Materials Today: Proceedings 80 (2023) 2824–2828

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

Facial emotion recognition methods, datasets and technologies: A literature survey

Prameela Naga [a], Swamy Das Marri [b], Raiza Borreo PhD [c]

[a] University College of Engineering, Osmania University, Hyderabad, India
[b] Chaitanya Bharathi Institute of Technology, Hyderabad, India
[c] Nasser Vocational Training Centre, Bahrain

ARTICLE INFO

Article history:
Available online 20 July 2021

Keyword:
Facial emotion recognition
Dataset
Literature Survey

ABSTRACT

Amidst of the technologies applied to analyze and recognize a facial emotion, there are still clustered challenges that needs to be addressed while building advanced emotion recognition models with more accurate results. This study summarizes the technologies used in the development of various models for facial emotion recognition. Because the advancement in technology is abruptly changing and improving, only the studies for the past two years were included. The studies earlier in the past two years are most likely redundant to the studies prior to it. The study includes discussion of the datasets commonly used by the researchers, the nature of the datasets and its content, and the process of generating the data. Moreover, the highlight of this literature survey is to gather the state-of-the-art technologies and various methods that are applied to dataset for achieving highest possible accuracy rate. However, there are still drawbacks that needs to be addressed.

Selection and peer-review under responsibility of the scientific committee of the International Conference on Nanoelectronics, Nanophotonics, Nanomaterials, Nanobioscience & Nanotechnology.

1. Introduction

Face is the most visible part of the human body, with which the identity of the person as well as their age or even their gender can be also be detected. As many studies in the face recognition conveys, facial emotion recognition rises and become well-known. Scientific research showed a very high potential for facial emotion recognition to be applied in a vast range of applications.

Many companies offering a vast range of consumer products take advantage of face emotion recognition technologies for recognizing customer feedback, identifying potential customers. These facial expressions become more valuable to the field of machine learning.

In the field of education, face emotions are used to determine the level of understanding of the students. Teachers have the freedom to teach their students in a manner that they are comfortable. They deliver the lessons using different platforms, digital or traditional way. The main objective is that at the end of each lesson, the students should understand the lesson. Using face emotion recognition, teachers can evaluate whether the students followed the lecture or not. Based on the feedback determined using the face emotion patterns, teacher may change his/her strategy in delivering the lecture.

Hospitals and healthcare institutions may also rely on the development of face emotion recognition. As patients, become immobilized, their ability to reach out for medical attention become difficult, and also the evaluation of the patients' pain condition relies on continuous monitoring of a medical staff. With face emotion recognition, what the patients' are feeling can be detected and evaluated.

According to the studies, there are 6 basic emotions that can be read by the current state of machine learning models namely – *anger, sadness, happiness, disgust, fear surprise and neutral*. Removing the neutral from the list, the 5 remaining emotions can be grouped into positive (surprise and happiness) and negative (anger, sadness, disgust and fear). The formation of these basic emotions can still be enhanced to reveal deeper and complex emotions.

There are several commercially available face recognition technologies including *Afefctiva Affdex, Microsoft Cognitive Services mod-*

E-mail addresses: prameelakotipalli@gmail.com (P. Naga), msdas_cse@cbit.ac.in (S.D. Marri), raiza.borreo@nvtc.edu.bh (R. Borreo)

https://doi.org/10.1016/j.matpr.2021.07.046
2214-7853/© 2021 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the International Conference on Nanoelectronics, Nanophotonics, Nanomaterials, Nanobioscience & Nanotechnology.

ule Emotion Recognition, Amazon Recognition Face Analysis, and *Neurodata Lab Emotion Recognition* attained their identities in the field of facial emotion recognition [2]. Despite of these, many independent researchers continues to devise their models to address unseen drawbacks.

2. Datasets

For any machine learning technique, datasets plays a major role. Datasets serves as the repository of attributes that are essential in face emotion recognition. Datasets are aggregated beforehand so that it will be ready for extracting and interpreting of the results Similarly, face emotion recognition algorithms also depends on the datasets. Datasets can be static images or videos or a combination of the two. Datasets can also be laboratory-controlled where pictures may be taken in a setup environment. There are also datasets that were naturally taken from the actual settings. The following sub-sections describes various standard datasets.

2.1. JAFFE database

There are 213 photos that are classified into seven facial expressions with added one neutral embedded in the Japanese Female Facial Expression (JAFFE) database. The data were posed by 10 Japanese female models. Each image has been evaluated on 6 emotional adjectives by 60 Japanese subjects. The database was planned and assembled by Michael Lyons, Miyuki Kamachi, and Jiros Gyoba. The emotions consist of *anger, disgust, fear, happiness, sadness, surprise*, and *contempt* [3,18,23].

