

Minimum AUs for Real-Time Facial Expression Recognition in Frame Sequence

Saghir Ahmed Alfasly
Department of Computer Science,
Kuvempu University,
Karnataka, INDIA

Suresha M., PhD
Assistant Professor,
Department of Computer Science,
Kuvempu University,
Karnataka, INDIA

ABSTRACT

Automatic emotion recognition is one of the most challenging tasks in computer vision and robotics. Although the verbal communication is an essential element of information exchange, the communication would be more effective and efficient by involving non-verbal communication including facial expression interpretation. Many approaches and methodologies have been proposed in terms of face segmentation, facial features extraction, and emotion classification. This article discusses a model for Facial Expression Recognition, which recognize the emotions from human facial expressions in live video acquired with web camera or using recorded videos. This article explains an entirely automated system for emotion recognition of six emotions (anger, disgust, fear, happiness, sadness and surprise) plus neutral state, involving image acquiring, preprocessing, face detection, segmentation, features extraction, encode AUs and finally classification. This system acquires images either with web camera or import recorded videos. Finally, it utilizes the minimum AUs for emotion classification with Rule-Based Classifier.

General Terms

AUs Action Units, Image processing, Pattern Recognition, Classification

Keywords

Facial Expression Recognition, Facial Action Units, Rule-Based Classifier.

1. INTRODUCTION

Facial expression is one of the atomic units of the visual human communication. Human communication can be smoother and the communicators react undoubtedly and much understandable with facial expressions as a research in social psychology has shown [1,2]. These results show the importance of the visual human communication especially facial expressions, which makes the human emotion recognition one of the current active fields in computer vision and artificial intelligence.

Human-Machine interaction, behavioral and medical sciences are some fields that facial expression recognition can be used.

Many models and algorithms have been developed to achieve good accuracy with efficient processing in real time in frame sequences, where some of these models use geometric features and local descriptors. However, some others use the FACS system which consists of 44 Action Units (AU). Each AU defines a set of muscle movements is defined by an Action Unit (AU). The classifier, that can be used, differs depending on the features and time factors such as Support Vectors Machines (SVMs) [3], K-Nearest Neighbor (KNN) [4,5], Bayes [6], Rule-Based, Tree based classifier or ANN.

The face features extraction based on Principle Component Analysis (PCA) or any other approach.

In this paper, we propose a real-time facial expression recognition model based on action units (AUs). This system extracts the facial features, encodes the face actions and classify of face actions in six basic emotion categories (anger, disgust, fear, happiness, sadness and surprise) plus the neutral state automatically, without any manual intervention.

2. RELATED WORK

Many researchers have been working on the computer vision approach for facial expression recognition since the early nineties, when Ekman proposed six basic facial emotions which refer to happiness, anger, surprise, fear, sadness, and disgust [7].

Most of facial expression recognition researches achieved on these six classes, but recently, efforts on emotion recognition turn to the recognition of complex and spontaneous emotions rather than those prototypical ones [8,9]. Fortunately, recent inventions in 3D technologies solve the problem perfectly by providing the means to measure and reconstruct 3D faces. Compared with 2D images, 3D faces hold more geometric shape data and are invariant to variables in recording conditions.

There are two kinds of features represent the AUs: Geometric-based features and appearance-based features. The shape of the face is represented based on the location of a set of landmark points conducted by means of ASM [10]. In the other side, modeling the face appearance changes can be conducted with appearance features such as Local Binary Patterns (LBP) [11], Histograms of Oriented Gradients [12] and Gabor wavelets [13]. Additionally, AUs can be detected and represented either on static frames or in a video (frame sequences) with temporal dynamics.

Tian et al. [14] used permanent features such as mouth, brows, eyes and transient features such as deepening of furrows to develop an Automatic Face Analysis system, which recognizes any changes in face muscles into action units (AUs). Valstar et al. [15] use the dynamics of a set of tracked fiducial points to recognize AUs in frame sequences. Jiang et al. [16] use LBP [11] and LBP-TOP [17] as descriptors for AU detection they have got a high accuracy. Tian et al. [18] use combination of Gabor wavelets and geometric descriptors, this combination obtained the best performance of AU recognition. Almaev and Valstar [19] use a combination of LBP-TOP and Gabor wavelets to develop a dynamic appearance based descriptor.

