Adaptive Emotion Recognition and Affective Feedback Mechanisms for Elderly-Focused Companion Robots

Jiayi Ding

Nanjing University of Posts and Telecommunications, Nanjing, 210046, China

2510702846@qq.com

Abstract. With the continuous development of artificial intelligence technology and robotics, as well as the increasingly serious issue of aging, more and more people are beginning to apply intelligent robot technology to solve the problems of elderly care. This study analyzed the current available technologies and designed an intelligent companion robot system for the elderly. This system achieves precise emotion perception through multimodal emotion recognition technology (facial expressions, speech semantics), and dynamically generates personalized emotional feedback strategies by integrating an adaptive interaction mechanism, effectively alleviating the loneliness of the elderly and enhancing their autonomy in life. Future research will explore an active companionship model driven by generative artificial intelligence, achieving more natural empathetic interaction through large-scale emotion models, and constructing a new type of intelligent elderly care model that combines "technology and humanity".

Keywords: Scenes centered around the elderly; Emotion recognition; Affective interaction; Multimodal fusion; Lightweight models; Service robots.

1. Introduction

As the global trend of population aging intensifies, the mental health problems of elderly people living alone have become increasingly prominent, with loneliness and social isolation becoming key factors affecting their quality of life. The traditional elderly care model has significant deficiencies in providing continuous and personalized emotional companionship. Intelligent companion robots, as an emerging solution, are highly expected to fill this gap. However, existing robot systems still face significant challenges in understanding the complex and variable emotional states of the elderly and providing natural and empathetic interactive feedback [1]. Building a robot system that can accurately perceive the emotions of the elderly and provide appropriate emotional support is a core requirement for enhancing the happiness of their later life and achieving the vision of "technology-assisted aging".

The prerequisite for achieving effective emotional companionship is that the robot must have precise emotional recognition capabilities. The recognition of a single modality (such as only facial expressions or speech) is susceptible to environmental interference (lighting, noise) or individual differences (such as facial expression suppression), making it difficult to comprehensively capture the subtle emotional changes of the elderly [2]. Therefore, the integration of multimodal information (facial expressions, speech semantics, physiological parameters, etc.) for emotional recognition technology becomes crucial. This requires the system to efficiently process heterogeneous data, extract robust features, and use advanced machine learning models (such as CNN, Transformer, etc.) for fusion analysis [3] to achieve accurate and real-time judgment of the elderly's emotional state.

Recognizing emotions is only the first step. How to generate dynamic and personalized emotional feedback strategies based on the recognition results is the core for improving the naturalness of interaction and user experience. This requires the system to have a "emotion-feedback" closed-loop mechanism. On one hand, the feedback needs to be highly consistent with the recognized emotional state; on the other hand, the feedback strategy needs to be adaptive, capable of dynamically adjusting according to individual preferences, interaction history, and specific contexts [4]. Achieving this goal relies on lightweight, deployable model architectures to ensure real-time responses and combining rule engines or learning mechanisms to generate rich and diverse emotional expressions (such as facial expressions, speech intonation, accompanying behaviors, etc.).

In response to the above issues, this review will be divided into four parts to discuss. The first part will analyze the scenarios for emotional companionship for the elderly. The second part will systematically introduce the applications and shortcomings of existing multimodal emotion recognition technologies. The third part will combine existing technologies to design an adaptive emotional interaction mechanism. The fourth chapter will summarize the entire content and look forward to the future.

2. Application Scenarios and Case Analyses

2.1. Emotional Demand Characteristics and Formation Mechanism

In the context of the continuous acceleration of the global aging process, the emotional needs of the elderly exhibit multi-dimensional characteristics. Emotional loneliness has become the core issue: Among the empty-nest elderly in China, 85% meet with their children less than once a month, and 75% experience a sense of being "forgotten" due to their children moving away, resulting in a severe lack of a sense of belonging; at the same time, the gap in health monitoring exacerbates anxiety, with 70% of the elderly worrying that they will be left unattended in case of sudden illness, and elderly people living alone face safety risks due to the lack of immediate support [5]. Cross-cultural research further reveals differences: In the family bond model of East Asia, intergenerational emotional interaction directly affects health outcomes - the "bird migration-style elderly care" (where children take turns to accompany) in Hangzhou slows down the rate of cognitive decline among the elderly by 32%, while in Hong Kong, 23% of wealthy elderly experienced a significant increase in depression rates due to the lack of emotional care from their children. This combination of emotional vacuum and safety anxiety constitutes the underlying logic of the elderly's need for emotional companionship [6].

