IJESRT: 8(1), January, 2019

ISSN: 2277-9655

International Journal of Engineering Sciences & Research Technology

(A Peer Reviewed Online Journal)
Impact Factor: 5.164

IJESRT

Website:www.ijesrt.com

Mail: editor@ijesrt.com

IJESRT
INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

METHODS FOR RECOGNIZING FACIAL EXPRESSION AND EMOTIONS USING VIDEO

Brahma Datta Shukla[*1] **& Pragya Singh Tomar**[2]
[*1&2]Institute of Computer Science, Vikram University, Ujjain

DOI: 10.5281/zenodo.4651243

ABSTRACT

Facial gestures and feelings are nothing but reactions to human beings' external and internal events in real life situations. Recognition of the end user's gestures and feelings from video streaming plays a very important role in human computer interaction. In such systems, the complex changes in human face movements need to be quickly monitored in order to provide the necessary response system. In order to avoid road collisions, the only real-time application is physical fatigue detection based on facial detection and expressions such as driver fatigue detection. Physical fatigue analysis or detection based on face expression is beyond the scope of this paper, but this paper reveals research on various methods recently proposed for facial expression and/or video recognition of emotions. This paper presents the methodologies in their comparative analysis in terms of feature extraction and classification used in methods of facial expression and/or emotion detection. The comparative analysis is conducted on the basis of accuracy, instrument of implementation, benefits and disadvantages. The result of this paper is the existing research gap and research problems for video-based facial detection and recognition systems that are still open to solving. Throughout this paper, the survey on recent methods is adequately addressed by considering potential research work.

KEYWORDS: Facial expressions, Frames, Emotions, Expressions, Fatigue, Feature Extraction, Classification.

1. INTRODUCTION

Today, human computer interaction-based systems are used in many real-time applications to track human actions from videos instantly and reliably. One such area is the realisation and monitoring of the expression of the human face and the identification of emotions from video streaming with a distinct intent, such as the detection of physical exhaustion. We incorporate the needs of facial emotion and expression recognition in upcoming sentences before going to address it more first.

In human-to-human speech, in addition to pronouncing a communication channel, the sound of mental, emotional, and even physical state is used in conversations about important information and facial expressions is the notion that the facial expressions of a person in their simplest form are a more subtle happy or angry thought[1], No signal will provide the computing context with feelings or absorption of all speaker desires from listeners, sympathy, or even what the speaker says, enabling our regular human user to stay at the forefront in the fabric can shift to absorb the forecast a generally establishment[2]. It is ubiquitous computing and environmental knowledge that future computing needs to accomplish. To recognise such interfaces and intentions, it is simple to naturally occur multimodal human-human communication-focused outcome to the end user, and as conveyed by feelings of social and emotional signals, the capacity to sense potential nonverbal acts and expressions will need to be created. Automatic recognition was motivated by research. Recognition of facial expression, machine vision, pattern recognition and research into human and computer interaction has been drawn to community notices Automatic detection of facial expressions such that affective computing technologies, including advanced tutoring systems, provide the essence of forms of the next generation computing device, patient tracking systems, etc. The individual's human face, various age groups, genders and other physical features vary from the cause[3].

In day-to-day experiences, emotions are important to human beings and they are used in daily life. In Human-Computer Interaction (HCI), Human Robot Interaction (HRI), and so on, emotion recognition has become a significant and fascinating area of research. The six basic feelings are: illness, happiness, fear, rage, sadness, and surprise. In various applications, computer graphics, automated driver fatigue detection, 3D/4D animation of avatars in the entertainment industry, psychology, video & text chat and gaming applications are included. Preprocessing, feature pulling out, and division are the identification of emotions from facial expressions using images.

In social interaction, the significance of the facial expression system is generally known and social intelligence system analysis has been an active research subject since the 19th century. In 1978, Suwa ET implemented facial expression recognition in Al.

The key point of face identification and alignment was the development of a facial expression recognition system. Feature extraction and classification, image standardization[4].

