

# Exploring the application of synchronous analysis of physiological signals and facial expressions in emotion recognition

**Langxuan Wang**

School of International Education, Guangdong University of Technology, Guangzhou, Guangdong Province, 510006, China

3233762603@qq.com

**Abstract.** In today's increasingly frequent human-computer interaction, the accurate recognition and response of intelligent systems to human emotions has become the key to improving the user experience. This paper aims to explore the application of physiological signals and facial expression synchronization analysis in emotion recognition. This paper adopts the method of literature review, and summarizes the acquisition techniques of physiological signals and facial expressions, synchronous analysis methods, and their application cases in emotion recognition. At the same time, the limitations and challenges of the existing technology are discussed. The review shows that the synchronous analysis of physiological signals and facial expressions has important application value in the field of emotion recognition, especially in mental health monitoring, educational feedback, and customer service automation. However, the practical application of the technology still faces many challenges, including the real-time nature of data collection, the adaptability of individual differences, and other identification problems. Therefore, future research needs to further optimize data acquisition technology and develop more accurate and personalized analysis algorithms.

**Keywords:** Physiological Signals, Facial Expression Recognition, Emotion Recognition, Multimodal Analysis, Deep Learning.

## 1. Introduction

In the wave of artificial intelligence, emotion recognition technology is gradually becoming a bridge between humans and computers, and its research and development status has received extensive attention from the academic community. In the field of emotion recognition technology, the existing research angles mainly focus on improving the accuracy of facial expression recognition, and explore how to integrate facial expressions and physiological signals through multimodal methods to enhance the comprehensiveness of emotion recognition[1]. In this research context, this study puts forward new insights into this field by carefully analyzing the existing achievements of scholars in the synchronous analysis of facial expressions and physiological signals. Facial expression recognition technology initially relied on a rule-based approach to recognition through predefined facial features. With the development of technology, machine learning methods have begun to be applied, using statistical models to learn expression patterns from data. In recent years, the introduction of deep learning

technologies, especially convolutional neural networks (CNNs) and recurrent neural networks (RNNs), has greatly improved the accuracy and robustness of facial expression recognition. CNNs excel at extracting facial features, while RNNs are able to efficiently process the temporal information of expressions[2]. Based on existing literature and data analysis, this paper will discuss the concept of multimodal emotion recognition, evaluate facial expression recognition techniques and physiological signal analysis methods, and their integration and application in emotion recognition. The practical application of multimodal emotion recognition technology will have a positive impact on many aspects of society, including improving the customer service experience, enhancing public safety monitoring, and promoting mental health management[3]. The significance of this study lies in the fact that it can not only provide new perspectives for emotion recognition theory, but also improve the naturalness of human-computer interaction and the accuracy of mental health assessment in practical applications.

## **2. The integration and development of facial expression recognition and physiological signal analysis technology**

Facial expression recognition and physiological signal analysis are the two pillars of the field of emotion recognition. The development of facial expression recognition technology has undergone a transformation from traditional methods to deep learning, and the collection and analysis of physiological signals provide intrinsic physiological indicators for emotional states.

Physiological signals, including heart rate, skin conductance, and brain waves, provide important complementary information for emotion recognition. Modern acquisition techniques are able to capture these physiological indicators in real-time and provide high-quality data for sentiment analysis through preprocessing and feature extraction techniques such as signal denoising and feature selection. The analysis of physiological signals not only enhances the objectivity of emotion recognition, but also improves the sensitivity of the system to different emotional states.

By integrating facial expression recognition and physiological signal analysis, researchers can build a more comprehensive emotion recognition system. The combination of the intuitiveness of facial expressions and the objectivity of physiological signals provides a richer source of data for understanding human emotions. This multimodal approach not only improves the accuracy of emotion recognition, but also opens up new possibilities for the application of emotion recognition technology in the fields of human-computer interaction, mental health assessment, and security monitoring.

