Intelligent Monitoring Technology and Safety Protection Analysis of Elderly Care Robots Based on Multi-Modal Perception

Ruizhe Ma

maruizhe726@gmail.com

DUT-BSU Joint Institute, Dalian University of Technology, Dalian, Liaoning, 116024, China

Abstract. As the process of aging the population of the planet becomes more and more significant, more of the elderly people will live alone, and elderly care robots which will assist them 24/7 can be used. This paper aims at exploring and discussing the development and use of essential technologies in hardware and software in three ways: in terms of vision, hearing, and vital signs monitoring. Interferometric fiber optic microphones and remote photoelectric plethysmographs can be incorporated on the hardware side and provide new solutions to an auditory reception and contactless measurement of elderly care robots, and several visual hardware accelerators will significantly enhance visual recognition capabilities. Deep neural networks, Transformer tracking and prediction models, which use local multi-scale features (Lo TPM) and the Kanade-Lucas-Tomasi (KLT) tracking algorithm also use the hardware, on the software side. This paper will attempt to offer a source to upcoming researchers.

Keywords: Elderly Care, Monitoring Technology, Multi-Modal

1 Introduction

With the challenges of dwindling birth rates and increasing ageing of the population slowly drying up both home-based and abroad, more and more aged individuals are without the attention and companions they need, which is potentially dangerous not only to their materialistic well being, but also to their psychological well being. Robots that are utilized in elder care are slowly replacing element within the elder care sector since they can supply 24 hours of companionship and offer a number of care and service characteristics, and have a very positive response in the market. Nevertheless, this is not always the case because the elderly population who needs care usually possesses issues, including unclear speech and slow or vague movements; therefore, identifying the voice and behavior pattern accurately is a major concern to elderly care robots. These are the main considerations of the real-time monitoring of the safety of the elderly.

The elderly care robots can observe and take care of the wellbeing of the seniors in several manners through the visual system, tracking system, and vital signs monitoring system. Nevertheless, the systems without full automation, even initially, where remote monitoring is performed by people, are still subject to scourges, including accidents because of the negligence

ICMEEA 2026, April 24-26, Singapore, Singapore

DOI 10.4108/eai.24-4-2026.2364922

of the operator. Hence, as a way of increasing the utility of the elderly care robots, increased studies are directed towards automated monitoring, tracking, and visual detection. To illustrate, scholars like Huang Ruqiang[1] have obtained results in monitoring the vital signs of seniors as a result of remote photoplethysmography and Zhang Xiaohan[2] has put forward a Transformer tracking and prediction model using local multi-scale features to allow robots to track them more precisely.

The purpose of this article is to discuss and summarize studies and implementation of eldercare robots in vision, hearing and vital signs monitoring domain. In this direction, the article presents the present position of research on each of the fields one by one with perspectives of hardware, software, and application and gives a view on the future of the eldercare robot in case of challenges and prospects.

2 Hardware

Forming the basis of robots, as the demand for their functionality continues to grow, simple hardware such as cameras with only a single function can no longer meet all needs. Based on the perceptual requirements of eldercare robots, the necessary hardware can roughly be divided into the following three aspects:

2.1Visual Inspection

Active 3D imaging technology used in robot vision, based on Time-of-Flight (ToF) and structured light, results in blurry color recognition, low imaging resolution, and is significantly affected by environmental factors. In contrast, passive binocular stereo sensors are more favored by researchers in service robot design due to their more accurate output of elements such as color information. On this basis, multiple researchers have achieved performance upgrades for vision systems from various perspectives, making significant contributions to the reception of visual information by robots.

Based on a strategy of algorithm-hardware co-desig[3], a high-energy-efficiency dedicated accelerator for Visual Simultaneous Localization and Mapping (SLAM) was proposed, targeting the Oriented FAST and Rotated BRIEF (ORB) feature extraction and matching algorithm. It successfully achieved increased frame rates, reduced Flip-Flop (FF) usage, and decreased lookup table usage, demonstrating significant low-power performance and real-time capability.

