

Autonomy of robots basing somatosensory interaction

Junxi Ye

Changjun Middle School International Department, Changsha, China

lan35zaocao@163.com

Abstract. Somatosensory interaction technology is an important pathway driving the development of artificial intelligence that meets societal needs, and the application of robots in human society is becoming increasingly widespread. This study aims to systematically review and evaluate the current state of artificial intelligence research based on somatosensory interaction technology, particularly its applications in the field of automated robotics. Using a literature review methodology, this paper elaborates on how automated robotics technology integrates somatosensory interaction to achieve natural human-machine interaction. It provides a comparative analysis of the progress made in current applications, existing limitations, and potential future development directions, followed by a comprehensive summary. The research indicates that somatosensory interaction technology offers a more intuitive and natural means of interaction for automated robots, achieving significant results across various fields. However, challenges remain in terms of technical precision, environmental adaptability, and cost. It is hoped that this study will provide valuable insights for the development of related technologies in China.

Keywords: Artificial Intelligence, human-computer interaction, somatosensory interaction

1. Introduction

Artificial Intelligence (AI) refers to the theories, methods, technologies, and application systems that enable machines to simulate, extend, and expand human intelligence. It encompasses various fields such as machine learning, deep learning, natural language processing, and computer vision, aiming to enable machines to think, learn, and solve problems like humans. Machine learning uses algorithms to allow machines to learn from data and improve performance, including supervised learning, unsupervised learning, reinforcement learning, etc. Deep learning employs deep neural networks to simulate the human brain, achieving more powerful learning and recognition capabilities, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs). Natural language processing enables machines to understand and process human language, including speech recognition, speech synthesis, semantic understanding, etc. Computer vision allows machines to "see" and interpret images and videos, such as image recognition, object detection, and image segmentation.

Human-Computer Interaction (HCI) refers to the process of information exchange between humans and computer systems, including input, output, and control of information. The goal of HCI is to make interactions between humans and computer systems more natural, efficient, and enjoyable. With the development of AI and motion-sensing interaction technologies, HCI methods are continuously evolving, transitioning from traditional keyboard and mouse interactions to more natural and intelligent voice interactions, gesture interactions, eye-tracking interactions, etc. Mainstream HCI technologies include Graphical User Interface (GUI), Voice User Interface (VUI), Haptic User Interface (HUI), and Somatosensory User Interface (SUI). In the future, HCI will place greater emphasis on user experience and personalized needs, integrating deeply with other technologies such as artificial intelligence, virtual reality, and augmented reality to create more intelligent and humanized interaction methods.

Somatosensory interaction technology allows users to interact directly with devices or environments through body movements without the need for complex control devices. This technology provides users with a more natural, intuitive, and immersive interaction experience, such as controlling TV playback by waving hands, mapping body movements to game characters, or freely exploring virtual reality environments. Mainstream motion-sensing technologies include camera-based motion sensing, sensor-based motion sensing, and deep learning-based motion sensing. However, motion-sensing interaction technology also faces several challenges. For instance, camera-based motion sensing is susceptible to environmental lighting and obstructions, devices equipped with motion-sensing technology are often expensive, limiting their widespread adoption, and it can be difficult to perform large movements in public spaces to meet recognition requirements.

In summary, technologies such as artificial intelligence (AI) and human-computer interaction (HCI) are profoundly influencing the development of somatosensory interaction technology, demonstrating remarkable effectiveness in enhancing the

convenience and well-being of residents. However, these technologies still exhibit certain limitations in their application toward advancing an intelligent society, making more in-depth and efficient research an urgent issue that needs to be addressed.

2. Robots and autonomy

As the global population ages, the demand for labor increases, and the development of intelligent society progresses, we have come to realize that robots may be the best solution to address these current challenges. However, the current level of automation and intelligence is still insufficient, making it difficult to effectively deploy them in certain roles. In recent years, with the rapid advancement of technology, motion-sensing technology has become increasingly widespread. Therefore, could the integration of robotics and motion-sensing technology further enhance the fulfillment of societal needs?

