Assistance methods and analysis based on facial expression recognition for people with disabilities

Haoming Zhu

Institute of Wollongong, Hong Kong, China

Zhuhaoming22@outlook.com

Abstract. In recent years, facial recognition technologies have demonstrated significant potential for assisting people with disabilities. How to accurately predict facial recognition for people with disabilities has always been a research hotspot. This paper reviews the current development of assistance methods based on facial expression recognition for people with disabilities. By summarizing relevant technological principles, application scenarios, and possible challenges in the future, this paper explores the possibilities and limitations of facial expression recognition technologies. It comprehensively analyzes relevant literature and summarizes design concepts for facial expression recognition technologies and assistance systems tailored to people with disabilities, as well as physiological differences in certain special groups. Furthermore, it examines the optimization and future directions for addressing scenario adaptation issues of facial recognition from the perspectives of current cutting-edge technologies and research, with the ultimate aim of improving the living experience of people with disabilities. It also provides a reference for enhancing the efficiency and feedback of assistance efforts.

Keywords: people with disabilities, facial expression recognition, assistance methods, AIMcT, facial expression

1. Introduction

People with disabilities face excessive obstacles in daily communication and life, posing a demand for alternative means to convey information. Among them, facial expression is one of the most important ones. Presently, artificial intelligence (AI) technologies have been developing rapidly, and facial expression recognition (FER) technologies offer new avenues for providing assistance to people in need. This type of technology can convert facial expressions into commands or signals, thereby helping people with disabilities to effectively convey their needs of the moment and receive help from people around in time. However, current FER technologies are primarily designed for able-bodied individuals and show limitations when applied to people with disabilities. Due to physiological characteristics and the specificity of usage scenarios, people with disabilities have higher requirements for this type of technology. Therefore, it is of great practical significance to study the assistance methods based on FER for people with disabilities. A review of relevant studies reveals a variety of design ideas in this field. Jain et al. designed the hands-free mouse system with a focus on the physical computer-interaction needs of people with disabilities [1]. They used the MediaPipe Face Mesh model to detect 468 facial key points, and mapped the user's nose to control cursor movement, while actions such as blinking or opening the mouth were used to perform mouse clicks and scrolling [1]. This system combines the real-time capture of OpenCV with simulated interaction of PyAutoGUI and achieves contactless interaction, showing marked practicality in everyday life and enhanced life experience. However, it relies on lighting conditions, and head posture may also affect the recognition of complex facial movements. Therefore, the issue of adaptability for people with disabilities who have diverse physiological characteristics should be further addressed. Kawaguchi et al. put forward a facial gesture recognition method based on electroencephalography (EEG) and self-organizing maps (SOM), which extracted the power of the α, β, and θ wave bands from EEG signals as features and used the SOM-Hebb classifier to perform facial gesture classification [2]. This method employs the EEG artifact signals generated by facial muscle movements to define various facial gestures, such as opening the eyes, closing the eyes, and clenching the teeth. It features a high accuracy rate, but its recognition speed needs improvement. The current speed is relatively slow at 5.7 seconds per recognition, and accuracy tends to decrease when recognizing multiple gestures, highlighting the necessary need to improve the recognition accuracy. In addition, the lack of real-time responsiveness needs to be addressed since even a one-second delay could be fatal for people with disabilities in dangerous situations. Furthermore, the scalability of the method needs to be enhanced. Wu et al. proposed the AI-enhanced Morse code translation system (AIMcT) based on images, which could generate Morse code via facial movement recognition,

with similar recognition accuracy between people with disabilities and those without [3]. The system operates contactlessly, requiring no physical interaction with the device for users with disabilities [3]. However, AIMcT has some issues that need to be addressed. If the person with a disability being scanned has a special condition (for example, having their lower jaw removed due to an illness), one facial recognition feature will be missing, which would reduce the success rate of the scan. Additionally, the system's adaptability to certain situations is limited. For instance, recognition may be problematic if there is significant movement or shaking during the scan. In research focused on the differences within special populations, Chen's group studied facial expressions of children with attention deficit and hyperactivity disorder (ADHD) [4]. The research team examined the differences in facial expressions between children with ADHD and children without ADHD and found that children with ADHD exhibit different behaviors depending on their state. For instance, when they are angry, their eyebrow movements are less pronounced than those of typical children, and when excited, they blink more frequently than their peers. However, the sample size of the research is relatively small, and a larger sample size is necessary for further research. Therefore, the conclusions cannot be considered definitive currently. The following sections of this paper present the practical applications of the aforementioned technologies in assisting people with disabilities, compare their recognition accuracy, real-time performance, and adaptability to different scenarios, and analyze the optimization direction of technologies targeting various physiological characteristics of people with disabilities. Subsequently, it discusses FER technologies for people with disabilities, design concepts for assistance systems, and physiological differences in special populations. By integrating current cutting-edge technologies, this study explores future development directions and provides insights to enhance the living experience and rescue efficiency for people with disabilities.