## **3. Synchronous analysis of facial expressions and physiological signals**

### *3.1. Theoretical basis of simultaneous analysis*

Simultaneous analysis of facial expressions and physiological signals relies on an in-depth understanding of time-series data. Time series analysis is a statistical method used to analyze data points at a point in time to extract meaningful patterns. In facial expression and physiological signal analysis, time series analysis can help us identify and understand changes in affective states over time. Multimodal data fusion is another key concept in simultaneous analysis. It involves combining data from different sources, such as video images and physiological sensor data, to obtain more comprehensive and accurate analytical results. Data fusion technology can improve the accuracy and reliability of emotion recognition.

### *3.2. Application cases of synchronous analysis in emotion recognition*

Synchronous analysis of facial expressions and physiological signals plays an important role in the field of emotion recognition, especially in the three key areas of mental health monitoring, educational feedback, and customer service automation.

In terms of mental health monitoring, this technology is able to monitor psychological states such as stress, anxiety, and depression in real time by analyzing individuals' facial expressions and physiological signals and using machine learning algorithms to identify abnormal changes in

emotional states so as to provide necessary interventions in a timely manner. Table 1 presents data on examples of the use of facial expression and physiological data in depression monitoring research:

**Table 1.** The use of facial expression and physiological data in depression monitoring research

Case studies	Research methods	Data source	Findings
Research on depression recognition based on facial expression analysis	Convolutional Neural Network (CNN)	Chinese depression database dataset EATD	The accuracy rate was 71.3%, and the proportion of facial expressions recognized was consistent with the situation of depressed patients [4].
Advances in the application of machine learning in the study of facial features in patients with depression	Machine learning	IEEE Xplore database	Facial expressions can present important nonverbal information that can aid in the identification and diagnosis of depression [5].
Chinese scholars have made progress in multimodal datasets for mental disorders and related research	Multimodal psychophysiological database MODMA	MODMA database	The dataset includes EEG experimental data and voice experimental data, which provides a new way to quantify the assessment of mental disorders [6].

Educational feedback is a method used by teachers to understand students' learning and feelings, and it helps teachers tailor teaching to the needs of students to improve learning outcomes. Modern educational technology uses biometric tools, such as facial recognition and heart rate monitoring, to analyze students' facial expressions and physiological responses, to assess teaching effectiveness and student engagement, and then adjust teaching strategies to improve learning efficiency and satisfaction. In the field of customer service automation, synchronous analysis technology can automatically monitor the emotional state of customers, realize personalized services, optimize service processes by analyzing customer emotional reactions, and improve customer satisfaction and loyalty. These applications not only improve the accuracy and efficiency of emotion recognition, but also provide new possibilities for innovation and development in related fields.

#### 4. Construction and optimization of multimodal emotion recognition model

##### 4.1. Multimodal data fusion strategy

The multimodal data fusion strategy is a key technology to construct an effective emotion recognition model. In the early fusion strategy, facial expressions and physiological signals are integrated in the data preprocessing stage, which helps to reduce the data dimension and strengthen the correlation between features. The late fusion strategy is to synthesize information from different data sources at the decision-making stage of the model, allowing each modality to contribute its information independently, which may improve the generalization ability of the model [7]. Table 2 illustrates the application of multimodal data fusion strategies in the field of emotion recognition:

**Table 2.** The application of multimodal data fusion strategies in the field of emotion recognition

Application examples	Convergence strategy	Description	Advantage	Challenge
Multimodal Emotion Recognition Challenge(MER 24)	Late fusion	Integrate information from different modalities at the decision-making stage	Allows independent contribution of information and improves generalization ability	There is a need for effective methods to synthesize decision-making outcomes of different modalities [8].
SkipcrossNets	Innovate the convergence strategy	Adaptively identify the best modal combinations at different stages	Increase feature interaction and improve the fusion effect	It is necessary to learn cross-modal learnable factors, which may increase the computational complexity [9].
Early fusion Strategies	Early fusion	Integrate facial expressions and physiological signals in the data preprocessing phase	Reduce data dimensions and strengthen feature correlation	Some modal-specific information may be lost.
Late fusion Strategies	Late fusion	Synthesize information of different modalities at the model decision layer	Maintaining data independence may improve accuracy.	The problem of time alignment of different modalities needs to be addressed.