2.2. Fer2013

One of the most widely used dataset is called FER that stands for Facial Expression Recognition. Composed of 35,887 images, the photos are scaled to 48x48 size setup. FER2013 has 28,709 photos for training, 3589 validation and 3589 testing images. FER2013 has three columns that define each photo. These columns are named as Emotion type in numerical format 0–6 that individually describe as anger, disgust, fear, happiness, sadness, surprised and neutral [24]. The second column is an array of numerical values representing the photos. The last column indicates the type of photo whether a training or a testing data.

2.3. Extended Cohn-Kanade (CK +)

Released in 2000 for automatic detection of facial expressions, there are 593 images taken from 123 people inside the dataset of Extended Cohn_Kanade (CK +). This dataset is composed of 593 sequences of images for 123 individuals. Thirty-one percent of the photos are male and 69% are female with age range of 18–50 years old [21]. In CK + dataset, the images were taken in a laboratory controlled environment. Datasets inside were labeled according to Facial Action Coding System (FACS) [14]. Using the Active Appearance Model (AAM) the face and the facial features were tracked from a video input. SVM was used to classify the facial expressions. See (Fig. 1.).

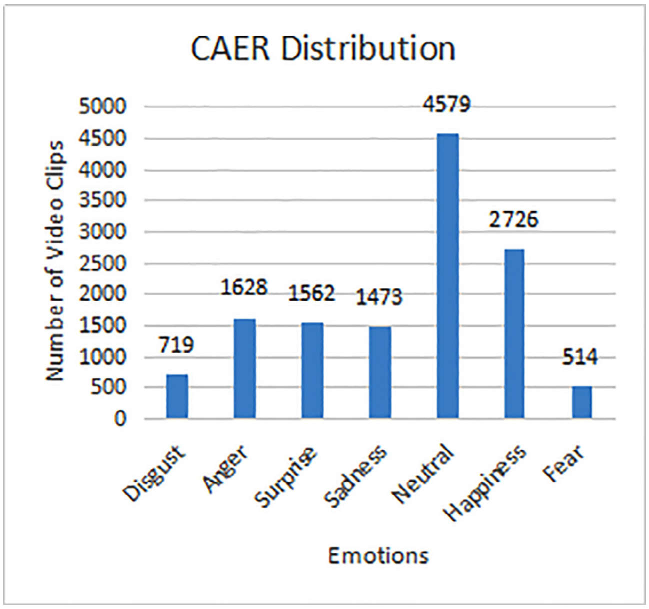

Fig. 1. Graph showing the CAER emotion distribution.

2.4. CMU-MultiPIE

With a total of 305 GB size, the CMU Multi-PIE face database contains 755,370 images of 337 people. These images were recorded up to four sessions for six months. Subjects were under 15 viewpoints or cameras simultaneously taking the images. The images of the subjects were taken under 19 illumination conditions. The frontal images were given the highest consideration by taking a high resolution [19]. The process of evaluating the emotions was based on AAM then PCA and LDA subspaces were computed on the remaining subjects.

2.5. AffectNet

AffectNet was produced due to the fact that image datasets before less consider the valence and arousal that describes images in a continuous dimensional model. With over 1,000,000 data inside, AffectNet is considered as the largest database for face images. About half of it were manually labelled with seven facial expressions and 68 face landmarks [7].

2.6. Iemocap

IEMOCAP is considered as one of the largest free multimodal resources for emotion detection [20]. There are 12 h of recorded audio-visual data captured from 10 humans that were in a conversation. IEMOCAP is an improved version of face database because it does not just consider the basic emotions sadness, anger, happiness and neutral but also annotated valence, dominance as well as the level of emotion activation according to a situation.

2.7. Raf-DB

With 29,672 images, RAF-DB is a collection of facial images gathered from the internet. There were 40 people who manually annotated the photos. The images has diverse characteristics in terms of age, ethnicity and gender. As the acronym of RAF-DB implies, the goal of this database is to evaluate real world complex expressions according to what type of facial expression it is. RAF-DB does not only consider basic emotions but also for compound emotions [17].

