Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/37/20230473

# Facial expression recognition with computer vision

**Huang Yu**

Leeds College, Southwest Jiaotong University, Chengdu, China, 610000

sc20hy@leeds.ac.uk

**Abstract.** Facial Expression Recognition (FER) is a specialized field within the domains of computer vision and pattern recognition, which is dedicated to the automated identification and examination of facial expressions. Facial expression recognition (FER) has attracted considerable scholarly interest in recent years owing to its diverse array of applications and its potential ramifications across multiple disciplines, such as psychology, human-computer interaction, marketing, and security systems. The objective of this study is to present a thorough examination of the scholarly progression of FER, elucidating the significant achievements, approaches, and obstacles encountered by researchers in this domain. The study presents a selection of databases that are appropriate for Facial Expression Recognition (FER) and conducts a comparative analysis of these databases. The primary methodologies are examined, and recommendations are provided for each stage. In conclusion, this research presents several suggestions for addressing both obstacles and potential in future research endeavors.

**Keywords:** Facial Expression Recognition, Human-Computer Interaction, Pattern Recognition.

## 1. Introduction

Facial expression recognition (FER) is a specialized area within the discipline of computer vision that is dedicated to the examination and analysis of facial expressions in order to infer the emotional states, intents, and cognitive processes of persons. It functions as a pivotal element in comprehending human behavior and has garnered significant momentum in recent decades. Numerous studies have demonstrated the practical significance of various interventions in the domains of medical care, advertising, and education. The optimization of a product by its data owner is contingent upon the happiness of the consumer. The automatic recognition of facial expressions exhibited by users or customers can be utilized to provide feedback to the system, hence enhancing the system's ability to deliver improved service. In recent times, there has been an increasing scholarly focus on the field of facial expression identification, with the objective of examining and comprehending the emotional states conveyed by individuals through visual media such as photos and videos.

The origins of FER research may be traced back to the 1970s, during which notable figures like Paul Ekman and Wallace Friesen initiated investigations on the universality of facial expressions and their connection to fundamental emotions. The research conducted by Ekman resulted in the creation of the Facial Action Coding System (FACS), which has gained significant recognition as a prominent instrument utilized for the analysis of facial expressions [1]. This study provided a basis for future research and aided in the creation of multiple databases that contain labeled datasets of facial

DOI: 10.54254/2755-2721/37/20230473

expressions. In recent times, there has been notable progress in the field of facial expression recognition, primarily attributed to the advancements in advanced deep learning models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) [2]. The utilization of facial landmark detection has been widely employed as a preliminary procedure, facilitating the precise recognition of facial expressions. Researchers have demonstrated improved feature extraction and comprehension of facial expressions by localizing face landmarks, including the eyes, nose, and mouth [3]. Researchers have endeavored to enhance the precision and resilience of face expression recognition systems by integrating other sources of information, including verbal speech and physiological data. Multimodal methodologies utilize complementary modalities in order to augment the process of emotion recognition [4].

The objective of this study is to present a thorough examination of the scholarly progression of FER, elucidating the significant achievements, approaches, and obstacles encountered by researchers in this domain. Additionally, suggestions are offered for potential avenues of future investigation.

## 2. Overview of Database

The utilization of FER databases is crucial for researchers and practitioners seeking to develop facial expression recognition systems that are precise, dependable, and resilient. In this section, the focus is on publicly accessible databases that encompass fundamental expressions extensively employed in our review study pertaining to the assessment of deep learning algorithms. In addition, there has been an introduction of recently released databases that offer extensive, diverse, and meticulously annotated datasets. The first one is CK+ [5]. The Cohn-Kanade+ (Ck+) database is widely used for facial expression recognition research. The dataset comprises a compilation of 593 image sequences depicting deliberately staged facial expressions. These sequences were obtained from a total of 123 individuals. The database encompasses a variety of emotions, encompassing anger, contempt, disgust, fear, happiness, sadness, and surprise, along with their respective severity classifications. Each sequence effectively captures the dynamic progression of face expressions, rendering it appropriate for the analysis of temporal changes. The second one is MMI (Machester Movilett Dataset) [6]. The MMI database is a comprehensive dataset for facial expression analysis, including both posed and spontaneous expressions. It covers diverse ethnicities, ages, and genders. The posed subset contains 150 individuals, each displaying six basic facial expressions (happiness, sadness, anger, disgust, surprise, and fear). The spontaneous subset captures dynamic facial expressions during natural interactions, making it valuable for real-world emotion analysis.