2.1. Robot and development

Robots are intelligent machines with functions such as perception, decision execution. They can work semi-autonomously or fully autonomously in various situations, helping humans complete dangerous or complex tasks, improving work efficiency and quality, and serving human life [1]. With the increasing demands of humans, the intelligent performance of robots has become important, and the concept of autonomous robots has been proposed. Autonomous robots are robots that come with various necessary sensors and controllers, and can independently complete certain tasks without external human input and control during operation. With the addition of autonomy, robots have become more intelligent and convenient, and can better assist humans in completing designated tasks.

The world's earliest machinery can be traced back to around 5000 years, when apes used stone tools to meet their daily needs. After thousands of years of development, humans learned to use bronze to make tools. Around 2000 years before the performance, ancient Chinese invented vehicles that used round wooden boards as walking parts. Around 1400 BC, humans only learned to use iron tools until 1495, Leonardo da Vinci drew the "Machine Warrior," which was later developed by a group of Italian engineers over a period of 15 years and is the earliest known humanoid machine.

The earliest robot in the world was Elektro, which was only used for performances at the 1939 New York World's Fair. He had a vocabulary of 700 words, could walk, speak, and even smoke. Later, Asimov's Three Laws of Robotics proposed in 1942 became the rules that robots must follow thereafter. In 1954, George Devol created the world's first programmable robot, followed by Joseph Engelberg creating the world's first industrial robot in 1959.

In 1968, the very first intelligent robot Shakey was born at Stanford Research Institute. In the following year, the first robot that can walk on two feet was developed by the Ichiro Kato Laboratory of Punmei University in Inada, Japan, which raised the robot's action to a new height. Afterwards, various types of robots emerged one after another, such as service industry robots that appeared in 1988, canine robots in 1999, vacuum cleaner robots in 2002, surgical robots in 2015, and so on. The most impressive one was Sophia, the world's first robot to hold national citizenship in 2017. As for now, the most advanced robotics technology is divided into three categories of robots from robotics companies, namely Atlas from Boston, Optimus from Tesla, and Figure One implanted with Open AI, and they are all humanoid robots. Apart from humanoid robots, other types of robots also have their own advantages, such as Boston Dynamics' robot dog Spot, which can help inspect industrial equipment and search and rescue, and the application fields of UAV mainly focus on military, reconnaissance and photography. At the same time, the development of AI has also brought unmanned driving and various conveniences to ordinary people, with the latest ChatGPT being the most representative, which can help robots to "think".

Nowadays, many restaurants have started using automated robots to meet people's needs for service. For example, in Shanghai, China, there are many restaurants that use automated robots to make dishes and serve tea and water to customers, basically reaching a level without human labor, fully reflecting the convenience brought by automated robots. So autonomous robots have great prospects in the service industry and time-consuming engineering, but they also need to meet the full perfection of precision, adaptability and other functions.

2.2. Autonomy of robot

With the increasing adoption of automated robots, they are expected to possess more functionalities to ensure superior performance. However, current technologies still exhibit significant limitations. The evaluation criteria for automated robots can be assessed based on the following three key aspects.

2.2.1. Task execution capability

Evaluating the task execution capability of automated robots requires consideration of the following four key aspects:

Precision and Accuracy – This metric is determined by parameters such as repeatable positioning accuracy and operational error margins (e.g., millimeter or micrometer-level tolerances).

Efficiency—This criterion assesses the robot's task completion speed and productivity per unit time (e.g., the number of parts processed per hour).

Reliability—Measured by the robot's Mean Time Between Failures (MTBF), indicating its ability to operate continuously without malfunctions.

Adaptability—This evaluates whether the robot can handle multiple task types (e.g., a flexible assembly-line robot switching between different products).

2.2.2. Intelligence and autonomy

The intelligence and autonomy of automated robots are determined by three core competencies: perception, decision-making, and learning capabilities. These dimensions are evaluated through the following metrics: The first is Perception Capabilities, are to automated robots what eyes, ears, and a sense of touch are to humans. They serve as the fundamental foundation and key driver for achieving truly intelligent, flexible, safe, and efficient operation. They fundamentally transcend the limitations of traditional automated equipment, which could merely perform preset actions. It assessed based on sensor performance, including visual recognition accuracy and force feedback sensitivity. The second is Decision-making Capability, is to automated robots what the brain is to humans, it is the central engine for achieving autonomy, adaptability, and intelligent decision-making. When robots possess perception capabilities (by gathering environmental information), decision-making capability enables them to analyze information, evaluate options, plan actions, and execute autonomously. This allows them to transcend simple automation and advance toward truly intelligent collaboration and independent operation. It is measured by the optimization level of path planning algorithms and emergency response strategies (e.g., obstacle avoidance success rate). The third is Learning Capability is to automated robots what the evolutionary mechanism is to humans, it is the ultimate engine for breaking predefined limits, achieving continuous optimization, and adapting to the unknown. It is evaluated according to the system's support for machine learning optimization (e.g., improving sorting accuracy through data-driven approaches).