2.8. Caer

All the above datasets focus on human face recognition and analysis of emotions only which are not suited for context-aware emotion recognition, hence the CAER was created. It consisted of 13,201 video clips from TV shows, data in CAER was manually

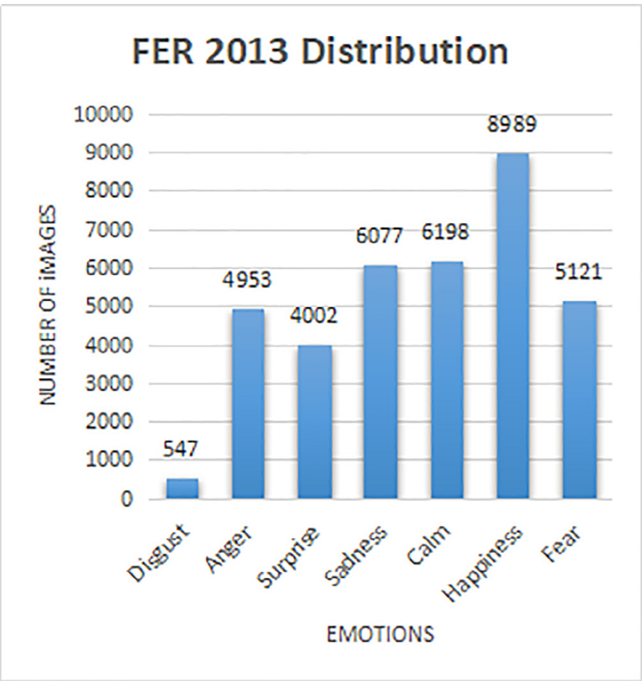

Fig. 2. Emotion distribution of FER 2013 Dataset.

annotated, blindly and independently, by six annotators according to seven basic emotion categories. All the video clips have audio and visual tracks. Three annotators evaluated each clip [10]. See (Fig. 2.).

2.9. iCV-MEFED

A human face would expresses one particular emotion but sometimes it expresses combination of emotions called compound emotions. In the *iCV-MEFED* dataset, a dominant and complementary emotion was used to describe each still images of 31,250 facial faces. The 125 subjects were asked to show 50 different emotions. There were 50 types of compound emotions produced [9].

2.10. Afew

The existing datasets [10,23] suggests that there are seven basic emotions, namely, anger, fear, disgust, happy, sad, surprise and neutral. Acted Facial Expression in the Wild (AFEW) categorize the images inside according to 7 basic emotions. The data type inside are dynamic and spontaneous videos that consists of 773 clips for training, 383 clips for validation and 653 clips for testing [10,24]. The video clips consists of movie fragments where actors were taken in a very natural setup including environmental noise [13].

2.11. Kdef

There are 4,900 images labelled with 7 basic face emotions (happy, sad, surprised, angry, disgust, afraid and neutral) in the KDEF dataset.. These images were taken from 35 male and 35 female participants. Dataset contains 5 different angles [11]. See (Table 1.).

3. Face emotion recognition methods and technologies

The method of recognizing face emotions undergoes several specific steps from face detection to emotion classification. Face detection is done by isolating the face from others by determining the existence of facial features, eyes, nose and mouth. From this point, the machine proceeds to classifying the emotions.

3.1. Face detection

Face detection is the ability of the computer to recognize a face. Haar feature-based cascade classifiers is simple and robust making it a very famous face detection model [1]. The Haar Cascade was used to detect the mouth and the eyes only [3].

Viola Jones algorithm is also a famous face detection model. It uses rectangular features to identify the human face in the image [4].

3.2. Feature extraction

Features are the key focal marks in the face that gives the shape and location of the important biometric parts of the face such as eyes, nose and mouth. Through clustering technology, similar features can be grouped together. Among the popular clustering techniques, K-mean algorithm is widely used method.

3.3. Data augmentation

In order to achieve desirable results in the emotion classification, data enhancement is conducted. One good practice is to augment the dataset in lower or higher contrast and brightness. Resizing of the images making it all uniform is also conducted.

Researchers also use filtering and edge detection techniques using Sobel edge detection method [3].

3.4. Emotion classification

SVM are supervised learning models with associated learning algorithms that analyse data used for classification and the regression analysis [5].

Like how human brains work, Convolutional Neural Network makes it easier and effective to analyze face emotions [11]. As a deep learning tool, CNN learns the patterns in the data. It could be the colored, bright or dark spots. Being a multi-layer algorithm, CNN also recognizes face landmarks. The computer sees only the two dimensional-array of an image composed of pixels. Each pixel is numbered and then compared if they matches or not. Then, CNN compares group or parts of the images rather than the whole image. CNN involves filtering, pooling, normalization and layering using Rectified Layer Units (ReLU). CNN is considered as a very robust algorithm for various image and object recognition tasks because of its hierarchical structure and powerful feature extraction capabilities from an images.