The third one is RAF-DB (Radboud Faces Database) [7]. The RAF-DB is a database created for facial expression recognition in real-world scenarios. It consists of 8,040 facial images from various subjects, including professional actors, posing six basic emotions (happiness, sadness, surprise, fear, anger, and disgust). Each image is labeled with emotion labels and rated in terms of valence and arousal, providing rich emotional annotations. The database captures diverse facial expressions under different variations of pose, illumination, and occlusion. The fourth one is FER 2013 [8]. FER2013 is a popular facial expression recognition database collected from the wild. It contains over 35,000 grayscale images categorized into seven emotions (anger, disgust, fear, happiness, sadness, surprise, and neutral). The dataset has a significant level of diversity and has been extensively employed for the purpose of evaluating facial expression recognition algorithms.

The fifth one is AffectNet [9]. AffectNet is a comprehensive database consisting of more than one million facial photos that exhibit a wide range of expressions. The set comprises seven fundamental emotions, alongside neutral and contemptuous terms. The database is capable of capturing nuanced expressions, so enabling a detailed examination of emotions. Furthermore, this technology offers characteristic labels including age, gender, and ethnicity, so allowing researchers to investigate potential relationships between demographic factors and face expressions. The findings indicate that the performance of CK+ surpasses that of other datasets [10].

Saurav S. [11], Shicp [12], Mohan K. [13], Anjani [14], Kumar [15] use some of the database in their research. The accuracy of CK+ is 98.48% [12],97.80% [13],98.85% [14],99.20% [15]. The

DOI: 10.54254/2755-2721/37/20230473

accuracy of MMI is 77.72 [11], 75.00% [14]. The accuracy of RAF-DB is 83.00% [11], 87.34% [12], 81.68% [13], 86.10% [15]. The accuracy of FER 2013 is 75.00% [11], 71.52% [12], 78.90% [13]. AffectNet is not used in their research.

## 3. Method

### *3.1. Pre-Processing*

Pre-processing is an essential procedure in any machine learning endeavors, encompassing not only FER but also other domains. Prior to extracting the pertinent features, it is imperative to ensure the quality of the data. The subsections offer many useful pre-processing techniques.

*3.1.1. Face detection.* The method of face detection is considered a crucial initial step in FER, with the objective of detecting and precisely determining the location of faces inside an image or video frame. Various robust face identification methods, such as Viola-Jones, Histogram of Oriented Gradients (HOG), and Convolutional Neural Networks (CNN), are employed in order to precisely identify and localize regions of interest on the face. These methods guarantee that subsequent FER algorithms are applied exclusively to the pertinent face regions. The algorithm employed is known as AdaBoost, which selectively extracts a limited set of crucial visual elements from a vast pool of potential information. The user's text does not contain any information to rewrite in an academic manner [16-20].

*3.1.2. Image processing.* Real-world scenarios frequently encounter diverse forms of noise, such as sensor noise, motion blur, and fluctuations in illumination. The process of applying a smoothing operation to an image allows for the extraction of pertinent patterns while simultaneously reducing the presence of unwanted noise. In this manner, the process of smoothing might enhance the resilience of the data that is to be subjected to analysis. Lighting conditions might cause facial photos to conceal the necessary facial characteristics needed for precise feature extraction [21-22]. By employing lighting normalizing techniques, such as histogram equalization or algorithms for illumination adjustment. The presence of these variances can pose difficulties in achieving precise feature extraction, as it necessitates the continual alignment of diverse expressions. The utilization of image processing techniques, such as face alignment and normalization, might contribute to the standardization of pictures by aligning facial characteristics in order to facilitate feature extraction (see figure 1) [23].

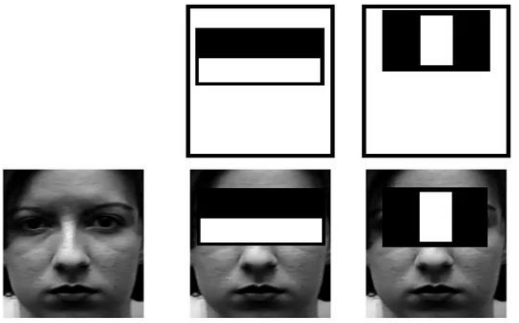

**Figure 1.** Example of smoothing an image [23].

### *3.2. Classification*

Recently, models have been utilizing both traditional machine learning methods and deep learning methods. The integration of two techniques in feature extraction and multimodal recognition has been increasingly prevalent in recent academic study. The preceding subsections outline the various methods discussed in this section.

DOI: 10.54254/2755-2721/37/20230473

*3.2.1. Traditional method.* Initially, the Local Binary Patterns (LBP) algorithm is employed to encode the grayscale texture patterns of facial images. This is accomplished by comparing the intensity values of the facial images with those of their neighboring pixels, as depicted in Figure 2. The binary patterns obtained from the analysis provide valuable insights into the local facial texture information, enabling the differentiation of various facial expressions. LBP-based methodologies provide a straightforward, effective, and resilient means of capturing distinctive texture characteristics for applications related to FER. Figure 2 presents an illustrative illustration of the aforementioned method.