2.2.3. Safety and collaborative performance

The potential hazards posed by automated robots to human operators have remained a persistent concern in industrial applications. Optimal safety performance not only ensures operator protection but also minimizes equipment damage, while collaborative capability determines the system's effectiveness in human-robot interaction. The following three aspects serve as critical evaluation metrics. Firstly, human-robot interaction safety ensures secure operation between users and the robot, preventing accidents. This requires compliance with international standards (e.g., ISO 10218) and optimization of critical parameters such as emergency stop response time and collision force thresholds. Secondly, environmental compatibility guarantees the robot's adaptability to diverse or specialized workspaces. This necessitates low operational noise levels, high energy efficiency, and minimal spatial footprint requirements. Thirdly, collaborative performance ensures seamless interaction between the robot and users. This demands fluid operation in hybrid work scenarios and adaptive behaviors in dynamic environments.

These three evaluation criteria comprehensively cover the full-spectrum assessment needs of robots, spanning from fundamental performance (what to do), intelligence level (how to do it), to social attributes (for whom to do it). This aligns with the current research trend in robotics technology that is shifting from "human replacement" to "human augmentation".

Numerous technologies can enhance automation standards. Currently, with the advancement of motion sensors and the growing demands for lifestyle and entertainment applications, motion-sensing technology has become a key focus area. This technology can enable automated robots to achieve more intelligent applications.

3. Somatosensory technology

Body-sensing technology refers to a technology that allows people to directly interact with surrounding devices or environments using body movements without the need for any complex control devices, allowing people to experience and interact with content in a immersive way. For example, when you stand in front of a TV, if there is a motion-sensing device that can detect your hand movements, if we wave our hands up, down, left, and right to control the fast-forward, rewind, pause, and stop functions of the TV, it is a direct example of using motion sensing to control peripheral devices, or mapping these four actions directly to the reactions of game characters, allowing people to have an immersive gaming experience. Other applications of somatosensory technology include 3D virtual reality, spatial mouse, gamepad, motion monitoring, health care, etc., which will have a large market in the future.

3.1. Technology

3.1.1. Inertial sensor

When using inertial sensing technology, the device will capture the motion acceleration, azimuth change, motion angle and other state signals collected by the sensors connected to the joints that need motion input, and integrate them together to obtain the spatial motion of each joint. This method is not hindered by the spatial field of view, but the cumulative error is large. However, with the development of MEMS (Micro-Electro-Mechanical System) technology, this technology has been widely used in more and more fields, because it has relatively low cost, simple experiment, and it's will not be influenced by other external interference, has low requirements for environment and space, collects data more accurately, and is less invasive to people's privacy.

3.1.2. Optical sensor

Optical sensing technology captures optical signals reflected from the human body either through active emission (primarily near-infrared light sources) or passive reception of ambient light. These signals are subsequently converted into electrical signals by image sensors. Depth perception constitutes its core function, primarily relying on three key techniques: the structured light method, which projects encoded patterns and analyzes their deformation to reconstruct depth maps through triangulation; the Time-of-Flight (ToF) method, which directly calculates distance based on the phase difference or propagation time of modulated light pulses; and the stereo vision method, which employs multi-view cameras to match feature points for 3D reconstruction. The generated depth maps, combined with RGB data, undergo processing by computer vision algorithms to achieve human body segmentation, skeletal tracking, and gesture recognition. These outputs are ultimately mapped into interactive commands. This technology offers advantages including non-contact operation, rich information content, and real-time performance. However, it still faces challenges such as interference from lighting conditions, occlusion issues, and high computational complexity, establishing it as a critical enabling technology in the field of natural human-computer interaction.

3.2. Improvements to robot

The application of motion-sensing technology has significantly enhanced the level of robotic automation. Many companies and research teams have utilized this technology to accomplish tasks that were previously unachievable or to improve existing processes.