Zeng, H and others first developed a bit-level computing strategy, and on this basis completed a fundamental processing unit with precise processability. They also proposed an efficient set of processing units, successfully achieving an energy efficiency of 6.05 TOPS/W and a comprehensive improvement in area efficiency[4].

Bao, ZS, and others put a VIT inference accelerator implementation on the KR260. Their most important improvements were 2.74 times to 3.21 times of energy efficiency enhancement, at an average throughput of 323.6 GOP/S by optimizing nonlinear operations with integer quantization and building a token stopping module on the basis of a better token stopping mechanism [5].

2.2Auditory reception

The hearing is yet another vital reception of information by robots and has received attention among the researchers due to low sensitivity of the conventional electret microphones used in robots besides not being resistant to electromagnetic interference. This has led to the research taking the direction of new microphone arrays and pickups. Other researchers opt to enhance the efficiency and accuracy of reception system, and others move towards enhancing the power consumption of the system in order to have a better adaptability of the overall robot.

The use of a self-built Fabry-Pertr interferometer fiber optic microphone conducted by Deng Xin et al. and optimization of the TDOA positioning algorithm using FPGA proved to have higher positioning precision and stability both indoors and outdoors, and represented a wide range of applicability [6].

Wang Xiaoqin et al. suggested low-power microphone, which relies on highly integrated micro-core control chip, which integrated high quality sound signal processing with small size, low power, and good sound pick up qualities, thus proving to be highly practical [7].

A technique proposed by Miaotailuo, F, and others to spatially upsample a limited number of spherically shaped microphone arrays in the form of few capsules has been suggested to produce high-order microphone array signals based on physical constraints, starting with a low-order microphones [8]. Their performance is on the state of art of spherical microphone array upsampling strategies using signal processing.

2.3 Vital signs monitoring

Being one of the important modules towards the actualization of the senescence care robots core functions, more researchers are moving the vital signs monitoring module out of the conventional contact based monitoring to non-contact monitoring components, whilst still upholding the monitoring precision. The fact that this shift enables real-time measurements of that of the aged without interfering with their movements, as well as diversification of the methods of monitoring vital signs makes it possible to ensure a multi-dimensional monitoring of vital signals.

With the help of the neural network models based on deep learning, Qiao, DH, and other scholars were able to measure the vital signs of people by blood pressure (BP), vital signs of the heart of an individual: heart rate (HR), variability of the heart rate (HRV), and even the oxygen saturation of the blood (SpO2) with the help of a user facial video taken by the smart cameras in the framework of remote photoplethysmography (rPPG). This approach enables the acquisition of non-contact vital signs with the help of smart cameras, which requires minimal hardware needs to operate elderly care robots, and shows high potential of application [9].

Mezherjouseresht, P., et al. have managed to measure the physiological parameters of a target at varying positions by directing radar beams at the vital signs [10]. In addition, they employed continuous wavelet transform (CWT) and algorithms involving the application of various mother wavelets to the phase of the combined signal in order to achieve respiratory and heart rate monitoring with fairly low error values. The approach is applicable even on indoor elderly care robots, which are widely applied, which evidences a high level of practicality. Simultaneously, it can be utilized to supplement a monitoring model based on smart cameras, described above, allowing to identify the vital signs of aged people in real-time in a variety of different views.

3 Software

As hardware systems continue their development, the creation of software is also undergoing development. Better-resolution cameras provide a framework towards adapting automatic following systems on elderly care robots; better resound microphone arrays, coupled with spatial recognition could significantly benefit voice interaction with the elderly; and an array of physiological monitoring tools with efficient and designed software system could result in the achievement of full safety monitoring of the elderly.

3.1 Intelligent Follow System

Due to the fact that the elderly care robots must be able to track the elderly at all times, the tracking system must be extremely precise. At present, the key research directions include different ways of cameras, ultra-wideband (UWB), and infrared sensors.