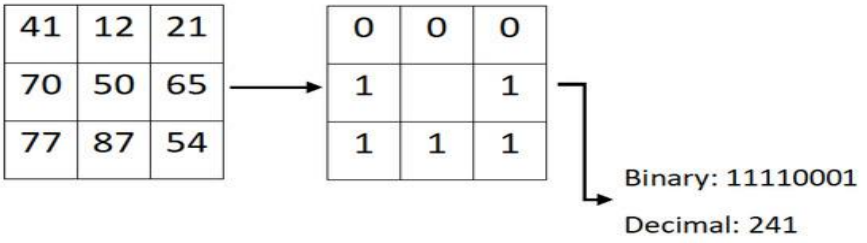


**Figure 2.** Example of encoding the grayscale texture patterns of facial images by comparing their intensity values with their neighbors through LBP [10].

Furthermore, the Local Graph Structure (LGS) approach conceptualizes facial expressions as graphical representations, where the nodes symbolize certain face landmarks or regions, and the edges encode the spatial connections among them. The presented graphs possess the capability to encompass geographical as well as contextual data, hence facilitating the accurate classification of face expressions. LGS-based methodologies have demonstrated favorable outcomes in effectively recording and evaluating patterns of facial expressions [17].

*3.2.2. Deep learning.* Convolutional neural networks (CNNs) are composed of a series of layers, including convolutional, pooling, and fully connected layers. These layers work together to acquire hierarchical representations of face expressions. The authors have demonstrated exceptional proficiency in capturing spatial patterns and have attained cutting-edge outcomes in facial expression recognition challenges [18].

Secondly, Recurrent Neural Networks (RNNs) are able to capture the sequential dependencies among facial expressions by employing recurrent connections. This allows RNNs to develop a context-aware comprehension of the data. Variants such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) have exhibited robust capabilities in capturing temporal dependencies in FER [19].

Thirdly, the utilization of 3D Convolutional Neural Networks (3D CNN) enhances the capabilities of Convolutional Neural Networks (CNN) by including both spatial and temporal information concurrently in the context of video-based FER. By integrating 3D convolutions, these networks effectively capture spatio-temporal patterns present in face expression sequences. According to previous research [20], 3D Convolutional Neural Networks (CNNs) have exhibited superior accuracy in the recognition of facial expressions in video data when compared to conventional CNNs.

In addition, attention mechanisms play a crucial role in allowing models to selectively concentrate on the most pertinent facial regions or frames. Attention-based models have the ability to boost the discriminative capability of the network by dynamically assigning weights to distinct spatial or temporal elements based on their relevance. The utilization of attention processes in FER has demonstrated encouraging outcomes as they guide the model's attention towards significant regions or frames within facial expressions [24].

DOI: 10.54254/2755-2721/37/20230473

## 4. Discussion

### *4.1. Current limitations in the development of recognition*

The initial issue pertains to the fundamental logical challenge in the field of recognition. The association between facial expressions and interior feelings is marked by inherent unreliability, limited precision, and a lack of generalizability. The consistency of facial expression changes in response to external circumstances varies across individuals, particularly among those who have been raised in culturally varied settings. Moreover, it is noteworthy that even within a homogeneous cultural context, individuals belonging to different age groups demonstrate diverse patterns of facial expressions when confronted with similar stimuli. It is imperative to acknowledge that several characteristics, including gender, occupation, and health status, play a significant role [25]. The intricate nature of the correlation between facial expression and emotion mandates the utilization of multi-modal algorithms as opposed to just relying on single-modal methodologies in order to achieve precise computations.

The second option exhibits a deficiency in terms of available datasets. There is a significant scarcity of population data sets that possess specific characteristics, such as age, career, or gender, pertaining to males or females. The datasets mentioned in the article all exhibit the same issue. Extensive gathering efforts are necessary, and ethical considerations must also be taken into account.

### *4.2. Ethic problem and outlook*

FER systems are dependent on the acquisition, examination, and interpretation of facial expressions, hence giving rise to apprehensions over matters of privacy and surveillance. The extensive implementation of FER technology in many settings such as public areas, workplaces, and personal devices has the potential to result in continuous surveillance and monitoring of persons without their awareness or explicit permission. The unauthorized access to personal data. Moreover, the absence of varied and inclusive datasets in the development and evaluation of FER systems might result in a disproportionate misinterpretation of expressions, hence perpetuating discrimination against specific demographics, including marginalized groups and persons exhibiting unusual expressions.

The research does not include the pipeline of multi-modal FER. However, this research field is of great significance in addressing the fundamental logical challenges in FER recognition. Therefore, future research will prioritize investigating this topic, specifically exploring multi-modal approaches such as image-audio and image-text-audio FER. The logical relationship between frames in a video will also be taken into consideration.