There are methods we use to classify facial expressions to speed up an effective number of algorithms. By using optical flow proposed for facial, effective algorithm faces motion detection, it should be based on either recognition detection technology that is focused on optical flow vector speed infusion technology. During time intervals, optical flow velocity represents the image changes that the algorithm operates on segmented image frames and we give vector based on their outcomes, the highest degree of equality that defines facial emotions. Job Algorithm (AU) encoded facial expression based on unite database operation. Using this approach, facial expressions can be defined to recognise that four kinds of expression[5] exist. Facial expression utilises the speed of emotion to recognise the first form. The second form of optical flow to define a picture frame using facial expression is the third type of facial expression to recognise the active model of shape to be used. A dynamic multidimensional view is confronted by the fourth type of neural networks[3] using facial expression to identify. It's hard work to model and build a model for face recognition. Several forms of different condition databases (expression, lights, etc.) for a different face [6] are available for facial detection. There are several methods of expressing facial expressions, such as differences in non-monotonic light, random noise and age changes, claiming to suffer from the drawbacks of this process, and defining conditions of expression. While some strategies, such as Gradient face, a variant of high discrimination power lighting, are still known for the capabilities of speech and age variation conditions[7].

There have been a variety of approaches proposed for video-based or image-based human facial expression and identification of emotions over the last decade. Such techniques vary in the methodologies and databases of facial expressions used. The identification accuracy and processing time is critical for every system, these two performance metrics describe the consistency and efficiency of the proposed method. Our aim in this paper is to study such recent facial expression recognition strategies with their advantages and drawbacks based on video inputs.

In reminder of sections, section is presenting the survey and study on recent video based facial expression and emotions recognition methods. Section III provides the comparative study of the approaches discussed in tabular form with a comparison of the accuracy graph. The current shortcomings and the study void are discussed in Section IV. Finally, the conclusion and prospective work addressed in section V.

2. METHODS

In this segment, analysis is discussed on the latest nine facial emotion/expression recognition methods using images. The observed approaches are from 2014, 2015 and 2016.

In[5], the author developed a novel method for finding facial motion and emotion recognition based on video evidence. The online statistical model (OSM) and cylinder head model (CHM) were combined with a 3D deformable facial model to monitor 3D facial movement in the sense of particle filtering. A fast and effective algorithm and a robust and accurate algorithm have been developed for facial expression recognition.

Retrieve facial animations sequentially. After the facial animation was gained, the awareness of static facial expression obtained from anatomical study described facial expression. The second was simultaneous access to facial animation and facial expression to boost usability and robustness of noisy input results. Subsequent facial expression was recognised in this approach through integrating static and acquired dynamic facial appearance information by preparation using a video archive for a multi-class expressional Markov process[5].

A novel video streaming face recognition and face monitoring technique was proposed by the author. The video frame shows no expression about a face's previous localization, nor does it make any comments regarding the posture. The rectangle-like window is drawn by measuring a video frame's top-left, top-right, bottom left, and bottom-right point in face picture outline. Any pre-processing mathematical tasks for a video frame are required to eliminate an error. In terms of counter boundary images, edge output is necessary. After this, to track the face position, the scalar and vector distance between all corner points of two consecutive frames are identified. The relocation of the corner point involves moving the spot and face direction into the next frame. The method used for video-based face recognition as well as tracking[6] is seen in Figure 1.

The user-oriented online video contextual advertising framework was developed by the author. Using Meta-data structure for video data storage and video-based face recognition using machine learning models from the camera with multimedia communications, this strategy was simply a union of networking streaming structure. From the individual captured images, the required object types are defined. This pictures will be analysed under predefined conditions. Based on the specified object type, the device uses the database of multimedia advertisement content and automatically selects and plays appropriate content. In addition, this tool examined current face recognition in video streaming and age assessment from approaches to face images[7].

For the identification of physical exhaustion, the author proposed a different method focused on video facial recording. Using facial videos obtained in a realistic setting of natural illumination, the successful non-contact device was developed to diagnose non-localized physical tiredness from maximum muscle activity, where participants were allowed to willingly shift their head, change their facial expression, and alter their posture. This methodology used a system of finding facial feature points by gathering a 'healthy tracking feature' and a 'supervised descent method' to solve the difficulties starting from realistic scenarios. In order to eliminate incorrect findings by removing low quality faces that appeared in a video series due to issues with natural lighting, head motion, and pose variance, a face quality measurement method was also integrated into this system[8].

As a pre-processor to the supervised ER methods in[9], the author suggested a lightweight online tool to resolve the limitations of conventional supervised ER methods in terms of precision and speed for the neutral vs. emotion classification. In[23], a custom model was created by using a series of reference neutral frames to learn the neutral presence of the user online, thus solving the problems posed by facial perceptions, lighting conditions, etc., when both learning and training exist on the same user. It is difficult to produce emotion-based reference model using emotion frames and it can also not generalise, since it needs several different kinds of emotion frames for model creation from a consumer, thereby also increasing computational complexity. In addition, the device will need a user interface that will direct the user before creating the reference model using all such frames to offer all kinds of emotions. Neutral frames are, thus, the preferred alternative for constructing reference models. The experiment was conducted to determine the probability of a user beginning with a non-neutral term while initiating a mobile phone application with 30 participants. No user was expressly asked to start with neutral [9].