On the DCF tracking framework, Transformer model, Zhang Xiaohan proposed a Transformer tracking prediction model (Lo TPM), and was also able to propose a distance measuring algorithm on the Orbbec Astra Pro depth camera, developing a successful method to make precise tracking and position measurement of pedestrians [2]. Moreover, based on an indoor setting, a vision-based follow robot system was also developed on the ROS system, and stable robot following was observed. This system is very compatible to the application need of the elderly care robots, therefore having immense application potential. Meanwhile, Gong Botao employed a deep neurological model that relies on convolutional autoencoders to handle position-based information that is collected by the camera on the robot and thus attained target following[11]. It is a deep-learning-based approach that allows the intelligent robot to independently track the targets, which maximizes the follow precision.

Liang Sichuang derived his work on the basis of Phase Difference of Arrival (PDOA) Ultra-Wideband (UWB) positioning technology. Through PDOA arrival phase difference positioning algorithm, he achieved the capability of the UWB base station in locating constrained mobile tags. This strategy made it possible to make UWB base stations miniature and viable to be installed on robots. Simultaneously with this, with an enhanced artificial potential field method, smooth and precise following in simulation was attained, eliminating problems of inaccessible target points and local minima. His works made the elderly care robots less difficult in the process of following elderly individuals in complex barriers of indoor surroundings [12].

3.2 Speech Recognition and Interaction

The speech recognition functions of the robot have also progressively got better based on the highly developed auditory reception module in the hardware section of this paper. Nevertheless, due to the fact that the better reception modules tend to capture more noise and interferences in the background, one of the focuses of the researchers in speech systems has been; how to remove these interferences and correctly identify the speech of the user.

Such relationship based on deep neural networks (DNN) there studied how clean speech and noisy speech correlate, with this correlation as a mapping function used to reconstruct improved speech out of noisy speech (Li Shiqi and others). Simultaneously, they also used a speech activity detection modeled with long short-term memory (LSTM) networks to more precisely approximate non-speech frames to make the DNN distinguish more between speech and noise [13]. This managed to achieve very high precision of the robot speech recognition in complex and noisy environments. The research is significant in reducing the effects of interference, e.g., noise caused by domestic appliances on elderly care robots, at the expense of using them, enhance the recognition rate of the voice of the elderly.

Another study approached this from a multimodal integration perspective. Hwang, JW and others proposed an audio-visual speech recognition (AVSR) model based on audio-visual information [14]. The model first performs audio-visual speech enhancement (AVSE) to enhance the target speech using audio-visual information and then utilizes the enhanced audio information along with visual information such as the speaker's lips or face. By integrating an AVSR model based on an integrator and hybrid decoding with a U-net-based AVSE model and recurrent neural network (RNN) attenuation (RA), the system achieves end-to-end training and successfully suppresses noise pollution. This model greatly improves the speech recognition capability of eldercare robots during face-to-face interactions with the elderly, showing high application potential.

3.3 Vital Signs Collection and Analysis

To ensure the safety of the elderly, in addition to close-range monitoring methods such as smartwatches, elderly care robots need to use a wider range of monitoring methods to obtain more comprehensive vital signs, thereby providing multi-faceted protection for the safety of the elderly.

Huang Ruqiang and others, based on remote photoplethysmography (PPG), used the KLT (Kanade-Lucas-Tomasi) tracking algorithm to perform real-time tracking of facial regions[1] . At the same time, they employed a maximum ratio merging algorithm to combine pulse wave signals from different pixel areas of the regions of interest, obtaining more robust pulse wave signals for measuring heart rate and respiration rate. This successfully enabled the remote acquisition of human physiological signals, with data showing high consistency with that generated by medical physiological monitoring vests. The algorithm is highly adaptable for monitoring functions in elderly care robots, demonstrating promising application prospects.