Here are some examples: Ahmad AKL and his partners collected nearly 20 kinds of attitudes using three-axis acceleration sensors, and carried out experiments and detection on them using dynamic time warping and nearest neighbor propagation algorithm, with a recognition rate of 98.4%; Song S K and his research partners used the same method to collect activities such as walking, sitting, standing, running and falling that people often do on human days. They used multi-layer perceptron to recognize the corresponding actions, and the recognition rate was 97.9%. V. Kosmidou and his research partners used acceleration and surface electromyography sensors to collect the relevant information of the hand. They used the sample entropy algorithm to recognize Greek sign language words, and the accuracy rate was as high as 92%; Julien pansiot and his research partners also use a three-axis acceleration sensor, but the object is the swimmer. They collect the swimmer's posture, calculate the elevation pitch and roll angle, determine whether the action is a stroke or a push, and then obtain the number of laps and segmentation time; Oliver Amft and his research partners used an isolated hidden Markov model to study a series of human actions including using tableware and eating and drinking when eating. With the help of four acceleration sensors to collect arm information, the recognition rate reached 94%.

- For reacting to change:

By using motion-sensing technology to perceive the environment, robots can respond differently based on varying conditions, thereby enhancing their own reactive capabilities.

Bourke and his research partners used a dual axis gyroscope to collect relevant data in order to distinguish between daily movements and falling movements, and then set the angle and the different order derivatives of the angle to achieve the corresponding purpose [2].

- For improving interaction:

With optical motion sensing, robots can collaboratively perform tasks with humans by tracking optical markers, thereby enhancing interactions between robots, humans, and the environment.

Collectively, by integrating inertial sensing technology with optical sensing technology, automated robots equipped with motion-sensing capabilities can more effectively execute diverse or specialized tasks. Their integration forms a robust and complementary sensory system, endowing robots with heightened autonomy, precision, robustness, and adaptability.

4. Application

Automated robots equipped with somatosensory interaction technology have achieved effective practical adoption across numerous fields. Below are three examples from mainstream application areas:

4.1. Medical applications

The KD9 Lower Limb Exoskeleton Gait Training System, developed by Qingdao Kangdao Medical, employs AI-based interactive technology to dynamically adjust gait patterns in real-time, assisting patients in achieving independent ambulation. The system records training data to optimize rehabilitation protocols, thereby aiding individuals with lower limb motor impairments in restoring walking ability.

Additionally, exergames (motion-based video games) have been utilized for postoperative rehabilitation in patients with lung cancer and thyroid cancer. Research indicates that exergaming interventions can significantly enhance patients' fatigue tolerance and quality of life.

In surgical applications, force-feedback robotic systems such as the Haption Virtuose 6D provide high-precision haptic feedback, enabling surgeons to practice virtual procedures (e.g., suturing, cutting) in simulated environments. This technology enhances surgical skills while minimizing operative risks. Furthermore, haptic interfaces (e.g., force-feedback arms) allow surgeons to remotely control robotic systems, addressing disparities in medical resource distribution.

For amputees, smart prosthetic limbs equipped with electrical stimulation can simulate natural tactile sensations by stimulating residual limb nerves, improving functional recovery and sensory feedback.

4.2. Entertainment applications

The entertainment robots developed by Jiangsu Borui Culture, integrated with VR technology, provide immersive experiences such as "Soaring Over Chongqing." Participants wearing VR headsets experience simulated rollercoaster movements through robotic arms, achieving a virtual-physical fusion high-altitude adventure.

In cinematic settings, D-BOX's haptic feedback seats synchronize with film content to deliver vibration and tilt effects, significantly enhancing viewer immersion. This technology has been widely deployed across Australia and New Zealand [3]. These applications collectively enable users to achieve heightened immersive experiences.

The "AI BOT Dynamic Robot Carnival" event at Shanghai IFC Mall featured robots capable of engaging in boxing matches and soccer games with human participants, demonstrating high agility and impact resistance while providing technology-enhanced sports entertainment. During the 2025 Spring Festival Gala, Unitree's H1 robots, utilizing AI motion control technology, performed the traditional Northeast Yangko dance in coordination with human dancers, executing complex maneuvers such as handkerchief twirling [4].