#### 4.2. Application of deep learning models

Deep learning models play a crucial role in the field of multimodal emotion recognition. CNN is widely used in facial expression analysis due to its powerful feature extraction ability, and by learning facial key points and expression patterns, CNN can achieve high-accuracy classification on standard expression databases such as FER-2013. For example, a study using CNNs for feature extraction of facial expressions achieved an accuracy rate of more than 70% on the FER-2013 database [10]. For time-series data of physiological signals, recurrent neural networks (RNNs) and long short-term memory networks (LSTMs) have shown advantages in processing such data. LSTMs are particularly suitable for analyzing physiological indicators such as heart rate variability (HRV) to identify an individual's emotional state due to their ability to capture long-term dependencies. In one study, the LSTM network modeled HRV data and successfully distinguished between heart rate patterns in calm and stressful states with an accuracy of more than 85% [11]. Effective model training and validation is key to ensuring the accuracy and robustness of deep learning models. This often requires a large amount of annotated data and a well-designed experimental protocol. For example, the SEMAINE database provides a wealth of facial video, voice recording, and physiological sensor data, providing a valuable resource for multimodal emotion recognition research [12]. In addition, by using data augmentation techniques and cross-validation methods, researchers are able to improve the model's ability to generalize to new data and reduce the risk of overfitting [13].

## 5. Discussion

Synchronous analysis technology faces the complexity of processing multi-source data in emotion recognition. The synchronization of facial expressions and physiological signals requires precise time alignment and phase matching, which is often difficult to achieve in practical applications [14]. In addition, differences in sampling rates and temporal resolution from different data sources create additional challenges for simultaneous analysis. In view of the challenges of synchronous analysis technology in emotion recognition, researchers have carried out a series of research attempts. For example, a cross-subject emotion EEG recognition method based on multi-source domain adaptation has been proposed, which solves the problem of inter-sample differences caused by differences between subjects through multi-source domain transfer learning, and the experimental results show that this method can effectively improve the accuracy of cross-subject emotion EEG recognition [15]. In addition, a 3D hierarchical convolutional fusion model is used to process multimodal physiological signal emotion recognition, which can fully explore the interaction between multiple modalities and describe sentiment information more accurately, and the experimental results show that the accuracy rate is as high as 98% on the DEAP dataset [16]. Although deep learning models perform well in emotion recognition, their generalization ability is still a key issue. Models tend to perform well on training data, but their performance degrades when faced with new, unseen data [17]. Improving the generalization ability of the model so that it can adapt to different environments and individual differences is an important direction of current research.

Emotion recognition systems need to process data from different individuals and different cultural backgrounds. The diversity of data presents difficulties in feature extraction and model training. How to design algorithms that can adapt to different data features is an important issue in the development of emotion recognition technology. In the field of emotion recognition, researchers are faced with the challenges posed by data diversity, including individual differences, cultural diversity, and the subjectivity of emotional expression. For example, different individuals may express the same emotional state in different ways, and people from different cultures may understand and express the same emotion differently. In addition, the collection of affective data often relies on subjective self-reporting, which can introduce bias. To address these challenges, more sophisticated algorithms are being developed that can extract richer and more reliable emotional features from multimodal data, such as combining physiological signals, facial expressions, speech, and text. At the same time, it needs to explore how to use big data and machine learning techniques to process and analyze these complex datasets, and how to improve the generalization ability of models through cross-cultural research. The accuracy of emotion recognition is highly dependent on the quality of the data. Noise during data acquisition, incomplete data recording, and inconsistent data annotations can all affect the performance of the model. Therefore, establishing a strict data quality control process is the key to improving the reliability of the emotion recognition system.

To further improve the performance of emotion recognition systems, future research needs to be more in-depth exploration in algorithm design, data processing, and application practice. In addition, interdisciplinary collaboration will be key to advancing the development of emotion recognition technology, and integrating research results in fields such as psychology, cognitive science, and data science will help achieve deeper human-computer interaction and broader social applications.