FER exhibits significant promise across many areas, promising a prosperous trajectory in the foreseeable future. The utilization of multimodal and real-time FER is expected to gain significant popularity in the future. The integration of multimodal and real-time FER in the healthcare industry is expected to bring about significant advancements in the field of mental health diagnosis and treatment. FER systems have the capability to effectively identify real-time facial expressions and other relevant data, such as voice, which can be utilized by healthcare professionals to discern emotional states and deliver tailored care and interventions. Additionally, the construction of additional databases specific to precise professions or age groups will be undertaken in order to mitigate the issue of overfitting in models.

## 5. Conclusion

FER has garnered considerable interest and demonstrated encouraging outcomes across several fields, such as psychology, human studies, and healthcare. The utilization of computer vision algorithms to automatically detect and interpret facial expressions has significant promise in enhancing human-machine interaction and comprehending human emotions. The study concludes that CK+ dataset exhibits higher accuracy compared to other datasets in the context of FER. Additionally, the article presents a widely recognized method within the field. In conclusion, the present state of FER research necessitates a thorough examination of the limitations and ethical concerns associated with this field.

DOI: 10.54254/2755-2721/37/20230473

## References


[1] Paul Ekman [Internet]. Wikipedia. Available from: https://en.wikipedia.org/wiki/Paul_Ekman

[2] Lopes, A.T., Teixeira, F., & Bernardino, A. (2021). Deep Learning Approaches for Automatic Facial Expression Recognition: An Overview. Sensors, 21(2), 530.

[3] Li, X., Lv, Z., et al. (2020). Deep Learning of Facial Landmarks for Robust Expression Analysis: A Survey. IEEE Transactions on Affective Computing, 11(5), 764-782.

[4] Kaya, H., Gürpinar, A., & Öz, A. (2020). Multimodal Emotion Recognition: State of the Art and Challenges. IEEE Access, 8, 186276-186293

[5] Lucey, P., et al. (2010). The extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression

[6] Valstar, M., et al. (2010). The first facial expression recognition and analysis challenge.

[7] Langner, O., et al. (2010). Presentation and validation of the Radboud Faces Database.

[8] Goodfellow, I. J., et al. (2013). Challenging common assumptions in the unsupervised learning of disentangled representations.

[9] Mollahosseini, A., et al. (2017). AffectNet: A database for facial expression, valence, and arousal computing in the wild.

[10] Facial Expression Recognition Using Computer Vision: A Systematic Review [Internet]. MDPI. Available from: https://www.mdpi.com/2076-3417/9/21/4678.

[11] SAURAV S,SAINI R,SINGH S.EmNet:a deep integrated convolutional neural network for facial emotion recognition in the wild [J]. Applied Intelligence, 2021 (16).

[12] SHI C P,TAN C,WANG L G.A facial expression recognition method based on a multibranch cross-connection convolutional neuralnetwork [J]. IEEE Access,2021 ,09:39255 -39274.

[13] MOHAN K,SEAL A,KREJCAR O, et al. FER-net:facial expression recognition using deep neural net[J].Neural Computing and Ap-plications,2021 (99): 1 -1

[14] ANJANI SUPUTRI DEVI D, SATYANARAYANA C.An efficient facial emotion recognition system using novel deep learning neuralnetwork-regression activation classifier [J].Multimedia Tools and Applications, 2021, 80(1 2):1 7543 -1 7568.

[15] KUMAR R J R,Sundaram S,ARUMUGA N,et al.Face feature extraction for emotion recognition using statistical parameters from sub-band selective multilevel stationary biorthogonal wavelet transform[J]. Soft Computing, 2021, 25 (7): 5483 -5501.

[16] Ahonen, T., Hadid, A., & Pietikäinen, M. (2004). Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(12), 2037-2041.

[17] Zhao, G., et al. (2010). Facial expression recognition from near-infrared videos. Human-Computer Interaction. Recognition. Springer, London, 706-715.

[18] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

[19] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

[20] Tran, D., et al. (2015). Learning spatiotemporal features with 3D convolutional networks. Proceedings of the IEEE International Conference on Computer Vision, 4489-4497

[21] Zhang, H., et al. (2020). Attention-based Convolutional Neural Network for Facial Expression Recognition. Applied Sciences, 10(8), 2817.

[22] Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1, I-511.

[23] Canedo, Daniel, and António JR Neves. "Facial expression recognition using computer vision: A systematic review." Applied Sciences 9.21 (2019): 4678.

[24] S. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. action unit and emotion-specified expression," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, 2010, pp. 94.

DOI: 10.54254/2755-2721/37/20230473

[25] COLONNELL V, CARNEVALI L, RUSSO P M, et al. Reduced recognition of facial emotional expressions in global burnout and burn-out depersonalization in healthcare providers [J].PeerJ, 2021, 09: 1 061 0 -1 061.