3.5. Real time emotion recognition

Facial emotion recognition is much easier to be tested in a controlled environment compared to uncontrolled one due to change in occlusion, illumination and pose [2]. But it is also applied to a much complex setup. Recognizing emotions in a real time environment is quite challenging due to the fast changes in the face angle, illumination and pose. Thus, a dedicated hardware device to process the images is recommended. Raspberry Pi can be placed in a robot to recognize emotions even in a robust environment allowing better accuracy and higher speed [15].

3.6. State of the art technologies, measures and models

CAER-Net is annotated for context-aware emotion recognition. In the study by [17] from CAER dataset with various length of

Table 1
Various face emotion datasets.

Sno	Name of the dataset	Features	Size	Type	Form of collection	Applications in Facial Emotion Recognition
2.1	**JAFFE**	213 posed images of Japanese female in .tiff format labelled with 7 basic expressions + neutral	1,000	Images	Wild	Virtual learning environment, facial landmark, EmotionalDAN [6]
2.2	**FER2013**	Photos represented by pixels are labelled with 0–6 according to the face emotion	35,887	Images	Wild	Convolutional Neural Network, Attentional Convolutional Network, Emotion recognition for video clips using CNN
2.3	**Extended Cohn-Kanade (CK +)**	Diverse images of individuals 18–30 years of age from African-American, Asian and Latin races.	500	Images	Laboratory controlled	SVM and NLPCA, Attentional Convolutional Network
2.4	**CMU-MultiPIE**	Subjects were taken photos under 19 illumination situations.	750,000	Images	Laboratory controlled	Raspberry Pi with ASM, Adaboost
2.5	**AffectNet**	Emphasis on valence and arousal that identifies images in a continuous dimensional model	1,000,000	Images	Wild	Context-aware emotion recognition, real-time emotion recognition
2.6	**IEMOCAP**	Addition of the level of emotion activation according to a situation	12 h	Videos	Laboratory controlled	Multi-modal emotion recognition
2.7	**RAF-DB**	30,000 crowdsourced annotated facial images	30,000	Images	Wild	CNN, Neighborhood features
2.8	***CAER***	Careful selection of video clips to identify emotions	13,000	Videos	Wild	Context-aware emotion recognition
2.9	***iCV-MEFED***	Compound emotion	31,250	Images	Laboratory controlled	Dominant and complementary emotion recognition
2.10	***AFEW***	dynamic temporal facial expressions from movies		Images	Wild	Video clips
2.11	**KDEF**	Photos were taken from 5 various angles.	4900	Images	Laboratory-controlled	Deep learning

videos, single non-overlapped consecutive 16 frame clips were randomly extracted. After the creation of the novel CAER dataset, a new model called CAER-Net was designed that focuses on both face and attentive context regions in the clips. This allows the recognition of human emotion from the images and videos.

SASE-FE is the first dataset of facial expressions that are either congruent or in-congruent because of the underlying emotional types.

3.6.1. Microsoft Hololens

Along with algorithms for face emotion detection, some popular hardware devices are used to increase the accuracy of the results. The Microsoft Hololens used a depth camera that enables 3D face detection. Developed as the first wireless and computer-controlled holographic smart glasses, Microsoft Hololens enables interface between user and digital data, and between user and holographic images in the real world. To enable the face detection using Hololens, an application called Microsoft Azure Face API is used.

3.6.2. Multi-modal systems

Combining different inputs such as image, video and voice is a new technique to determine the exact emotion of a person. Multi-modal systems are used to compensate the low accuracy rates for some emotions [13]. The voice containing vocabulary, syntax and use of words can bring a deeper emotion from a person.

3.6.3. Strrn

In data analysis, spatial temporal, is used when data is collected across both space and time. It can also be applied to image processing. For face emotion classification, spatial–temporal recurrent neural network or STRRN is a novel deep learning framework that combines feature learning of spatial and temporal information of signal sources [22].

3.6.4. Stationary wavelet entropy

Extracting the features of a face involves getting specific focal points that changes whenever the person changes an emotion. These focal points are the basis of what emotion is expressed and stationary wavelet entropy is used to extract these features [16].

3.6.5. Island loss

As described in paper of [8], island loss is used to enhance the discriminative power of the deeply learned features. It is designed to reduce the intraclass variations while enhancing the inter-class differences.

