Proceedings of the 4th International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2026.38119

Tactile Perception for Embodied Intelligent Dexterous Hands: Sensor Principles, Sensor Placement, and Multimodal Fusion

Xiaoyi Gong

Hebei University of Technology, Tianjin, China
Xiaoyi.Gong@student.lut.fi

Abstract. To perform these movements better, dexterous hands need to be able to feel contact with objects and the surrounding environment. This article mainly introduces tactile technology for dexterous hands, and it discusses these technologies from the view of embodied intelligence. The article mainly introduces five types of tactile sensors: piezoresistive, capacitive, piezoelectric, triboelectric, and optical sensors. These sensors are categorized by their different working principles. This article compares their sensitivity, sensing range, response speed, ability to detect small details, and the difficulty of integrating them into the system. Secondly, the deployment strategies of sensors at fingertips, palms, full hand surfaces, and soft-bodied hands, as well as the engineering constraints they face, are analyzed. Thirdly, multimodal fusion methods of touch, vision, and force, as well as their roles in enhancing grasping stability, are discussed, along with preliminary explorations of combining tactile information with visual-linguistic-action models. The study finds that piezoresistive sensors achieve a 92.8% classification accuracy in fabric texture recognition, outperforming human subjective classification. Soft-bodied humanoid dexterous hands can achieve all grasping poses in the GRASP classification by integrating distributed tactile sensors. Based on literature analysis, this paper points out that the core bottlenecks of current research include insufficient long-term stability of sensors, lack of interfaces between tactile representation and basic models, and limited computational power for real-time closed-loop control. It also proposes future research directions such as self-powered sensing, tactile basic models, and simulation-to-reality transfer.

Keywords: tactile perception, dexterous hand, embodied intelligence, multimodal fusion, VLA model

1. Introduction

Embodied intelligence emphasizes that intelligent agents possess physical bodies and acquire intelligence through active interaction with the physical world [1, 2]. Embodied intelligence is not only based on symbols and pictures, but also operates in an unstructured real-world environment; thus, it is more demanding of perception, decision-making and execution. As the first end-effector for interaction with the physical world, a dexterous hand should have good performance for the

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.38119

embodied intelligence system to carry out fine manipulation [3]. Earlier research on dexterous hands focused primarily on motion. Kinematic Model and Trajectory Planning are examples of this. However, if one wishes to perform a task well with skilled hands, it must be known how things feel to the touch in addition to the ability to move. The people's hands are also missing this function. Many Touch Receptors are located in the skin. Quickly acquire information on the contact force, surface features, temperature and slip of an object. The sense of touch helps people develop fine motor skills, such as holding an egg or playing the piano. Vision can provide the shape and position of an object, but it cannot fully sense the force and other mechanical information at the point of contact in real-time [4]. Therefore, Touch can address this problem. Contact information of the hand with objects can be used to inform intelligent judgment and movement decisions.

Flexible Tactile Sensors have been developing rapidly in recent years [3, 5]. Various Sensors are employed to measure force in different ways. Piezoresistive sensors detect changes in the resistance of sensitive materials under pressure to measure force. They have a simple structure and are relatively low in cost. Capacitive sensors are based on detecting changes in the distance between electrodes or in the dielectric constant. They can be used to create electronic skin. Piezoelectric sensors are based on the positive piezoelectric effect, have a very high-frequency response, and are suitable for detecting dynamic signals such as vibration and sliding. Triboelectric Sensors are contact-electrification and electrostatic-induction devices that do not need an external power supply. An optical sensor captures an image of elastomer deformation with a micro-camera to reconstruct the distribution of contact forces in space at the micrometer level [4]. Tang Chaoquan and others have developed a bionic finger with good performance in tactile perception [4]. Liu Yibo and others have systematically studied the research progress of soft humanoid dexterous hands [6].

However, the enhancement of hardware performance does not automatically lead to an increase in system ability. Currently, research still has three problems. First of all, there is a problem of deploying sensors in the confined space of a dexterous hand and ensuring their long-term stable operation. The second is that a map has not yet been built for the tactile information to high-level decision-making. Most of the current work is based on tactile signals for binary grasp judgment, and their application in object attribute recognition and online strategy adaptation has not been fully explored. The third is that the visual-language-action model is primarily based on visual information, and the introduction method and representation learning for tactile data have yet to be investigated [5].