Concurrently, AI-powered motion-sensing fitness has emerged as a prominent application in recent years. Shimi Network's motion-sensing algorithms, combined with wearable devices like smartwatches, integrate fitness movement recognition with AI coaching guidance to facilitate interactive home-based exercise experiences.

4.3. Life applications

The "Capsule Interface" technology developed by H2L enables users to remotely control robots through muscle activity to perform delicate tasks, such as organizing objects or cleaning [5]. Additionally, robots can engage in natural interactions by recognizing human gestures, facial expressions, and speech, facilitating applications such as assisting children with learning, providing companionship for the elderly, and participating in household entertainment activities (e.g., dancing and gaming).

In educational settings, motion-sensing interactive robots can be employed in practical courses, such as programming and robotic operations, allowing students to directly control robots via gestures or movements to complete experiments. Furthermore, robots are now being utilized in public safety and service roles. For instance, Chengdu has deployed robotic "police officers" for routine patrols, equipped with autonomous navigation and anomaly detection capabilities, while also supporting remote command through motion-sensing interaction [6]. This reduces the workload of law enforcement personnel. Similarly, in public spaces such as shopping malls and airports, robots can provide guided services through gesture recognition or voice interaction [7].

As the technology matures, such robots are expected to become more prevalent and integrated into a broader range of daily life scenarios.

5. Conclusion

As a significant development direction in the field of artificial intelligence, intelligent somatosensory interaction technology enables natural human-machine collaboration through multimodal perception and intelligent decision-making, demonstrating broad application prospects in areas such as medical rehabilitation, home services, and immersive entertainment. Based on high-precision sensing, deep learning algorithms, and real-time feedback systems, this technology can accurately recognize users' movement intentions (responding to external stimuli) and achieve personalized interaction (e.g., large-scale human-machine collaboration), significantly enhancing the naturalness and intelligence of human-machine interaction. (This improvement in the intelligence level of robots is reflected in three aspects.)

However, the current technology still faces numerous challenges, including hardware-related cost and compatibility issues, insufficient environmental adaptability, limited algorithmic precision, high computational demands of system architectures, and the lack of a standardized framework. Particularly notable bottlenecks exist in data collection, multimodal command processing, abstract reasoning capabilities, and continuous learning mechanisms. Future breakthroughs should focus on lightweight markerless sensing technologies, robust cross-domain recognition algorithms, optimization of end-edge-cloud collaborative computing architectures, and the construction of cross-scenario action semantic libraries. Simultaneously, research on neuro-inspired learning, distributed intelligence, and lifelong learning mechanisms needs to be strengthened to propel intelligent somatosensory interaction technology toward greater efficiency and intelligence. This will facilitate a paradigm shift from passive execution to active situational understanding (ensuring task intent automation and safety guarantees for intelligent agents) and provide robust support for the development of an intelligent society.

References

- [1] Wikipedia contributors. (2019). Autonomous robot. In *Wikipedia*. https://en.wikipedia.org/wiki/Autonomous_robot
- [2] Yang, H. T. (2016). Research on human action recognition based on inertial sensors [Master's thesis, Southeast University].
- [3] GlobeNewswire. (2025). D-BOX and HOYTS deepen collaboration to expand premium cinema experiences across Australia and New Zealand. *Business Upturn*. <https://www.businessupturn.com/brand-post/d-box-and-hoyts-deepen-collaboration-to-expand-premium-cinema-experiences-across-australia-and-new-zealand/>
- [4] Thepaper.cn. (2025). Zero-frame hand-to-hand silk fan dance: Unveiling the Spring Festival Gala robot's "cyber folk dance". https://m.thepaper.cn/baijiahao_30118476
- [5] Malayil, J. (2025). Japan builds glove that steers drones remotely with hand gestures. *Interesting Engineering*. <https://interestingengineering.com/innovation/hand-gesture-drone-control>
- [6] ISPR. (2025). Japan's new tech turns your body into a remote control for humanoid robots. <https://ispr.info/2025/07/01/japans-new-tech-turns-your-body-into-a-remote-control-for-humanoid-robots/>
- [7] Fang, T. (2025). Shenzhen Cultural Industries Fair: Jiangsu exhibitors showcase entertainment robots take you "up mountains and down seas". *Jiangnan Times*. http://www.jntimes.cn/zdxx/202505/t20250526_8489133.shtml