3.6.6. Deeply supervised CNN

Hybrid model is the combination of two or more models in order to achieve higher accuracy rating in classifying face emotions. Feed Forward Neural Network and Naïve Bayes were both used in training the dataset. First, the inputs will be processed by feed forward neural network then the output from which will be given as an input to the Naïve Bayes algorithm [15].

Feature extraction is also applied to preprocessed images using Active Shape Model. From a series of features detected, it can be reduced to fewer numbers. The Euclidean distances of the remaining features are computed and then combined to form a feature vector that is fed to AdaBoost for classification [19].

3.6.7. Facial expression sentence generating model

What else could be more fascinating when the face emotion recognized is interpreted into a more meaningful output set of words called sentence? Text-based facial expression description using several essential elements were used to describe comprehensive facial expressions. These elements are gender, facial action units, and intensities. To verify the results, the researchers created comprehensive facial expression sentence generating model along with facial expression recognition model for a single facial image.

3.6.8. Compound emotion recognition

Humans can recognize the six basic emotions only that can be expounded to more facial expressions called compound emotion. Compound emotion is characterized by the combination of basic emotions such as happily-disgusted or sadly-fearful. The two emotions are labeled as dominant and complementary emotion.

4. Comparison metrics for evaluating various emotion methods

4.1. Accuracy

For one certain emotion, dataset, training data and testing data, a model created might not be copied with another method [12]. The goal of every model is to achieve higher accuracy results than the previous studies related to it that is based on confusion matrix. There are many factors affecting the accuracy like the dataset.

4.2. Average processing time

The time it takes to recognize emotions in real time is significant in telling whether the model is effective or not. Average processing time and accuracy can be used side by side. In robust environment, such as wild, hardware devices like Raspberry Pi are used.

4.3. Quality of datasets

Datasets are very important in the testing of the accuracy of the models. The data must be properly labeled. However, some existing datasets are still not well-defined just like the FER2013. Although considered as a huge dataset, one of the challenges on this dataset is the unbalanced data. In addition, this dataset has invalid samples like non-face images, wrong cropped face, and expression labeling mistakes [6]. Moreover, the lack of samples from AFEW dataset shows that emotion recognition on video clips has not been completely solved [23]. Confusion in the face emotion classification is also common. Fear and surprise, highly considered as opposite emotions were most likely to have an error because of the similarity in facial features [4].

5. Conclusion and future scope

Studies focusing on facial emotion recognition have gone too far. Datasets, face detection algorithm and classification algorithm were presented in different ways depending on the objective of the study. Choosing a dependable dataset is very important as it contributes to the level of accuracy. One must consider the type of data inside and the size of data. Datasets can be classified as laboratory controlled or naturally setup. Some datasets were labelled manually by validators and some were categorized by computer algorithm like SVM, PCA and LDA. Large datasets are more advantageous and useful generating to higher accuracy results.

The methods used in these studies are very much alike in terms of face detection which is the first phase in face emotion recognition. Haar cascades and Viola Jones algorithm are basically used for face detection. Augmentation or the process of enhancing the photos to get a more visible face features is also applied.

Classifying the faces as to what emotion it belongs can be done using the Convolutional Neural Network (CNN) and Support Vector Machine (SVM). Along with these algorithms, development of classifiers is given focus to enhance the accuracy levels.

In the future, further studies on mapping of the various elements in fulfilling facial recognition to different studies are suggested. In this way, we can see how the datasets affect the performances of the each of the steps in face emotion recognition and vice versa.