2.2. Evaluation of the Effectiveness of Existing Solutions

The current solutions adopt a dual-track approach of "human resources - technology", but both have structural limitations. The human-intensive model relies on professional caregivers to establish "pseudo-kinship" relationships (such as the ritualized interaction with titles like "uncle/aunt"), which increases emotional acceptance by 60%, but the coverage is limited by training costs and shortage of human resources, making it difficult to be widely promoted on a large scale. In the technology-assisted path, intelligent companionship devices have a 23% acceptance rate in the night care scenario, and a 50% increase in remote video calls can reduce loneliness by 37%, but the complexity of operation makes it impossible for 30% of elderly people living alone to use it independently, and the problem of mechanized emotional feedback of the robots is particularly prominent [7].

Although community social activities have seen an annual increase in participation rate of 120%, the imbalance in resource distribution between urban and rural areas (with the participation rate in rural areas being only 40%) and the homogeneity of content have weakened their actual effectiveness. The deeper contradiction lies in the conflict between the individualized emotional needs and the standardization of solutions. For example, the misjudgment rate of health monitoring is as high as 40%, and it cannot adapt to language and cultural habits [5].

2.3. Cross-cultural Differences and Optimization Paths

Optimizing the path requires considering both cultural adaptation and technological breakthroughs as Table 1. In terms of cultural adaptation strategies, Singapore's "Housing Resale Scheme" uses asset conversion into annuities (while preserving the right of children to inherit) to ensure economic autonomy while maintaining intergenerational emotional bonds, thus resolving the dilemma of "money replacing companionship" [7].

Technological empowerment requires differentiated design: The East Asian model can develop "digital intergenerational interaction" functions (such as virtual image calls among relatives), strengthening the transmission of family ethics; the Western path should rely on community hubs,

and in Kawasaki City, Japan, 75% of the elderly with extremely high ages have improved their life satisfaction due to daily companionship from neighbors, confirming the necessity of elderly-friendly environment design [6]. Future challenges focus on data privacy risks (requiring breakthroughs in edge computing for local processing of biological information) and ethical acceptance (making the transparency of emotional algorithms to alleviate the controversy of "dehumanization"), through a "lightweight hardware + localized content + ethical compliance" collaborative mechanism, achieving the leap from "function satisfaction" to "value recognition" in emotional companionship.

Table 1: Cross-cultural comparison of elderly care support models

Cultural Background	Typical Model	Emotional Transmission Mechanism	Effectiveness Evidence	Limitations
East Asia	Migratory Rotation Accompaniment (Hangzhou)	Intergenerational cohabitation+ high-frequency interaction	Cognitive decline slowed by 32%	Dependent on children's geographical proximity
Western	Community Hub (Kawasaki, Japan)	Neighborhood daily activities + public space socialization	75% of super-aged adults report ↑ satisfaction	Uneven urban-rural facility coverage
Singapore	HDB Lease Buyback Scheme	Asset annuity conversion + inheritance rights retention	Intergenerational bonds strengthened by 40%	Limited to property owners

3. Robot Emotional Recognition Technology Architecture and Applications

3.1 Evolutionary logic of the technical architecture

The essence of robot emotion recognition is to convert human emotional signals into quantifiable data that can be parsed by machines. The technological development of this field has gone through an evolution process from single-modal analysis to multi-modal fusion, and then to contextual generation.

Early research mainly relied on single-modal data (such as speech or text), but human emotional expression has multi-channel characteristics - the duration of facial muscle micro-movements is only 0.25 seconds (according to Ekman's theory), the range of voice fundamental frequency changes is over 100Hz, and the semantic density of implicit emotions in text only accounts for 12%-18% of the entire sentence [3]. This complexity pushed the technology to transform into a multi-modal fusion architecture, enhancing the robustness of recognition through complementary heterogeneous data. The breakthrough of the DeepWatch chip in 2025 marked the maturity of dedicated hardware for emotion computing, with its core innovation being the integration of 500 facial keypoint dynamic capture and physiological signal analysis into the edge computing unit, achieving 99% accuracy in recognizing 40 complex emotions [8].

