration a distribution of weights is updated such that, the weights of each incorrectly classified example are increased (or alternatively, the weights of each correctly classified example are decreased), so that the new classifier focuses more on those examples [19]. The last part of image pre- processing module is cropping the face images to get the area enclosed by rectangle of the Viola-Jones object detector for further processing.

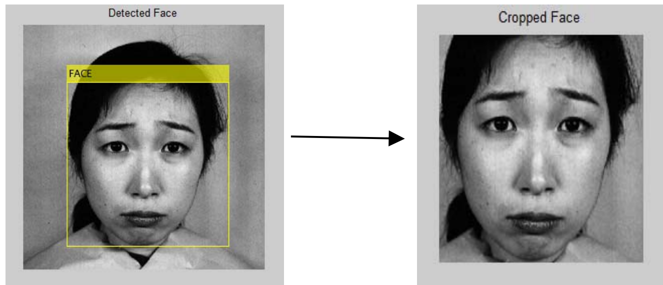
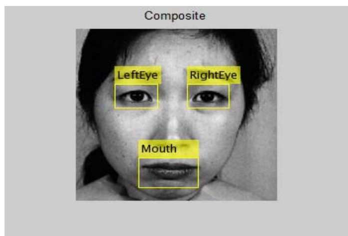


Fig.2: Cropping of face detected by Viola-Jones algorithm

B. Feature Detection

In this step we want to recognize the positions of eyes and the mouth in a given face image. This is the second step to reduce the region of interest first being the cropping of face image. . The Viola-Jones algorithm is used for feature detection. The detected face attained from pre-processing is first divided into three parts such that one part contains the left eye, another contains the right eye and the third part contains the mouth. This is done to overcome the limitation of cascade object detector based on Viola Jones algorithm, which sometimes detects mouth as an eye or vice versa. This method does a multi-scale search and chooses the facial landmark candidates through thresholding. A composite image of the three parts is achieved at the end with detected features. There is the possibility of multiple candidates (multi-size and/or different position) for which we choose the candidate with the largest size. The reason is that it is observed that smaller-size candidates tend to be less reliable, like a small eye candidate localized on the eyebrow [20].

(a) (b)



(c)

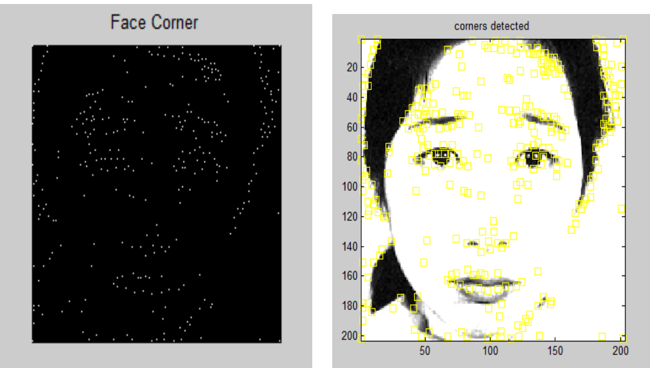
Fig.3. (a) Division of Face (b) feature detection in separate parts (c) Combining the parts of face after detection

C. Corner Detection

Corner detection is an approach used to extract certain kinds of features and infer the contents of an image. Some uses canny edge detector to quantify amount and orientation of furrows [22] but we used corner detection. Corner detection is used in motion detection, image registration, video tracking and object recognition [23]. Corner is the intersection of two edged. The significance of this step is that corner detection gives us plausible landmarks as many facial landmarks are corners and by having all corners in an image, landmarks can be chosen from them by further computation. Harris Corner Detection algorithm, which is used by us, detects the corners in any given image and is an improvement over Moravec's corner detector by considering the differential of the corner score with respect to direction directly, instead of using shifted patches.It is applied to the entire pre-processed image and calculates the points of change in all directions. Change of intensity for the shift [*u, v*]: [24]

$$E(u,v)=\sum_{x,y} w(x,y)\left[I(x+u,y+v)-I(x,y)\right]^2$$

The function w is the window function, the function I(x + u, y + v) is the shifted intensity and the last is the original intensity. The window function is Gaussian which reduces the noise. The corner threshold value for my system has been chosen as 600.