2. Facial expression recognition technologies for people with disabilities

This section discusses the FER technologies for people with disabilities. Under normal conditions, FER involves capturing subtle changes in a person's facial expressions and then translating these movements into information or commands that can be understood by the receiver. Due to various factors, people with disabilities may have limited physical mobility and, in some cases, are unable to freely control their facial muscles. Current technologies mainly revolve around three key areas: facial feature extraction, model training, and interaction, such as the multi-feature point fusion technology [5]. The following discussion will introduce the basic principles of the technology and analyze its advantages and disadvantages.

2.1. The multi-feature point fusion technology

By detecting positional changes in three facial feature points, namely eyebrows, eyes, and mouth, a facial expression feature vector is constructed, which is then used with classification algorithms to recognize user commands. For example, Dillague et al. proposed a method that combines these three feature points [6]. Compared to conventional recognition technologies that only involve eyes and mouth, this technology is more comprehensive and can recognize expression details with higher accuracy. This technology adopts the Viola-Jones algorithm for face detection. Combined with the Dlib library, it divides the face into 68 facial feature points for extraction and focuses on recognizing movements in the eyebrows, eyes, and mouth areas. By calculating the Euclidean distances between feature points, namely the degree of mouth opening or the tilt angle of the eyebrows, it quantifies various expression changes. For example, the distance between the upper and lower lip feature points is used to determine the mouth's open or closed state. Integrating with Artificial Neural Network (ANN) and Backpropagation Algorithm for model training, it can achieve classification of different facial expressions such as happiness and fear. One of the advantages of AIMcT is that it enhances recognition robustness, which can be used for people with disabilities, such as patients with facial paralysis. Meanwhile, quantifying the distances between feature points makes the recognition process interpretable and adaptable to different users' physiological characteristics. It also features relatively low computational complexity, supports real-time processing, and can be used with household devices. However, it demonstrates several shortcomings. For example, if facial images are unclear or under conditions of strong or low lighting and head tilt, the recognition performance declines. The positioning of feature points is relatively fixed. If the face is deformed, there is a risk of misidentification. It also suffers from a lack of generalizability since separate training is needed for different types of disabilities, such as facial paralysis, resulting in higher costs due to the requirement for more training.

2.2. Action unit analysis technology

Based on the Facial Action Coding System (FACS), action unit analysis technology can break down expressions into 42 individual action units (AUs) and recognize complex expressions by detecting combinations of these AUs [7]. According to Kawaguchi et al. this technology is suitable for FER in populations with special disabilities [2]. It uses convolutional neural networks (CNN) to automatically detect the activation intensity of AUs, captures their changes with the aid of temporal analysis to distinguish between fleeting expressions and sustained states, and supports visual outputs. It can be used to interpret subtle or

incomplete expressions, suitable for those who have weaker control over their facial muscles. Meanwhile, it outputs AU intensity values rather than single emotion labels, providing flexible data and enabling personalized features. Additionally, with this technology, the cost of assistive devices for people with disabilities can be further decreased. The major shortcoming of this technology is that it requires a substantial amount of FACS-labeled data for model training, while the expression data of people with disabilities is insufficient. Then, the recognition accuracy for complex AU combinations is relatively low, as a single expression may involve multiple AUs, which can lead to misidentification. Moreover, the high computational load makes it unsuitable for low-end embedded devices.