2.1 Emotion recognition

The most common approach to model emotions in computer vision is to rely on six basic emotions, namely anger, contempt, disgust, fear, happiness, sadness and surprise along

with Neutral state. Notice that these are the six basic emotions defined by Ekman [20] with the addition of contempt. As in the case of AUs classification, emotions can be detected on static frames or frame sequences using geometric-based and appearance based descriptors or a combination of both. Similarly to AUs recognition, facial expressions can be detected on static frames or frame sequences using either geometric based or appearance based descriptors or a combination of both [21].

Zhang et al. [22] concluded that appearance based features (Gabor wavelets) outperformed the geometric based features after conducting a comparison between both kinds of descriptors. Add to that they proved a slight improvement by

On the other hand, Shan et al. [25] use appearance descriptor approach model. The LBP with multi-scaled sliding windows have been used for emotion detection. Bartlett et al. [13] used a bank of Gabor filters for emotion recognition in a real-time streaming, where a set of Aus together defines an emotion. In [26] Pantic and Rothkranz developed an expert system for predict the emotion based on the presence and absence of Aus.

3. METHODOLOGY

Developing an entirely automatic system for facial expression recognition is the main objective of our research. So, we can list the major functionalities that the system conducts:

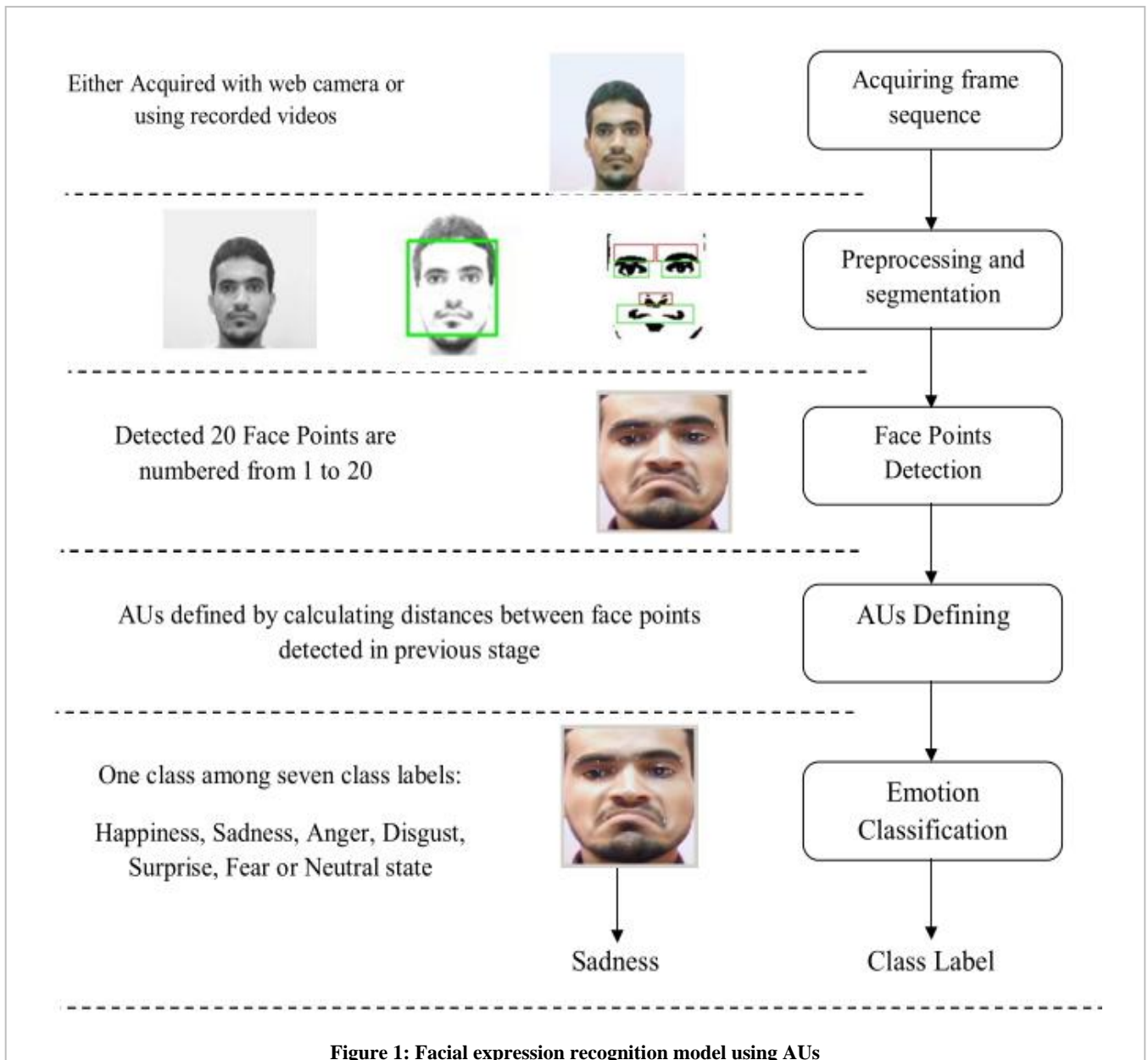


Figure 1: Facial expression recognition model using AUs

combining both kinds of features than just using Gabor wavelets.

Recently, Jeni et al. [23] dynamic geometric descriptors have been used to recognize emotions. Kotsia and Pitas [24] use grid-tracking and deformation system, based on deformable models, tracks the grid in consecutive video frames over time.

1. Frame sequences acquiring either with web camera or recorded videos.
2. Automatic Preprocessing and segmentation stages.