(a) (b)

Fig.4. (a) Corners detected (b) Detected Corners mapped to face

D. Facial Landmark Detection

In this section the main contributions of the paper are presented [25]. The results of the feature detection and corner detection are combined to give a resultant image in which the corners detected in the eyes and mouth region are enclosed by a rectangle.

Fig.5. Combination of feature and corner detection

From the corners lying inside the rectangle we have to choose the facial Landmarks. We detect a total of 12 facial landmarks out of which half are primary (left and right landmarks) and (top and bottom landmarks) half are secondary. The procedure to find landmarks is repeated for all rectangles demarcating features in the image (w is the width and h is height of the rectangle) .The values x and y mentioned in the procedure are the points, to which the landmark we want, lies the closest. These values have been determined after trying different values and picking the ones which give best result. The steps of the procedure are as follows:

1. *Cropping the image:* The portion of the image marked by rectangular boundary is cropped and the colored boundary is removed from all four sides of the cropped image to avoid its interference during further computation.

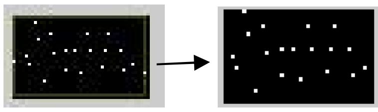
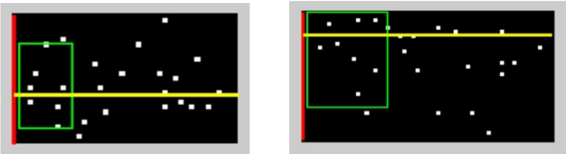


Fig.6. Removing boundary from detected feature

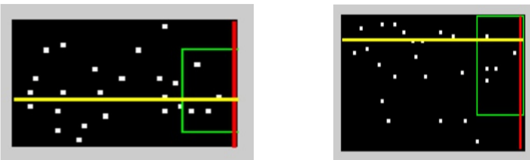
2. *Locating Left landmark for eyes and mouth:* The left primary facial landmark for the eyes is detected by first cropping the region of interest from the detected feature. The region of interest for eyes has width 1 to w/4 since by knowledge of geometry the left landmark will lie to the left portion of the rectangle and height is h/4 to 7*h/8 as landmark will lie around middle portion of the rectangle. The ROI for left facial landmark in the mouth is height 1 to 3*h/4 because the mouth corner is observed to lie in the upper left region and width is 1 to w/3.
   Then the algorithm 1 is used to choose a single point from the ROI by passing values a = 1 for both eyes and mouth and b = 5* h/8 for eyes and h/5 for mouth to it.



(a) (b)

Fig.7. (a) eye feature (b) mouth feature (The green rectangular box represents the ROI and red line represents the value a while as b is represented by the yellow line.)

3. *Locating right landmark for eyes and mouth:* For right primary facial landmark the ROI has width 3*w/4 to w since by knowledge the right landmark will lie to the extreme right and height h/4 to 7*h/8 as landmark will lie around middle portion of the rectangle demarking eye. The ROI for the right facial landmark of the mouth is height 1 to 3*h/4 and width is 3*w/4 to w. Then the algorithm 1 is used to choose a single point from the ROI with values a = w for both eyes and mouth and a = 5* h/8 for eyes and h/5 for mouth.

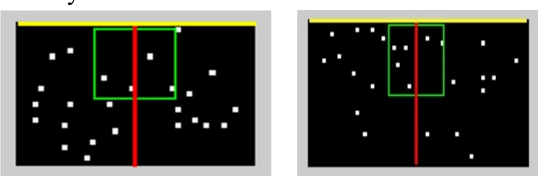


(a) (b)

Fig.8. (a) eye feature (b) mouth feature (The green rectangular box represents the ROI and red line represents the value a while as b is represented by the yellow line.)

4. *Locating top landmark for eyes and mouth:* For the top facial landmark for the eyes sub-region is w/3 to 2*w/3 is selected which is the middle region and 1 to h/2 is selected which is the top region of the rectangle demarking eye. The ROI for the mouth is 1 to h/2 and width is 3*w/8 to 5*w/8.