2.3. Extraction based on facial feature point

Wu et al. suggested converting facial movements into Morse code, then translating them into text and commands to enable contactless communication [3]. The principle is to use OpenCV for automatic calibration and lighting compensation to ensure accurate extraction of facial features. Similar to the method described in section 2.1, it also uses the Dlib library to track mouth feature points, determining "open" and "closed" mouth states based on changes in the distance between two points, which are then interpreted as "dots" and "dashes." By incorporating a fuzzy time recognition algorithm, it automatically adjusts the decision threshold. Finally, the detected actions are matched to the international Morse code table and converted into commands composed of letters or Arabic numerals to control peripherals such as a computer mouse. Its operation is contactless, meaning that no sensors are required, and people with disabilities do not need additional training to use it. The adaptive ambiguity processing algorithm allows it to adapt to irregular movements, accommodating the user experience of certain special patients, offering greater fault tolerance. Moreover, its costs are lower. This technology also demonstrates several drawbacks. It relies on a single type of movement. If the user is unable to perform certain mouth actions, the number of recognition cues decreases. Even if eye actions and other movements can be performed, the recognition accuracy will still decline. In the meantime, the Morse code transmission speed is 10 characters per minute, making smooth communication difficult. Therefore, the encoding speed needs to be improved. Significant camera shake can also lead to misidentification. This highlights the need to fix the device's position, indicating its inability to adapt to certain environments.

2.4. Multimodal fusion recognition

Single-modality recognition has limitations. Deshmukh et al. proposed to integrate multimodal fusion recognition with facial expressions and other cues to enhance the robustness of recognition [8]. It achieves facial identity authentication using the Face Recognition library, tracks gestures and maps them to mouse actions using MediaPipe, then executes complex instructions combined with voice commands, thereby developing a "face + gesture + voice" three-modal interaction system [9]. It allows different modalities to complement each other and improve fault tolerance. However, the accuracy of the voice module decreases in noisy environments, gesture recognition is somewhat dependent on lighting conditions, and it also requires hardware investment. In addition, Bian et al. (2024) proposed multimodal large language models, providing new insights for understanding the deeper semantics of complex expressions [10].

3. Practical applications

FER technologies are widely applied in real-world scenarios, covering various fields. In the medical field, there are also some examples of its use in assisting people with disabilities. For instance, the hands-free mouse system proposed by Jain et al. [1] can track the facial key points of users in real-time and use facial organs as control peripherals. The cursor is controlled through nose movements, while peripheral operations are carried out using eye and mouth actions within this system, making it extremely suitable for people with disabilities in hospitals, rehabilitation centers, and other similar facilities. There is also a translation system that generates Morse code based on facial movements, possessing the potential to become the mainstream assistive technology used by people with disabilities, with its high recognition accuracy [3]. FER technologies can also assist young patients with ADHD, enabling early detection of their condition and timely intervention and treatment [4]. They can also be used to verify patient information and prevent impersonation. At Zhongshan Hospital Xiamen University, facial recognition is employed to ensure accurate patient identification. Moreover, these technologies can be used in mental health to promptly assess a patient's facial expressions, helping to prevent self-harm or suicidal behavior. They can also be used in other scenarios, such as attendance systems for students with disabilities [11], enabling them to check in smoothly.

4. Limitations

There are several challenges for FER technologies. The first is the environmental adaptability. Lighting, background, and facial occlusions can all reduce recognition accuracy; in strong backlighting, the recognition rate of features like the mouth or

eyebrows decreases when using multi-feature point fusion [5,6]. The second is the network condition, which may affect the process speed, ultimately leading to delayed patient treatment [2]. This poses a challenge for real-time performance improvement. The third is the weak recognition ability for non-frontal poses; tilting, turning the head sideways, or shaking the head all increase the error rate [11]. Besides, FER technologies rely heavily on data; currently, there is insufficient data collected for facial expression recognition in people with disabilities, which has resulted in limited generalizability that needs improvement [7].

5. Future prospects

The prospects for FER technology development mainly involve technological optimization and scenario adaptability. Technology-wise, it should focus on multimodal fusion, integrating voice and physiological features such as hand gestures tracked by MediaPipe in addition to facial expressions. This will improve recognition accuracy and the success rate of command recognition in noisy environments. The response time of FER technologies should be improved to increase the success rate of emergency assistance for people with disabilities, reducing it to two seconds, one second, or even less. Meanwhile, emphasis needs to be placed on adapting technologies to different scenarios, so that people with disabilities can still use the related functions in environments with significant shaking, such as on buses. Lighting compensation factors should be optimized so that people with disabilities can be recognized accurately even in areas with strong or low brightness. The database of facial recognition should be improved to not only accommodate different types of disabilities but also enhance facial recognition from various angles, which will enhance the recognition accuracy when users look down, turn their heads, or move their heads.