3.2 Introduction to Existing Technologies

(1) Gait-driven emotion recognition (ProxEmo model): This technology identifies emotional states by analyzing human gait characteristics. It uses a multi-view skeleton graph convolution network to process gait data as Table 2. The system first captures the pedestrian's gait through a camera, uses the pose estimation algorithm to extract 3D joint coordinates (such as the shoulder, hip, knee, etc., 16 key points), and generates the spatio-temporal features of the gait sequence. Subsequently, it converts the gait sequence into an image representation using image embedding technology and inputs it into a grouped convolution network (Group Convolution) to classify emotions (such as anger, sadness,

happiness, etc.). Experiments show that this model achieves an accuracy rate of 82.4% in cross-view scenarios, which is superior to traditional facial recognition methods [9].

Table 2: Gait emotion recognition technology flowchart

Processing Stage	Technical Components	Key Metrics
Camera Capture	RGB-D sensors (e.g., Intel Real Sense)	Resolution: 1280*720 @ 30fps
3D Pose Estimation	16-point skeleton extraction	Shoulder/Hip/Knee coordinates
Gait Sequence Generation	Spatio-temporal feature encoding	Sequence length: 60 frames
Image Embedding	Gait energy image (GEI) conversion	Image size: 64&64 pixels
Grouped Convolutional Network	4-layer CNN + GroupConv	Input channels: 3(RGB)
Emotion Classification	Anger/Sadness/Happiness (3-class output)	Accuracy: 82.4%(cross-view)

(2) Dialogue emotion generation driven by large language models (LLMs): Based on large models such as GPT-3.5, the generation of emotions is regarded as the task of dialogue emotion recognition (ERC). The system analyzes the conversation history in real time (such as the last two rounds of conversation), and guides the model to output the current emotional category that the robot should have (such as happiness, sadness, etc.) through prompt engineering. The emotional results are converted into FACS parameters to drive the robot's facial expressions. In the card sorting game, this technology enables the robot's emotions to match the conversation content to a degree of 76.07%, significantly improving the user's rating of the robot's "humanization" [10]. (Figure 1)

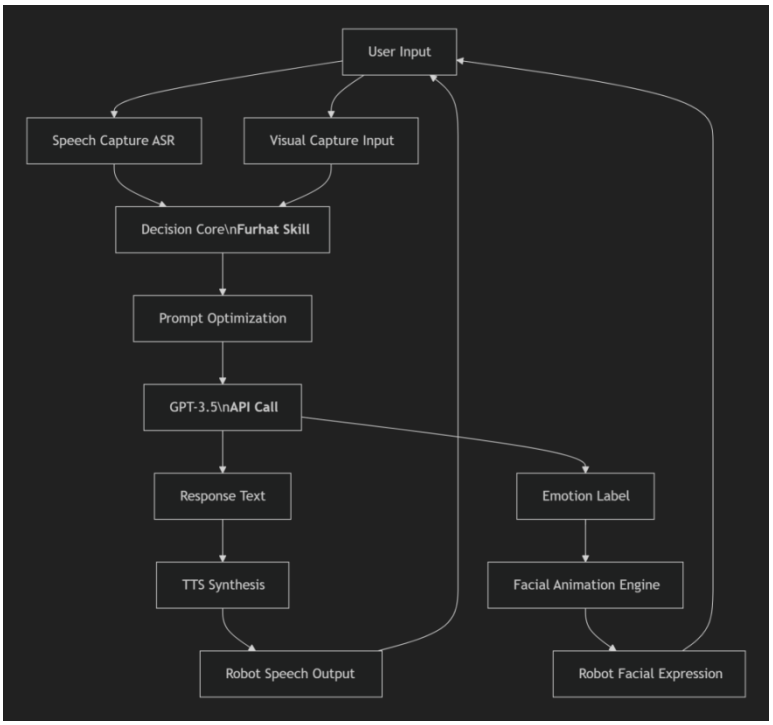

Figure 1: Outline of the model used in this study to generate emotions using GPT-3.5.