   Then the algorithm 1 is used to choose a single point from the ROI with values a = w/2 for and b = 1 for both eyes and mouth.

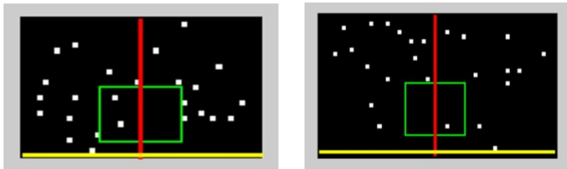


(a) (b)

Fig.9. (a) eye feature (b) mouth feature (The green rectangular box represents the ROI and red line represents the value a while as b is represented by the yellow line.)

5. *Locating bottom landmark for eyes and mouth:* For the bottom facial landmark for the eyes sub-region w/3 to 2*w/3 is selected which is the middle region and h/2 to 7*h/8 is selected which is the bottom most region of the rectangle demarking eye. The ROI for the mouth is h/2 to 6*h/7 and width is 3*w/8 to 5*w/8.

   Then the algorithm 1 is applied to choose a single point from the ROI with a = 1 for mouth and h for eyes and a = w/2 for both eyes and mouth.



(a) (b)

Fig.10. (a) eye feature (b) mouth feature (The green rectangular box represents the ROI and red line represents the value a while as b is represented by the yellow line.)

- ALGORITHM 1:-

1. After the corners have been detected, the binary image produced after corner detection is overlapped by the eye and the mouth rectangles.
2. The area of the image enclosed by the three rectangles is then cropped one by one.
3. From the cropped rectangle, first the colored boundary of the classifier is removed to allow easy extraction of corner points.
4. The left landmark is chosen on the basis of its distance from the left boundary and from the line dividing images into two equal halves horizontally.
5. The same is done to get right fiducial point; the only difference is that the distance is calculated from the right boundary.
6. In case two points are close to each- other, then the sum of their vertical and horizontal distances is compared. The one with the least value is selected.
7. Taking the center of the line joining the left and right landmarks and searching for the extreme corners both top and bottom, we get two non-fiducial points.
8. These points are then mapped again on the original face image.
9. The process is repeated for all the rectangles and the corresponding landmarks are extracted respectively.

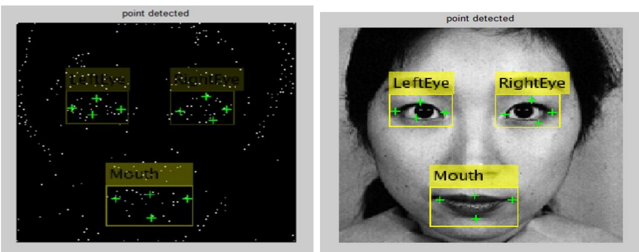
This Iterative best fit algorithm aims to find the required landmark from a no. of possible landmark. Each possible landmark is checked and the one chosen fits the requirements best. To find a landmark from the sub region, the points with minimum value of x and y coordinate are chosen. For any one point to be chosen as the desired landmark it must have minimum value of both x and y thus being closest to the required position. If one possible landmark has minimum x value and one has minimum y value then one of the two is chosen by adding values of x and y for the possible landmarks then the one which has the least overall sum is chosen and its coordinates are returned.

III. EXPERIMENTS AND RESULTS

There are various databases that were used by researchers such as FERAT database, have been used by many researches [26, 27, 28]. Our proposed algorithm for the detection of facial landmarks is tested using the JAFEE Database. The Jaffee database is a Japanese Female Facial Expression database [25] containing 213 images of 7 facial expressions which are posed by 10 Japanese female models giving different poses. Different expressions shown by different models are shown by the following figure

Fig.11. JAFEE Database

To determine that the facial landmarks every image is is checked and each landmark is given one of the three labels- *right, bearable* or *wrong*. Right is given if landmark is detected at the correct position .



(a) (b)

Fig.12. (a)Landmarks detected (b)Detected Landmarks mapped to the face.