6. Conclusion

The aforementioned applications of FER technologies in assisting people with disabilities have provided multiple solutions to difficulties encountered in various everyday situations. By reviewing some of the existing literature, it is found that previous studies focus on recognizing single facial expressions, while recent studies are shifting towards multimodal fusion and personalized adaptation. Centered on the recognition of facial key points, EEG-assisted analysis, as well as Morse code translation and conversion, current research aims to significantly improve assistance for people with disabilities in real-life scenarios. From the perspective of technology, FER technologies for people with disabilities have become more precise and diverse. For instance, three-point extraction and a dynamic tracking algorithm are used to capture the user's subtle facial expressions, enabling contactless communication for people with disabilities. Furthermore, the multimodal fusion technology can effectively address the limitations of a single modality. However, existing technologies present significant limitations: scenario adaptability and real-time performance need further improvement; while multimodal fusion enhances accuracy, it also increases computational costs; coverage of special groups is incomplete, and research data on patients in extreme conditions is relatively scarce. Therefore, it is important to address these challenges for further optimization, including improving response speed and scenario adaptability, and conducting more research on special groups with more updated and mature approaches for improvement. In summary, although FER technologies have demonstrated evident potential, their applications under the current environment require more advanced technology innovation and development. By continuously optimizing algorithms, diversifying application scenarios, and considering minority groups, these technologies can better assist people with disabilities and ultimately improve quality of life.

References

[1] Jain, R., Bhavana, M. N., Kale, A., Rastogi, S., & Thakur, P. (2025). Hands free mouse using facial expression for physically disabled people. In *2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL)* (pp. 1097–1101). https: //doi.org/10.1109/ICSADL65848.2025.10933200

[2] Kawaguchi, T., Ono, K., & Hikawa, H. (2024). Electroencephalogram-Based Facial Gesture Recognition Using Self-Organizing Map. *Sensors*, 24(9), 2741. https: //doi.org/10.3390/s24092741

[3] Wu, C.-M., Chen, Y.-J., Chen, S.-C., & Zheng, S.-F. (2023). Creating an AI-Enhanced Morse Code Translation System Based on Images for People with Severe Disabilities. *Bioengineering*, 10(11), 1281. https: //doi.org/10.3390/bioengineering10111281

[4] Chen, Y., Ma, Y., Fan, X., Lv, J., & Yang, R. (2024). Facial expression recognition ability and its neuropsychological mechanisms in children with attention deficit and hyperactive disorder. *Journal of Zhejiang University (Medical Sciences)*, 53(2), 254 - 260. https: //doi.org/10.3724/zdxbyxb-2023-0390

[5] An, Y., Lee, J., Bak, E., Pan, S. (2023). Deep Facial Emotion Recognition Using Local Features Based on Facial Landmarks for Security System. *Computers, Materials & Continua*, 76(2), 1817–1832. https: //doi.org/10.32604/cmc.2023.039460

[6] Dillague, J. D. O., Juico, J. H. A., Yu, N. S. L., & de Goma, J. C. (2024). Detection of facial expressions based on three feature points using image processing with artificial neural networks. In *2024 5th International Conference on Industrial Engineering and Artificial Intelligence (IEAI)* (pp. 29–33). https: //doi.org/10.1109/IEAI62569.2024.00014

[7] Wyman, A., & Zhang, Z. (2025). A Tutorial on the Use of Artificial Intelligence Tools for Facial Emotion Recognition in R. *Multivariate Behavioral Research*, 60(3), 641–655. https: //doi.org/10.1080/00273171.2025.2455497

[8] Deshmukh, M., Pothawale, A., Bedre, P., & Dange, A. (2024). Facial recognition and gesture-based PC interface with voice command integration. In *2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS)* (pp. 842–847). https: //doi.org/10.1109/ICUIS64676.2024.10866854

[9] Zhou, Y., Liang, Y., & Tan, P. (2023). Design of an Intelligent Laboratory Facial Recognition System Based on Expression Keypoint Extraction. *IEEE Access*, 11, 129805–129817. https: //doi.org/10.1109/ACCESS.2023.3329575

[10] Bian, Y., Küster, D., Liu, H., & Krumhuber, E. G. (2024). Understanding Naturalistic Facial Expressions with Deep Learning and Multimodal Large Language Models. *Sensors*, 24(1), 126. https: //doi.org/10.3390/s24010126

[11] Karki, N., Anwar, A., Ur Rehman, I., Husamaldin, L., & Saadati, P. (2023). Smart attendance monitoring system using face recognition for people with disabilities (PwDs). In *2023 IEEE International Smart Cities Conference (ISC2)* (pp. 1–7). https: //doi.org/10.1109/ISC257844.2023.10293664