3. Automatic extraction of the face point model and facial features.

4. Automatic defining AUs based on facial features.
5. Automatic emotion classification into one of the basic six emotion categories (Happiness, Sadness, Anger, Disgust, Surprise or Fear) plus Neutral state.

The general workflow of our system stages can be obviously illustrated in figure 1.

3.1 Face Model

Face model consists of 20 points called the facial feature points, so each point represents one important location on human face. These points are numbered from 1 to 20 as it illustrated in Figure 2.

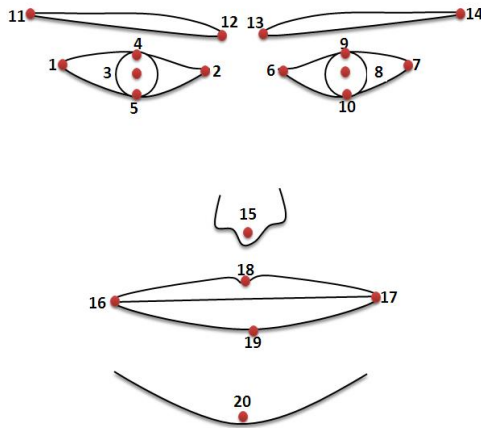


Figure 2: Human Face Model

Face points are described in Table 2.

3.2 Features

Features extracted are represented as the distances among different defined face feature points in Figure 2. Eight features are defined in Table 1.

Table 1, Features calculated from face points

Feature	Feature Description
f1	Distance from 12 to 15
f2	Distance from 13 to 15
f3	Distance from 1 to 16
f4	Distance from 7 to 17
f5	Distance from 2 to 16
f6	Distance from 6 to 17
f7	Distance from 16 to 17
f8	Distance from 12 to 13

Table 2, Face points definition

Point	Point Description
1	Left eye outer corner
2	Left eye inner corner
3	Center of left eye iris
4	Top of the left eye

5	Bottom of the left eye
6	Right eye inner corner
7	Right eye outer corner
8	Center of right eye iris
9	Top of the right eye
10	Bottom of the right eye
11	Left eyebrow outer corner
12	Left eyebrow inner corner
13	Right eyebrow inner corner
14	Right eyebrow outer corner
15	Center of Nose
16	Left corner of the mouth
17	Right corner of the mouth
18	Top of the upper lip
19	Bottom of the lower lip
20	Tip of the chin

3.3 Action Units (AUs)

Action Units are built based on eight features defined previously in Table 1. However, only ten AUs have been built to represent facial expressions. These AUs have been chosen to minimize the number of AUs as much as it is possible to promote the efficiency and the effectiveness of the system as well as providing an acceptable level of results accuracy.