Another research direction for remote monitoring methods is millimeter-wave radar. Kang Xuanyuan and others proposed a signal processing method based on millimeter-wave radar, which successfully achieved effective separation and identification of respiration and heartbeat signals by using static clutter cancellation technology, phase unwrapping and differential enhancement algorithms, as well as variational mode decomposition (VMD) adaptive signal processing. This method is also suitable for the non-contact health monitoring required by elderly care robots, and it has relatively small error values, showing considerable application potential [15].

4 Application

Currently, the development of elderly care robots is taking place worldwide, with a considerable number of products being produced and put into use. The current products can be roughly divided into two categories: compact emotional companion types and larger practical function types. Emotional companion elderly care robots mainly provide psychological care and companionship for the elderly through voice interaction, while functional elderly care robots focus on monitoring and ensuring the safety of seniors living alone, or providing physical assistance to elderly people with mobility difficulties through specific physical forms (such as table-type robots).

Fig.1. ElliQ appearance (Picture credit : Original)

The representative emotional companion robot, ElliQ, is shown in Fig. 1. The core function of this robot is to provide proactive companionship based on the daily habits and emotions of the elderly, aiming to address social isolation and feelings of loneliness. In addition, it can link with health monitoring devices to track the elderly's physical data in real time and provide health management functions such as medication reminders and doctor appointments.

A representative robot in the category of practical functions is the Samsung Bot Care. This elderly-care robot can autonomously patrol indoors and use sensors and cameras to monitor the elderly's vital signs, such as heart rate, respiratory rate, and sleep quality, without physical contact. It can also automatically send alerts to the outside world if it detects a fall, abnormal inactivity (long periods without movement), or abnormal vital signs.

There is no superiority between the two; rather, they complement each other to provide comprehensive care for the elderly. As shown in Table 1, the two types of robots have their own unique core functions based on different core technologies, and they can achieve their main advantages in different scenarios.

Table 1. Multidimensional Comparison of Bot Care and ElliQ

Comparison Dimension	Samsung Bot Care	ElliQ
Core technology	Computer vision, autonomous navigation, non-contact biosensors	Context-aware AI, affective computing, behavior analysis, powerful third-party service integration
Core Function	1. Vital signs monitoring (heart rate, breathing, sleep). 2. Abnormal alerts	1. Proactive interaction and reminders. 2. Health and lifestyle

	(falls, prolonged inactivity)	management (integrating data from third-party health devices). 3. Digital connection bridge (simplifying video calls and sharing messages and photos)
Main Advantages	- Proactively identify risks (especially for emergencies). - Non-contact, unobtrusive monitoring. - Provide objective health trend data.	- Extremely high user engagement and emotional connection. - Highly simplified operations, reducing the digital divide.

5 Future Prospects

At present, elderly care robots still face many challenges. First is the high cost brought about by the need to meet multiple functional requirements. At the same time, they face ethical issues, such as how the elderly being cared for should interact with these robots. Moreover, due to the current limitations of AI, more research is needed on how to communicate more smoothly with the elderly. Even so, in the long run, elderly care robots are bound to become one of the crucial tools in an aging society, and researchers around the world are making efforts from various directions to advance this field. Now, with the advent of the era of big data and the rapid development of artificial intelligence technology, the field of elderly care robots has also taken center stage. With the aid of deep convolutional neural networks and AI, both the collection of audiovisual data and the accuracy of internal algorithms have greatly improved. As researchers from various fields gradually discover the immense potential contained in elderly care robots, more and more innovations are being applied to them, achieving multifaceted functional expansion, whether it is companionship for the elderly on an emotional level or assistance in daily life. Overall, elderly care robots are still on a path of gradual improvement, and with the continuous efforts of many researchers, they are sure to overcome current challenges and better serve human society.