TABLE1. ANALYSIS OF DIFFERENT LANDMARKS

S.N	Landmarks	Correct	Bearable	Wrong
1.	Left (eyes)	83% (L)+ 88%(R)	6%(L)+ 7%(R)	11%(L)+ 5%(R)
2.	Left (mouth)	81%	19%	0%
3.	Right (eyes)	84% (L)+ 81%(R)	12%(L)+ 16%(R)	5%(L)+ 2%(R)
4.	Right (mouth)	91%	5%	0%
5.	Top (eyes)	84%(L)+ 70%(R)	12%(L)+ 21%(R)	5%(L)+ 9%(R)
6.	Top (mouth)	66%	14%	20%
7.	Bottom (eyes)	79% (L)+ 74%(R)	14%(L)+ 14%(R)	7%(L)+ 12%(R)
8.	Bottom(mou th)	88%	7%	5%

Since the bearable label means that there will be no effect to the further computation like expression detection thus an overall accuracy of 93% has been achieved. Only the detection accuracy of the top landmark for mouth is significantly lower due the fact that in an open mouth or smile showing teeth Viola-Jones Feature Detection fails to bind the rectangle over the whole mouth thus the sub- region for the search of top landmark becomes faulty. This aggravation marks a big disadvantage in the general use of classifier that is based on the object detection approach.

CONCLUSION AND FUTURE WORK

In this paper we proposed a reliable landmark detection algorithm with low detection time. After this we can apply Deformation methods as Deformation methods applied to facial expression recognition include DCT [29], Gabor wavelets [29], neural networks [29] and Active Appearance Models the usage of Haar like features reduces the search area significantly in very less time without any complex computation. This also increases the whole performance of the system. The designed algorithm has a big potential with the opportunity of further improvements. . If the face in distorted significantly like in the mouth curled up to one side in disappointment or when the lips are parted significantly, the feature detection shows error in mouth detection which affects the landmarking and can be removed by using a correction method. The last improvement that will be listed here is an increase of the number landmarks to detect. That makes an implementation of an identification or verification system possible.

REFERENCES

[1] Çeliktutan, Oya, Sezer Ulukaya, and Bülent Sankur. "A Comparative Study of Face Landmarking Techniques", EURASIP Journal on Image and Video Processing, 2013.

[2] K Liu, A Weissenfeld, J Ostermann, X Luo" Robust AAM building for morphing in an image-based facial animation system" ,In Proc.of Int .Conf. on Multimedia and Expo.,(Hannover, Germany, 2008), pp. 933–936

[3] U Park, AK Jain," 3D face reconstruction from stereo-images", In Proc. Of Int .Workshop on Video Processing for Security (Quebec City, Canada, 2006), p. 41

[4] F Dornaika, F Davoine, , "Online appearance-based face and facial feature tracking ", In Proc. of Into .Conf. on Pattern Recognition (Washington, DC, USA, 2004), pp. 814–817

[5] Humid Luann, Mohammed Ouanan, Brahim Aksasse, "Facial landmark localization: Past, present and future", 4th IEEE International Colloquium on Information Science and Technology (CiSt), 2016.

[6] K.H. Park, H.E. Lee, Y. Kim, Z.Z. Bien, "A Steward Robot for Human Friendly Human-Machine Interaction in a Smart House Environment*,"* In IEEE Transactions on Automation Science and Engineering*,* vol. 5, no. 1, pp. 21-25, 2008.

[7] C. Lisetti, S. Brown, K. Alvarez, A. Marpaung, "A Social Informatics Approach to Human-Robot Interaction with a Service Social Robot*,"* In IEEE Systems, Men, and Cybernetics. Special Edition on Human-Robot Interaction, vol. 34 no. 2, 2004

[8] C. Du, Q. Wu, J. Yang, and Z. Wu, "SVM based ASM for facial land- marks location," In Proc. IEEE International Conference on Computer and Information Technology (CIT'08), Nov 2008, pp. 321–326

[9] F. Cootes and C. J. Taylor, "Active shape models - smart snakes*,"* In Proc. British Machine Vision Conference (BMVC'92), Leeds, UK, set 1992, pp. 266–275

[10] T. F. Cootes, G. J. T Edwards, and C. J. Taylor, "Active appearance models," In Proc. of the European Conference on Computer Vision (ECCV'98), Freiburg, DE, Jun 1998, pp. 484–4.