## 6. Conclusion

This paper reviews the application and progress of synchronous analysis of physiological signals and facial expressions in the field of emotion recognition, and clarifies the important role of multimodal emotion recognition technology in improving the naturalness of human-computer interaction and the accuracy of mental health assessment. The integration of facial expression recognition technology and physiological signal analysis has significantly improved the accuracy and robustness of emotion recognition through advanced technologies such as deep learning. However, in research and writing, the scope of research is mainly limited to the published literature and may not cover the latest research results. Because the research methodology relies on literature reviews rather than empirical data, there

is not much experimental validation, which may limit the in-depth evaluation of technical performance. Although the application of multimodal emotion recognition is discussed, the prediction of future challenges may not be comprehensive. Future research should focus on algorithm innovation to improve the accuracy of simultaneous analysis and the generalization of models. At the same time, data diversity and quality control should be strengthened to ensure the reliability of the emotion recognition system and the privacy and security of user data.

## References

- [1] Wang, Shuai, et al. Multimodal Emotion Recognition From EEG Signals and Facial Expressions[J]. IEEE ACCESS, 2023, 11:33061-33068.
- [2] Yan Zhao et al. Attention-Based CNN Fusion Model for Emotion Recognition During Walking Using Discrete Wavelet Transform on EEG and Inertial Signals[J]. BIG DATA MINING AND ANALYTICS, 2024, 7(1):188-204.
- [3] Chen Xinyi, Tao Xiaomei. Emotion recognition method based on multimodal physiological signal feature fusion[J]. Computer Integrated Manufacturing Systems, 2023, 40(06):175-181+186.
- [4] Chen Kunlin, Hu Defeng, Chen Nannan. Research on Depression Recognition Based on Facial Expression Analysis[J]. Computer Era, 2023(10):70-74.
- [5] Xin LI, Qing FAN. Application progress of machine learning in the study of facial features of patients with depression[J]. JOURNAL OF SHANGHAI JIAOTONG UNIVERSITY (MEDICAL SCIENCE), 2022, 42(1): 124-129.
- [6] Cai, H., Yuan, Z., Gao, Y. et al. A multi-modal open dataset for mental-disorder analysis. Sci Data 9, 178 (2022).
- [7] Bei Pan et al. A review of multimodal emotion recognition from datasets, preprocessing, features, and fusion methods, Neurocomputing, Volume 561, 2023, 126866, ISSN 0925-2312
- [8] Zhijing Xu, Yang Gao. Research on cross-modal emotion recognition based on multi-layer semantic fusion[J]. MATHEMATICAL BIOSCIENCES AND ENGINEERING, 2024, 21(2).
- [9] Xinyu Zhang et al. SkipcrossNets: Adaptive Skip-cross Fusion for Road Detection[Z]. arxiv, 2023.
- [10] Syed Muhammad Daniyal et al. An Improved Face Recognition Method Based on Convolutional Neural Network[J]. JISR ON COMPUTING, 2024, 22(1).
- [11] Miaohua Zhang et al. "Robust tensor factorization using maximum correntropy criterion," 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 2016, pp. 4184-4189, doi: 10.1109/ICPR.2016.7900290.
- [12] Wen Peiyu et al. Research and Application of Multimodality in Emotion Recognition[Z]. Applied Science and Technology, 2024,51(1):51-58.
- [13] Sun Hao et al. Robustness evaluation of multi-source remote sensing image depth recognition model against attacks[J]. Journal of Remote Sensing, 2023, 27(8):1951-1963.
- [14] Li Yuchi, Li Haifang, Jie Dan, et al. Phase synchronization analysis of emotional EEG based on complex networks[J]. Computer Engineering and Applications, 2017, 53(18):230-235.
- [15] GAO Hanbing. Cross-subject emotional EEG recognition based on multi-source domain adaptation. Chinese Journal of Intelligent Science and Technology[J], 2021, 3(1): 59-64
- [16] LING Wenfen. Multi-modal physiological signal emotion recognition based on 3D hierarchical convolution fusion. Chinese Journal of Intelligent Science and Technology[J], 2021, 3(1): 76-84. doi:10.11959/j.issn.2096-6652.202108
- [17] Ge Liangzhu. Research on Generalization Performance and Model Selection of Deep Learning [D]. Tianjin University, 2019