6 Conclusion

This paper summarizes and analyzes the current research status of vision, hearing, and vital signs monitoring in elderly care robots, and explores the key technologies and development prospects of elderly care robots. Firstly, in terms of vision, various specially designed visual accelerators significantly improve the frame rate and energy efficiency of visual recognition in elderly care robots, while system designs such as LoTPM also significantly enhance accuracy and automatic tracking capabilities. In terms of hearing, the introduction of the FPI optical microphone and integrated sound acquisition design, combined with DNN and AVSR algorithms, strengthens the robot's anti-interference ability in noisy environments. Finally, in terms of vital signs monitoring, research mainly focuses on two directions: rPPG and radar beam technologies, both of which allow non-contact measurement of various vital signs. At present, how to better integrate the above research achievements into a complete elderly care robot and enable coordinated operation of each module still has considerable room for development. This research will allow the functionality of elderly care robots to advance to a greater extent.

References

[1] Huang, R.Q., Su, W.H., Zhang, S.Y.: Research on non-contact visual measurement methods of human vital signs for rescue robots. Medical and Health Equipment. pp. 26–31 (2020)
[2] Zhang, X.H.: Research on intelligent control system of vision-based follow service robot. Master's Thesis. Shenyang University of Technology (2023)
[3] Qi, X.Y., Liu, Y., Hao, S., et al.: Design of high-energy-efficiency visual SLAM hardware accelerator. Integrated Circuits and Embedded Systems. pp. 51–59 (2024)
[4] Zeng, H., Mao, W., Zhang, S., et al.: A precision-scalable vision accelerator for robotic applications. Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS 2024). pp. 1–5 (2024)
[5] Bao, Z., Li, H., Zhang, W.: A vision transformer inference accelerator for KR260. Proceedings of the 8th International Conference on Computer Science and Artificial Intelligence. pp. 245–251 (2024)
[6] Deng, X., Xiong, L.S., Dong, Z.F., et al.: Development of fiber optic microphone array for auditory navigation of mobile robots. Optics and Optoelectronic Technology. pp. 118–128 (2023)
[7] Wang, X.Q., Qiao, S.S., Hu, X.Y., et al.: Design and research of low-power pickups in robot auditory systems. Micro-Nano Electronics and Intelligent Manufacturing. pp. 11–15 (2020)
[8] Miotello, F., Terminiello, F., Pezzoli, M., et al.: A physics-informed neural network-based approach for the spatial upsampling of spherical microphone arrays. Proceedings of the International Workshop on Acoustic Signal Enhancement (IWAENC 2024). pp. 215–219 (2024)
[9] Qiao, D., Ayesha, A.H., Zulkernine, F., et al.: Revise: Remote vital signs measurement using smartphone camera. IEEE Access. pp. 131656–131670 (2022)
[10] Mehrjouseresht, P., Babarinde, O.J., El Hail, R., et al.: Wide-angle vital signs measurements by adaptive FMCW radar. Proceedings of the IEEE MTT-S International Microwave Biomedical Conference (IMBioC 2023). pp. 91–93 (2023)
[11] Gong, B.T., Wang, K.P., Zhang, L.: Research on autonomous following algorithm for intelligent robots based on deep learning. Electronic Design Engineering. pp. 25–29 (2022)
[12] Liang, S.C.: Research on intelligent follow robots based on UWB technology. Master's Thesis. Guilin University of Electronic Technology (2024)
[13] Li, S.Q., Zhou, Y.M., Zheng, X.Y., et al.: Research on speech enhancement algorithms for service robots in complex noise environments. Sensors and Microsystems. pp. 35–39 (2025)
[14] Hwang, J.W., Park, J., Park, R.H., et al.: Audio-visual speech recognition based on joint training with audio-visual speech enhancement for robust speech recognition. Applied Acoustics. pp. 109478 (2023)
[15] Kang, X.Y., Yuan, J.J., Cheng, Y.Z.: Research on millimeter-wave radar vital signs monitoring method based on VMD and phase difference enhancement. Mechanical and Electrical Engineering Technology. pp. 1–7 (2025)