Table 3, Action Units (AUs)

AU	Description	Mapped on Model
1	Mouth corners pulled up	Decreased f3 and f4
2	Mouth corners pulled down	Increased f3 and f4,
3	Raised brows	Increased f1 and f2
4	Lowered brows	Decreased f1 and f2
5	Mouth corner pulled inward	Decreased f7
6	Mouth corner pulled outward	Increased f7
7	Mouth corners pulled up2	Decreased f5 and f6
8	Mouth corners pulled down2	Increased f5 and f6
9	Eyebrows pulled inward	Decreased f8
10	Eyebrows pulled outward	Increased f8

3.4 Preprocessing

In this stage of the frame sequence preprocessing, the system removes noises and enhances frames for preparing them to next stage. So, this stage includes frame sequence read, color to gray frame conversion. Some filters for noise removing, background removing and edge enhancement have been used. Figure 4 illustrates the output of this stage.

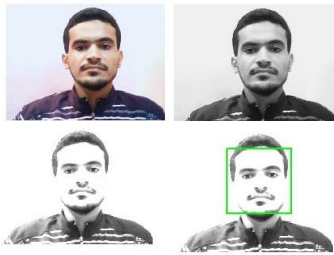


Figure 3: Preprocessing Stage

3.5 Segmentation

This stage consists of many sub stages such as face detection using Viola-Jones algorithm, then pixels sweeping approach [27] has been used for face features detection: eyes, eyebrows, mouth and nose, which shows high accuracy . Figure 6 illustrates these functionalities.

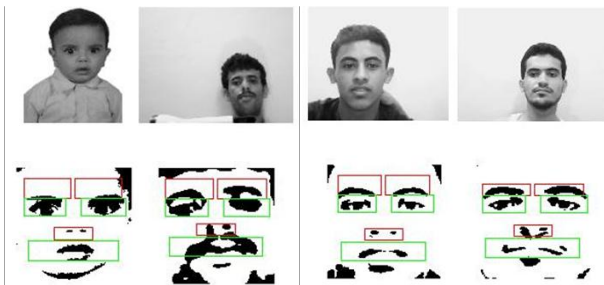


Figure 4: Segmentation Stage

3.6 Face points and features extraction

In this stage, following the face parts segmentation (eyes, eyebrows, nose and mouth), human face feature points should be determined accurately. After finding all needed feature points, mentioned in the table 2, the system calculates the feature values which are represented as distances among specified face feature points. This can be done using Euclidean Distance or any other distance metric.

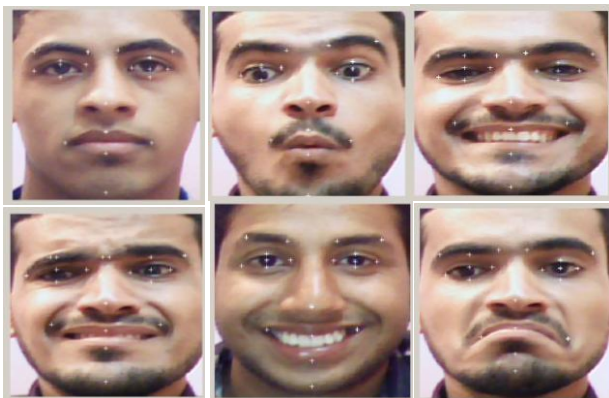


Figure 6: Samples show locations of face Classification

To perform the AU recognition task, the researchers have used the rule-based classifier, due to two factors; first one the small set of features extracted and utilized, Table 4 shows and organizes the minimum AUs for each facial expression. Second factor refers to the ease of utilizing these features with rule-based classifier.