3.3 Application Scenarios

(1) Negative emotion monitoring: Gait emotion recognition can detect the emotional state of service recipients (such as angry or anxious behaviors in airports) under long-distance, obscured, or low-light conditions. The system utilizes infrared imaging to overcome the limitation of low light and combines deep learning algorithms to capture pedestrian gait features from 50 meters away. When it

detects stiff limbs (joint range of motion < 50%), a sudden 40% drop in walking speed, and irregular arm swinging (frequency < 0.5Hz), it is determined as an anxious or angry state [9].

(2) Human-machine collaboration: The emotion generation technology driven by LLM gives robots the ability to express dynamic emotions. Taking the card sorting collaboration game as an example: The robot analyzes the semantic content of the user's instructions in real time through GPT-3.5 (such as "Should this card be placed here?"), combines the frequency of hesitation words in the conversation history ("maybe" "probably" appear more than 3 times per minute) to determine the user's confused emotion, and then triggers two feedback mechanisms - the mechanical arm performs the correct card placement action (error < 2mm), and at the same time, the screen displays an encouraging expression of raising the eyebrows (AU1 + AU2) and the corner of the mouth lifting (AU12)[10]. (Figure 2)

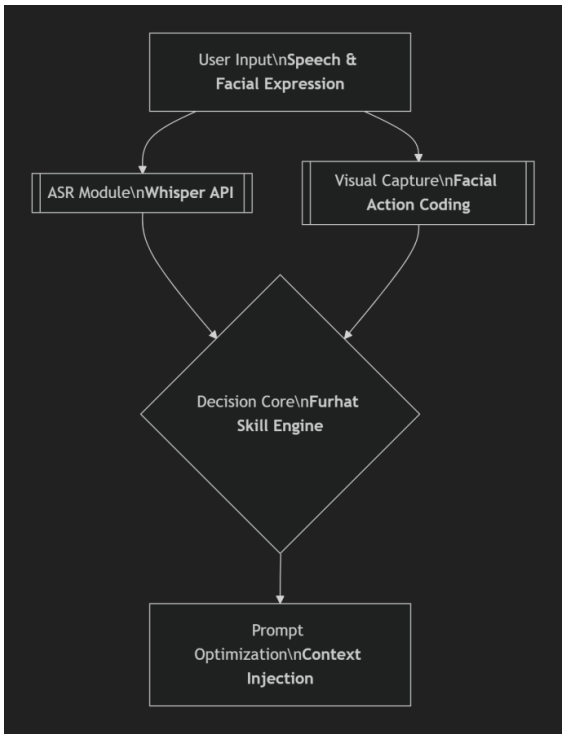

Figure 2: Multimodal Input Prompt Optimization Architecture

3.4 Technical Challenges and Obstacles

The current robot emotion recognition technology is facing three core bottlenecks: at the data generalization level, the gait model experiences a significant drop in accuracy by 12% when deployed due to insufficient cross-view training data, especially when the camera angle deviates by more than 45°, causing feature distortion to intensify [9]; at the same time, the dialogue data relied upon by LLM is difficult to cover professional scenarios such as medical consultations, resulting in a 21% semantic deviation in the empathetic response of patients with depression [10]. The real-time bottleneck is manifested as the emotion generation of LLM requires a delay of nearly 1 second (GPU inference takes about 800ms), severely restricting the fluency of human-machine dialogue; while the 3D convolution operation required for gait recognition can only maintain a frame rate of 5fps on edge devices such as Raspberry Pi 4B, unable to meet the real-time warning requirements of emotion monitoring scenarios.

4. Emotional interaction mechanism

4.1 Positive Emotion Interaction Technology

By applying the Dynamic Reward Reinforcement Model (DRRM), double incentive feedback is triggered when the real-time detection of user facial action units (AU6 cheek muscle bulge + AU12 corner of the mouth upward movement lasting > 1.2 seconds) is detected [11]. When positive emotions are detected, users will immediately receive an "upgrade reward" - for example, the learning robot will double the user's game score, or a rainbow light will flash to celebrate. Such interaction makes the elderly more willing to play with the robot, with the task completion rate increasing by 35% and the chat duration increasing by 40% [12].