[11] Lecture Notes in Computer Science, Springer, 2015.

[12] AA Salah, N Alyüz, L Akarun, "Registration of 3D face scans with average face models", J. Electron. Imag. 17(1), 011006 (2008)

[13] Advances in Face Detection and Facial Image Analysis, Springer, 2016.

[14] Tistarelli, M.. "Active vision-based face authentication", Image and Vision Computing, Elesvier, 20000301.

[15] M. Pantic and L.J. Rothkrantz, "Automatic analysis of facial expressions: The state of the art*,"* In IEEE Trans. Pattern Anal. Mach. Intell. vol. 22, no. 12, pp. 1424-1445, 2000

[16] Mi Young Nam. "Facial landmark detection system using interest-region model and edge energy function", 2009 IEEE International Conference on Systems Man and Cybernetics, 10/2009.

[17] Seyed Mehdi Lajevardi. "Contourlet structural similarity for facial expression recognition", IEEE International Conference on Acoustics Speech and Signal Processing, 2010

[18] Dibeklioglu, H., A. A. Salah, and T. Gevers. "A Statistical Method for 2-D Facial Landmarking", IEEE Transactions on Image Processing, 2012.

[19] Khaizuran Abdullah. "QAM modulations over wavelet based OFDM channel for facial expression recognition", Australasian Telecommunication Networks and Applications Conference (ATNAC), 2009.

[20] G.M. Beumer, Q. Tao, A.M. Bazen, R.N.J. Veldhuis. "A Landmark Paper in Face Recognition", 7th International Conference on Automatic Face and Gesture Recognition (FGR06), 2006.

[21] Lecture Notes in Computer Science, Springer, 2012.

[22] J. Canny," A Computational Approach to Edge Detection*",* IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 8, no. 6, June 1986.*

[23] Yao, Na, Tiecheng Bai, Xia Jiang, and Haifang Lv. "Improved Harris Corner Detection for Chinese Characters", Fourth World Congress on Software Engineering, 2013.

[24] Samer R. Saydam. "Contourlet Based Interest Points Detector", 20th IEEE International Conference on Tools with Artificial Intelligence, 2008.

[25] Lecture Notes in Computer Science, Springer, 2013.

[26] Siu-Yeung Cho. "Facial emotion recognition by adaptive processing of tree structures", Proceedings of the ACM symposium on Applied computing - SAC 06 SAC 06, 2006.

[27] HC, Wu CY, Lin TM ," Facial Expression Recognition Using Image Processing Techniques and Neural Networks Advances in Intelligent Systems & Application",. Springer-Verlag Berlin Heidelberg, pp 259–267,2013.

[28] Ma L, Khorasani K ," Facial expression recognition using constructive feed forward neural networks", *IEEE* Trans Syst Man Cybern B: Cybern 34:1588–1595,2004.

[29] P. J. Phillips, H. Moon, P. J. Rauss, and S. A. Rizvi, "The FERET evaluation methodology for face-recognition algorithms," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 10, pp. 1090–1104, 2000.

[30] Stephen Moore. "Automatic Facial Expression Recognition Using Boosted Discriminatory Classifiers", Lecture Notes in Computer Science, 2007

[31] Martin Bockeler, Xuebing Zhou "An Efficient 3D Facial Landmark Detection Algorithm with Haar-like Features and Anthropometric Constraints*,"* in IEEE International Conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt.5-6 Spt. 2013, INSPEC Accession Number 13826547, pp. 1-8.

[32] M. A. Nicolaou, H. Gunes, and M. Pantic, "Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space*,"* IEEE Transactions on Affective Computing, vol. 2, no. 2, pp. 92–105, 2011.