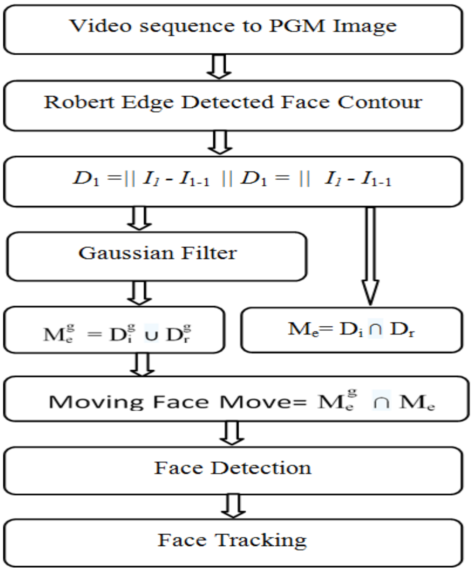
Video sequence to PGM Image
Robert Edge Detected Face Contour
$D_1 = || I_l - I_{1-1} ||$ $D_1 = || I_l - I_{1-1}$
Gaussian Filter
$M_e^g = D_i^g \cup D_r^g$
$M_e = D_i \cap D_r$
Moving Face Move= $M_e^g \cap M_e$
Face Detection
Face Tracking

Figure 1: System Architecture of Method

As a pre-processor to the supervised ER methods in[9], the author suggested a lightweight online tool to resolve the limitations of conventional supervised ER methods in terms of precision and speed for the neutral vs. emotion classification. In[23], a custom model was created by using a series of reference neutral frames to learn the neutral presence of the user online, thus solving the problems posed by facial perceptions, lighting conditions, etc., when both learning and training exist on the same user. It is difficult to produce emotion-based reference model using emotion frames and it can also not generalise, since it needs several different kinds of emotion frames for model creation from a consumer, thereby also increasing computational complexity. In addition, the device will need a user interface that will direct the user before creating the reference model using all such frames to offer all kinds of emotions. Neutral frames are also the preferred alternative for constructing reference models. The experiment was performed with 30 participants to determine the probability of a user beginning when a mobile phone programme is introduced with a non-neutral expression. No user was expressly asked to start with neutral[9]

The author implemented functions to classify the emotions of the BU-4DFE database's video streaming. Videos of 101 topics, 6 BU-4DFE database feelings, used by the authors. The apex frame of a video series is dynamically located by this feature. The Euclidean distance is identified between apex and neutral frame attribute points and their difference in the corresponding neutral and apex frame is calculated to form the vector of the feature that is given to the emotion recognising classifier. With this technique, only two frames and 39 attribute points were used. This minimises the function vector scale. Using SVM (Support Vector Machine) and NN (Neural Network) with various kernels, the classification was completed. The classification time for SVM was determined where, at low computation time, Gaussian RBF, Gaussian RBF (Soft margin) and sigmoid kernel performed reasonably better. This method's preliminary findings suggest that its precision is higher than previous approaches, but has not yet been tested in real-time datasets. The methodology followed by this system is seen in Figure 2[10].

Provided the data matrix X m-by-n, which is treated as the row vector m (1-by-n) x1, x2... The Euclidean distance, xm, between the xs and xt vectors is defined as follows:

$$D=(x_s-x_t)(x_s-x_t) \quad (1)$$

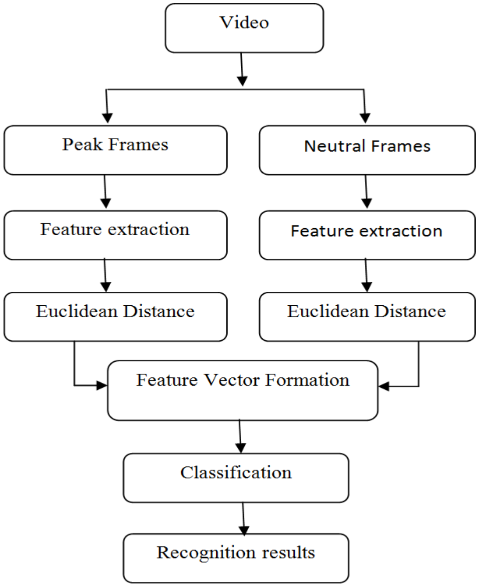

Figure 2: Framework for emotion recognition Using Video Sequences

The author proposed spatiotemporal feature extraction on video samples for facial expression recognition. The proposed spatiotemporal texture map (STTM) is able to record, with poor computational complexity, slight spatial and temporal variations in facial expressions. Next, the Viola-Jones face detector is used to identify the face and frames are removed to eliminate unnecessary background. The proposed STTM, which incorporates spatiotemporal details separated from the three-dimensional Harris corner function, then models facial features. To remove the complex characteristics and present the features in the form of histograms, a block-based approach is used. The characteristics are then categorised by the support vector machine classifier into classes of emotion and voice. The experimental findings indicate that the suggested solution shows the highest success on datasets containing posed expressions, unforced micro expressions, and close-to-real-world expressions, respectively, compared to state-of-the-art approaches with an average recognition rate of 95.37, 98.56, and 84.52percent[11].

The author suggested video frames for accurate facial expression and emotion recognition. In real-world natural contexts, innovative methods called Severe Sparse Learning (ESL), which have the potential to learn a dictionary (set of basis) and a non-linear classification paradigm together, are robustly recognised for facial emotions. This method combines the Extreme Learning Machine (ELM) discriminative power with the sparse representation reconstruction property to allow precise classification when faced with noisy signals and incomplete data recorded in natural settings. In addition, this study introduces a novel distinctive and pose-invariant local spatio-temporal descriptor. This system is capable of achieving state-of-the-art precision of identification on datasets of both acting and random facial emotions[12].

The author adds the identification of another video-based facial expression. The author first merged the LBP-TOP features (which could also be called condensed 3D-LBP) and the Gabor feature representation in this article, inspired by the success of VLBP, to characterise complex facial expression sequences. Then, for grouping, SVM is adopted. The enlarged Cohn-Kanade database (CK+) experiments illustrate the promising success of this method[13].

3. COMPARATIVE STUDY

The tabular structure study of techniques studied in the above section of this paper with their advantages and drawbacks is described in this section. The graph of accuracy is provided after tabular (table 1) review.

ISSN: 2277-9655
Impact Factor: 5.164
CODEN: IJESS7

Table 1: Comparative study of emotion/expression recognition techniques

Paper Title	**Methodology**	**Advantages & Disadvantages**
Spatiotemporal Feature Extraction for Facial Expression Recognition	STTM (spatio-temporal texture map), SVM, Processing of blocks	Advantages: This approach outperforms multiple state-of-the-art attribute extraction approaches dependent on presentation. Disadvantages: The system of classification is simple and not respected under dynamic head movements
Video-Based Facial Recognition Using Histogram Sequence of Local Gabor Binary Patterns from Three Orthogonal Planes	LGBP-TOP, SVM, Gabor Filters	Advantages: Stable and less difficult process Disadvantages: Specifications to increase precision performance.
Dynamic Facial Emotion Recognition from 4D Video Sequences	Euclidean distance, Apex Frames, Neural Network, SVM	Advantages: Relative to complicated approaches, precision is increased. Disadvantages: Not yet measured in real-time terms.
Neutral Face Classification Using Personalized Appearance Models for Fast and Robust Emotion Detection	CLM, Patch processing, LBP KE points	Advantages: The computational benefit of using the suggested system as a pre-processing unit. Disadvantages: under sudden pose variations, CLM fitting can also not be exact.
Robust Representation and Recognition of Facial Emotions Using Extreme Sparse Learning	ESL, ELM	Advantages: Improved results in tough scenarios. Disadvantages: For both function extraction and classification, higher computing costs.