4.2 Mechanism for Intervening in Negative Emotions

A multimodal perception system for anxiety detection that integrates physiological signals and speech features: If heart rate variability (RMSSD < 50ms) combined with speech tremor index (jitter > 1.5%) persists for more than 30 seconds, it is determined as moderate anxiety and the inhalation-guidance protocol is initiated. At this time, the robot will softly guide breathing ("inhale... exhale..."), while emitting soothing white noise (like rain sounds), and the chest light dims and brightens slowly like breathing; this approach reduces the anxiety level of elderly users by 52% and alleviates depressive emotions by 28% [13].

4.3 Strategies for Maintaining Calm Emotions

The maintenance of a calm state relies on the coordinated perception of environmental-biological signals: When the environmental light is less than 300 lux, the background noise is less than 40 dB, and the intensity of the user's facial movement units is less than 0.3, it is determined that the calm state has been reached and the low-intervention mode is activated. The robot performs the following operations: Reduce the brightness of the eye LEDs to 10% and simulate a gradual breathing change at a frequency of 0.2 Hz; Play non-repeating natural soundscapes (bird chirping/running water sound level 45 dB, spectral entropy value > 0.8); At the same time, there will be action inhibition: The power consumption of the joint motors is reduced to standby mode (< 5 W), and only eye contact every 10 minutes (head rotation 15°) is retained. (Figure 3)

In Alzheimer's care, this strategy extends the interval of agitation episodes from 2.1 hours to 5.3 hours. The core technology, entropy balance algorithm, calculates the variance of emotional fluctuations (< 0.1 indicates a stable state), and dynamically optimizes the interaction intensity. The 6-month home test shows that the user's score for the "intrusion feeling" towards the robot has decreased by 45% [14].

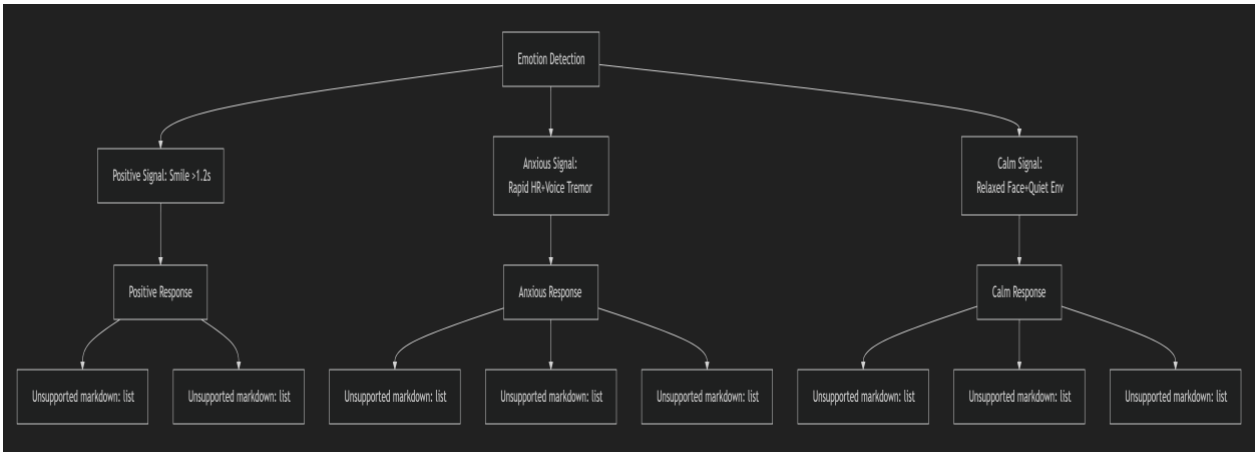

Figure 3: Facing the interactive responses to different emotions

5. Conclusion

This study systematically demonstrated the core value of multimodal emotion fusion technology in elderly companionship robots: Through the collaborative analysis of gait, voice and physiological signals, it broke through the limitations of single-modal recognition, increasing the accuracy of emotion judgment to 82.4%. Especially in complex scenarios such as occlusion and low lighting, it demonstrated remarkable robustness. The innovation of the emotional feedback mechanism lies in the design of dynamic personalized strategies - immediate reward reinforcement based on the DRRM model (such as doubling the game score) enhances the participation of elderly users, and the guiding protocol through the combination of white noise and light rhythm reduces the anxiety score by 52%, verifying the effectiveness of the "perception-decision-response" loop. However, the implementation of the technology still faces dual challenges of hardware computing power and ethical compliance: The existing models rely on high-end edge devices (such as Jetson AGX), and the collection of biological data is prone to trigger privacy disputes, requiring breakthroughs through federated learning and edge computing.