CRediT authorship contribution statement

N. Prameela: Conceptualization, Methodology, Software, Data curation, Writing - Editing. **M. Swamy Das:** Supervision, Validation, Visualization, Writing - review & editing. **Raiza Borreo:** Writing - original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A.S. Aljaloud, H. Ullah, A. Alanazi, Facial emotion recognition using neighborhood, Int. J. Adv. Computer Sci. Appl. 11 (1) (2020) 299–306.
[2] A. Mahmood, S. Hussain, K. Iqbal, W.S. Elkilani, Recognition of facial expressions under varying conditions using dual-feature fusion, Hindawi Math. Prob. Eng. 2019 (2019) 1–12.
[3] D. Yang, A. Alsadoon, P.W.C. Prasad, A.K. Singh, A. Elchouemi, An emotion recognition model based on facial recognition in virtual learning environment, Procedia Computer Sci. 125 (2018) 2–10.
[4] E. Dandil, R. Ozdemir, Real-time facial emotion classification using deep learning data science and applications, 2(1), 2019, 13-17.
[5] G. Tonguç, B. Ozaydın Ozkara, Automatic recognition of student emotions from facial expressions during a lecture, Computers Edu. 148 (2020) 103797, https://doi.org/10.1016/j.compedu.2019.103797.
[6] H.-D. Nguyne, S. Yeom, G.-S. Lee, H.-J. Yang, I.-S. Na, S.-H. Kim, Facial emotion recognition using an ensemble of multi-level convolutional neural networks, International Journal of Pattern Recognition and Artificial Intelligence, vol. 33, no. 11, 2019.
[7] I. Tautkute, T. Trzcinski, A. Bielski, I know how you feel: Emotion recognition with facial landmarks, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, 1878-1880.
[8] J. Cai, Z. Meng, A. Khan, Z. Li, J. O'Reilly, Y. Tong, Probabilistic attribute tree in convolutional neural networks for facial expression recognition, ArXiv abs/1812.07067, 2018.
[9] J. Guo et al., Multi-modality network with visual and geometrical information for micro emotion recognition, in 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, 2017, pp. 814-819, 2017.
[10] J. Lee, S. Kim, J. Park, K. Sohn, Context-aware emotion recognition networks,“ in CVF International Conference on Computer Vision (ICCV), 2019, Korea Seoul, 20142-10151.
[11] M.A. Ozdemir, B. Elagoz, A. Alaybeyoglu, R. Sadighzadeh, A. Akan, Real time emotion recognition from facial expressions using CNN architecture, in Medical Technologies Congress (TIPTEKNO) Izmir, Turkey, 2019, 1-4.
[12] M. Ley, M. Egger, S. Hanke, Evaluating methods for emotion recognition based on facial and vocal features, in Poster and Workshop Sessions of the European Conference on Ambient Intelligence 2019, Rome, Italy, 2019, pp. 84-93.
[13] M. Malygina, M. Artemyev, A. Belyaev, O. Perepelkina, Overview of the advancements in automatic emotion recognition: comparative performance of commercial algorithms, PsyArxivprints, 2019.
[14] P. Lucey et al., The extended Cohn-Kanade DATASET (CK+): a complete dataset for action unit and emotion-specified expression, in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, San Francisco, CA, 2010, pp. 94-101.
[15] R. Shahane, K.R. Sharma, S. Siddeeq, Emotion recognition using feed forward neural network & naïve bayes, Int. J. Innov. Technol. Explor. Eng. (IJITEE) 9 (2) (2019) 2487–2491.
[16] S.-H. Wang, P. Phillips, Z.-C. Dong, Y.-D. Zhang, Intelligent facial emotion recognition based on stationary wavelet entropy and Jaya algorithm, Neurocomputing 272 (2018) 668–676.
[17] S. Jyoti, G. Sharma, A. Dhall, Expression empowered residen network for facial action unit detection, in IEEE International Conference on Automatic Face and Gesture Recognition 2019, Lille, France, 2019, pp. 262-269.
[18] S. Minaee, A. Abdolrashidi, Deep-emotion: facial expression recognition using attentional convolutional network, ArXiv abs/1902.01019, 2019.
[19] S. Palaniswamy, S. Tripathi, Emotion recognition from facial expressions using images with pose, illumination and age variation for human-computer/robot interaction, J. ICT Res. Appli. 12 (1) (2018) 14–34.
[20] S. Tripathi, H. Beigi, Multi-modal emotion recognition on iemocap dataset using deep learning, arXiv preprint arXiv:1804.05788, 2018.
[21] T. Kalsum, M. Majid, S. Anwar, Emotion recognition from facial expressions using hybrid feature descriptors, The Institute of Engineering and Technology, pp. 1003-1012, 2019.
[22] T. Zhang, W. Zheng, Z. Cui, Y. Zong, Y. Li, Spatial–temporal recurrent neural network for emotion recognition, IEEE Trans. Cyber. 49 (3) (2019) 839–847.
[23] X. Zhongzhao, Y. Li, X. Wang, Z. Liu, Convolutional neural networks for facial expression recognition with few training samples, in 37th Chinese Control Conference (CCC), China, 2018, pp. 1-6.
[24] Y. Fan, J. Lam, V. Li, Multi-region ensemble convolutional neural network, Springer International Publishing, 2018.

Further Reading

[1] T. Lui, L.K. Tianhao, M. Wang, Multi-feature based emotion recognition for video clips, ICMI'18, 2018, Boulder, CO, USA 630-634.