Table 4, Action Units (AUs)

No.	Expression	AUs
1	Happiness	(1+6)or (7+6)
2	Sadness	2 or 8
3	Anger	(2 or 8) + 9 + 4
4	Disgust	7 + 5 + 9 + 4
5	Fear	(1 or 7) + 4 + 9
6	Surprise	3 + 10

4. RESULTS AND DISCUSSION

The researchers conducted their experiments on both types of frame sequence, live videos and ten recorded videos of ten persons. The chart, in figure 5, shows the classification accuracy for the six facial expressions along with neutral state. The accuracy of recognition differs from one facial expression to another, where some expressions are recognized with a high accuracy and some other recognized with medium accuracy. The expressions recognized with high accuracy are - happiness, anger and surprise. On the other hand, the expressions recognized with less recognition accuracy are - sadness, disgust and fear, this can be interpreted due to the high similarities among these expressions

5. CONCLUSION

In this paper, a fully automatic system is presented and developed for a real-time AUs classification. As the results have been achieved, it is obvious to conclude that conducting a facial expression recognition system based only on AUs will not achieve high accuracy for all emotions. Therefore, a combination of different techniques and approaches can be used in order to promote the system performance.

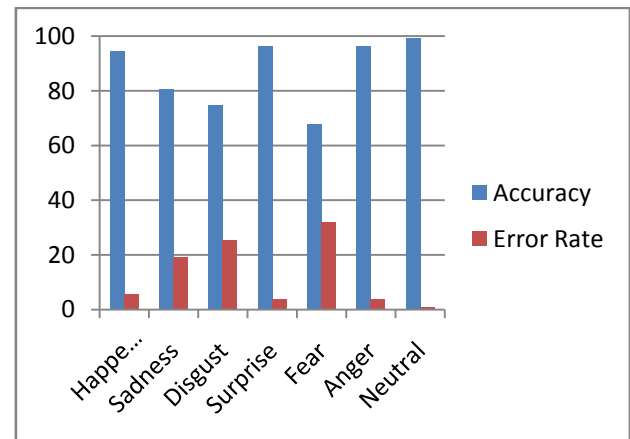


Figure 5: Accuracy of Emotion Recognition

6. REFERENCES

- [1] E. Boyle, A.H. Anderson, A. Newlands, The effects of visibility on dialogue and performance in a co-operative problem solving task, *Language and Speech* 37 (1) (1994) 1–20.
- [2] G.M. Stephenson, K. Ayling, D.R. Rutter, The role of visual communication in social exchange, *Britain Journal of Social Clinical Psychology* 15 (1976) 113–120.
- [3] S. Roweis, L. Saul, Nonlinear dimensionality reduction by locally linear embedding, *Science* 290 (2000) 2323–2326.