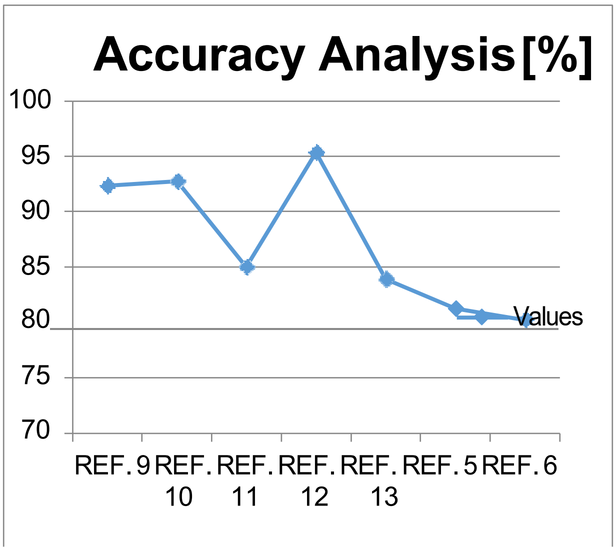

Figure 3: Accuracy Comparison of Different Methods

4. RESEARCH GAP

We reviewed the new techniques for video-based facial expression and identification in the above sections in order to discover their shortcomings and further scope of improvement. The approaches are investigated and contrasted above. For any technique, the most significant output criterion is consistency of identification. From Figure 3, we observed that we have the following findings on the precision efficiency of various methods:

- The stable approach has low success in terms of accuracy of identification.
- Approaches of greater precision (in the 90s), with reliability and robustness for identification.
- Most methods for processing time are not tested, which is also important for all methods of facial expression and emotion detection.
- Some video-based approaches are analysed using databases based on videos.
- The average accuracy is about 95 percent, and also needs to be more improved.

5. CONCLUSION AND FUTURE WORK

The purpose of this paper is to present comparative analysis using various approaches on different techniques of video-based facial and expression recognition. Automatic facial expression and identification of emotions have played an important role in our everyday communication and computer science over the last decade, such as human-computer interaction systems, biometrics, and surveillance, etc. A great deal of profound and fruitful study has been carried out in this field in recent years. With their benefits, drawbacks and precision efficiency, the methods analysed are from 2015 and 2016. In this paper, the existing methods of research problems are described. New method architecture should be undertaken for future work in order to enhance the robustness, efficacy and precision of identification.

REFERENCES

[1] Samad, Rosdiyana, and Hideyuki Sawada. "Edge based Facial Feature Extraction Using Gabor Wavelet and Convolution Filters." In MVA, pp. 430-433. 2011.

[2] Thai, Le Hoang, Nguyen Do Thai Nguyen, and Tran Son Hai. "A facial expression classification system integrating canny, principal component analysis and artificial neural network." arXiv preprint arXiv: 1111.4052 (2011).

[3] Sisodia, Priya, Akhilesh Verma, and Sachin Kansal. "Human Facial Expression Recognition using Gabor Filter Bank with Minimum Number of Feature Vectors." International Journal of Applied Information Systems, Volume 5 – No. 9, July 2013 pp. 9-13. [4] Meher, Sukanya Sagarika, and Pallavi Maben. "Face recognition and facial expression identification using PCA." In Advance Computing Conference, 2014 IEEE International, pp. 1093- 1098. IEEE, 2014.

[4] Jun Yu & Zengfu Wang, "A Video-Based Facial Motion Tracking and Expression Recognition

System", Springer Science Business Media New York 2016.
[5] Aniruddha Dey, "Contour based Procedure for Face Detection and Tracking from Video", 3rd Int'I Conf. on Recent Advances in Information Technology I RAIT-20161, 2016.
[6] Le Nguyen Bao, Dac-Nhuong Le, Le Van Chung and Gia Nhu Nguyen, "Performance Evaluation of Video-Based Face Recognition Approaches for Online Video Contextual Advertisement User-Oriented System", © Springer India 2016.
[7] Mohammad A. Haque, Ramin Irani, Kamal Nasrollahi, Thomas B. Moeslund, "Facial video-based detection of physical fatigue for maximal muscle activity", IET Computer Vision, 2016.
[8] Pojala Chiranjeevi et.al "Neutral face classification using personalized appearance models for fast and robust emotion detection", IEEE Transactions on Image Processing. 2015.
[9] Suja P et.al, "Dynamic Facial Emotion Recognition from 4D Video Sequences", ©2015 IEEE.
[10] Siti Khairuni Amalina Kamarol et.al, "Spatiotemporal feature extraction for facial expression recognition", IET Image Process., pp. 1–8 & The Institution of Engineering and Technology 2016.
[11] Seyedehsamaneh Shojaeilangari et.al, "Robust Representation and Recognition of Facial Emotions Using Extreme Sparse Learning", IEEE Transactions on Image Processing, 2015.
[12] XIE Liping et.al, "Video-based Facial Expression Recognition Using Histogram Sequence of Local Gabor Binary Patterns from Three Orthogonal Planes", Proceedings of the 33rd Chinese Control Conference, July 28-30, 2014, Nanjing, China**.**