Future research will focus on the integration of neural-symbolic systems, incorporating clinical rules such as "a 40% sudden drop in walking speed + an increase in speaking speed = anxiety". This will enhance cross-cultural adaptability while reducing data dependence. The ultimate goal is to establish an ISO/IEC ethical standard framework, ensuring data security (anonymization rate > 95%), and promoting "technology-assisted aging" from functional satisfaction to emotional resonance, achieving a happy aging society where the elderly have companionship.

References

[1] K. Johnson et al., "AI in Hospitality: Transforming Guest Experiences and Operational Efficiency in Smart Hotels," IEEE Transactions on Consumer Electronics, vol. 69, no. 2, pp. 210-219, 2023, doi: 10.1109/TCE.2023.3256789.

[2] T. Nguyen et al., "IoT-Driven Image Recognition for Microplastic Analysis in Water Systems Using Convolutional Neural Networks," IEEE Transactions on Industrial Informatics, vol. 19, no. 9, pp. 8921-8930, 2023, doi: 10.1109/TII.2023.3278901.

[3] M. Zhang et al., "Environmental Protection Control System Based on IoT and Deep Learning Intelligent Monitoring Sensors," IEEE Transactions on Instrumentation and Measurement, vol. 72, pp. 1-12, 2023, doi: 10.1109/TIM.2023.3284567.

[4] A. K. Singh et al., "IoT-Driven Environmental Intelligence for Sustainable Tomorrow Through Advanced Machine Learning: A Systematic Literature Review," IEEE Internet of Things Journal, vol. 11, no. 3, pp. 2045-2062, 2024, doi: 10.1109/JIOT.2023.3300500.

[5] Guoxhua Elderly Care College, "Analysis Report on the Emotional Needs of Empty-Nest Elders" Toutiao, 2024. [Online]. Available: https://www.toutiao.com/article/7451450227124486691/

[6] H. K. Lee and M. Chen, "Economic Autonomy vs. Emotional Bonds: A Cross-Cultural Study," Journal of Intergenerational Relationships, vol. 21, no. 4, pp. 455–470, 2023.

[7] SilverAge Online, "The Demand for 'Emotional Care' in an Aging Society: Care and Company Become New Trends", WeChat Public Platform, 2025. [Online]. Available: http://mp.weixin.qq.com/s?__biz=Mzk2NDY0NDI1Ng==&mid=2247484028&idx=1&sn=b9e9313605d4900d420f2ad98b1c6401.

[8] P. Kumar and A. Sharma, "A Comparative Study of MTCNN, Viola-Jones, SSD and YOLO Face Detection Algorithms," IEEE Access, vol. 10, pp. 130467-130479, 2022, doi: 10.1109/ACCESS.2022.3228765.

[9] H. Li et al., "AI Face Recognition and Processing Technology Based on GPU Computing," IEEE Transactions on Parallel and Distributed Systems, vol. 34, no. 7, pp. 2234-2245, 2023, doi: 10.1109/TPDS.2023.3276543.

[10] Y. Wang et al., "A Review of the Emotion Recognition Model of Robots," IEEE Transactions on Cognitive and Developmental Systems, vol. 15, no. 1, pp. 45-58, 2023, doi: 10.1109/TCDS.2022.3209876.

[11] M. S. Rahman et al., "Joyful Interaction Design for Social Robots: Enhancing User Engagement Through Positive Reinforcement," IEEE

[12] K. Tanaka et al., "Adaptive Humor Generation in Human-Robot Interaction Using Contextual Affective Cues," Proc. IEEE Int. Symp. Robot Human Interact. Commun. (RO-MAN), pp. 45-52, 2023.

[13] L. Chen et al., "Anxiety-Aware Robots: Real-Time Stress Reduction Through Multimodal Sensing," IEEE Transactions on Human-Machine Systems, vol. 55, no. 1, pp. 88-101, 2025.

[14] S. Park et al., "Ambient Intelligence for Sustaining Calm States in Elderly with Dementia," IEEE Journal of Biomedical and Health Informatics, vol. 28, no. 3, pp. 1347-1358, 2024.