- [4] B. Fasel, J. Luettin, Automatic facial expression analysis: a survey, *Pattern Recognition* 36 (1) (2003) 259–275.
- [5] Y.L. Tian, T. Kanade, J. Cohn, Evaluation of Gabor wavelet-based facial action unit recognition in image sequences of increasing complexity, in: *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, 2002, pp. 229–234.
- [6] G. Littlewort, M. Bartlett, I. Fasel, J. Susskind, J. Movellan, Dynamics of facial expression extracted automatically from video, in: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Workshop on Face Processing in Video*, 2004.
- [7] Ekman, P., 1993. Facial expression and emotion. *Am. Psychol.*, 48(4):384-392.
- [8] Zeng, Z., Pantic, M., Roisman, G.I., et al., 2009. A survey of affect recognition methods: audio, visual, and spontaneous expressions. *IEEE Trans. Patt. Anal. Mach. Intell.*, 31(1):39-58. [doi:10.1109/TPAMI.2008.52]
- [9] Vinciarelli, A., Pantic, M., Heylen, D., et al., 2012. Bridging the gap between social animal and unsocial machine: a survey of social signal processing. *IEEE Trans. Affect. Comput.*, 3(1):69-87.
- [10] P. Lucey, J.F. Cohn, I. Matthews, S. Lucey, S. Sridharan, J. Howlett, K.M. Prkachin, Automatically detecting pain in video through facial action units, *IEEE Trans. Syst. Man. Cybern. B Cybern.* 41 (2011) 664–674.
- [11] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: Application to face recognition, *Pattern Anal. Mach. Intell.*, *IEEE Trans.* 28 (2006a) 2037–2041.
- [12] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, *Computer Vision and Pattern Recognition*, 2005. *CVPR 2005. IEEE Computer Society Conference on*, IEEE, 2005, pp. 886–893.
- [13] M.S. Bartlett, G. Littlewort, I. Fasel, J.R. Movellan, Real time face detection and facial expression recognition: development and applications to human computer interaction, *IEEE Computer Vision and Pattern Recognition Workshop*, 2003. *CVPRW'03.*, 2003, p. 53.
- [14] Y.I. Tian, T. Kanade, J.F. Cohn, Recognizing action units for facial expression analysis, *Pattern Anal. Mach. Intell.*, *IEEE Trans.* 23 (2001) 97–115.
- [15] M.F. Valstar, I. Patras, M. Pantic, Facial action unit detection using probabilistic actively learned support vector machines on tracked facial point data, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2005. *CVPR Workshops*, IEEE, 2005, p. 76.
- [16] B. Jiang, M.F. Valstar, M. Pantic, Action unit detection using sparse appearance descriptors in space-time video volumes, *Automatic Face & Gesture Recognition and Workshops (FG 2011)*, 2011 *IEEE International Conference on*, IEEE, 2011, pp. 314– 321.
- [17] G. Zhao, M. Pietikainen, Dynamic texture recognition using local binary patterns with an application to facial expressions, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (2007) 915–928.
- [18] Y.I. Tian, T. Kanade, J.F. Cohn, Evaluation of Gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity, *Proceedings Automatic Face and Gesture Recognition*, 2002., 2002, pp. 229–234.
- [19] T.R. Almaev, M.F. Valstar, Local Gabor binary patterns from three orthogonal planes for automatic facial expression recognition, *Affective Computing and Intelligent Interaction (ACII)*, 2013 *Humaine Association Conference on*, IEEE, 2013, pp. 356–361.
- [20] P. Ekman, W.V. Friesen, Constants across cultures in the face and emotion, *J. Pers. Soc. Psycho.* 17 (1971) 124.
- [21] D. Sanchez-Mendoza et al. Emotion recognition from mid-level features, *Pattern Recognition letters.* 67 (2015) 66–74.
- [22] Z. Zhang, M. Lyons, M. Schuster, S. Akamatsu, Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron, *Automatic Face and Gesture Recognition*, 1998. *Proceedings. Third IEEE International Conference on*, IEEE, 1998, pp. 454–459.
- [23] L.A. Jeni, D. Takacs, A. Lorincz, High quality facial expression recognition in video streams using shape related information only, *Computer Vision Workshops (ICCV Workshops)*, 2011 *IEEE International Conference on*, IEEE, 2011, pp. 2168– 2174.
- [24] I. Kotsia, I. Pitas, Facial expression recognition in image sequences using geometric deformation features and support vector machines, *Image Process.*, *IEEE Trans.* 16 (2007) 172–187.
- [25] C. Shan, S. Gong, P.W. McOwan, Robust facial expression recognition using local binary patterns, *IEEE International Conference on Image Processing*, 2005. *ICIP 2005*, IEEE, 2005, pp. II–370.
- [26] M. Pantic, L.J. Rothkrantz, Expert system for automatic analysis of facial expressions, *Image Vis. Comput.* 18 (2000)8
- [27] Alfasy S.A, Suresha M, A simple approach for facial features detection, *International Journal of Advanced Research in Computer and Communication Engineering*, 5(6):154-158, June 2016.