Existing reviews either focus on sensor materials and principles or concentrate on the mechanical design and control of dexterous hands, and rarely address the issue of tactile perception and dexterous manipulation from a system level [3, 6, 7]. Due to this reason, the tactile-perception technology has not been widely applied to dexterous hands. To solve this problem, the first few types of tactile sensors and their respective advantages are introduced here. The Position of the Sensor is constantly changing. In the past, only the fingertips were used; now, however, the entire hand has been adopted. At the same time, sensors also began to be applied to rigid structures and, later on, were placed in soft structures. In addition, the article will also present some other types of information and explain how they can be combined to help improve the grasp of objects. Based on the above analysis, it can offer some reference for the future design of tactile-perception systems for dexterous hands.

2. Tactile sensor principles and performance

Tactile sensors can be seen as the basic hardware that dexterous hands use to feel things. Based on their different ways of sensing contact, these sensors can be divided into different types. At present,

DOI: 10.54254/2755-2721/2026.38119

there are five main types that are commonly used or have good potential for use in dexterous hands: piezoresistive, capacitive, piezoelectric, triboelectric, and optical sensors [3]. These tactile sensors are different in many aspects, such as sensitivity, sensing range, response speed, the ability to detect small changes, and the difficulty of putting them into the system. Some sensors perform well in laboratory tests; however, when applied to actual dexterous hands, their performance requires further testing and evaluation [7].

2.1. Piezoresistive sensor

Piezoelectric sensors are generally employed in tactile detection for dexterous hands. Examples of commonly used pressure-sensitive materials are carbon nanotubes, graphene and metal nanoparticles. When an external force is applied to these materials, the internal conductive structure of the materials changes. Therefore, the resistance will change. Measure the change in resistance to find the size of the external force. Due to its simple structure, low cost and ease of signal output, this type of sensor is widely used [7]. They can also be organised in a high-density array and placed on the fingertips of nimble fingers. Arrays can find small surface features and improve the recognition accuracy of object textures.

Tang Chaoquan and others have developed a bionic finger based on the piezoresistive effect to sense contact pressure. The linearity of pressure and voltage variation in the range of 0 to 5.0 Newtons is 0.997, and the sensor has good linearity in the commonly used force range [4]. In the hardness perception experiment, when the bionic finger pressed on wood blocks, foam plastics, and sponges, the pressure curve change rates were L_wo, L_fo, and L_sp, respectively. The magnitude relationship among them is $L_wo > L_fo > L_sp$, which is completely consistent with the actual hardness order of the materials, verifying that the piezoresistive sensor can quantitatively reflect the hardness of objects through the pressure curve change rate [4]. Furthermore, in fabric texture recognition, based on the support vector machine model, using 10 feature parameters including peak mean, average power, dominant frequency, power spectrum centroid frequency, and intensities of six frequency bands, the average recognition accuracy for six types of fabrics, namely linen, gauze, cotton, lycra, polyester fiber, and silk, reached 92.8%, which is higher than the human subjective classification accuracy of 88.8% [4]. However, this type of sensor suffers from nonlinearity, hysteresis, and baseline drift under long-term cyclic loading, with a trade-off relationship between sensitivity and measurement range.

2.2. Capacitive sensor

Capacitive sensors can detect whether an external force is applied by measuring changes in the distance between two electrodes or changes in the material's dielectric constant. They can be connected with electronic skin. They can not only detect pressure, but also sense whether an object is coming close. In this way, dexterous hands can know the distance and direction of an object before they actually touch it [3]. This feature is very useful in manufacturing, especially in some fine assembly tasks that need to know the object's position before contact [3]. Liu Yibo et al. reported that capacitive sensors have high sensitivity, fast response, and low power consumption. But their measurement circuits are more complex, and they can be affected by electromagnetic interference [6]. Boutry et al. developed a bionic electronic skin using a capacitive sensor array. The sensor design copies the interlaced structure between the dermis and epidermis of human skin. It can measure and distinguish between normal and tangential forces in real time, has high sensitivity, and a response speed of the order of milliseconds [6]. Capacitive Sensors are also susceptible to

DOI: 10.54254/2755-2721/2026.38119

electromagnetic interference (EMI). Signal stability is reduced in a complex environment, and the signal conditioning circuit is also relatively complex.

2.3. Piezoelectric sensor

Piezoelectric sensors are based on the direct piezoelectric effect of piezoelectric materials. A force is applied to the material, and therefore polarisation and a surface of free charges occur. Therefore, the response speed of these sensors to changes in force is relatively fast. They are suitable for detecting some changing tactile signals, such as vibration, sliding, and texture information obtained by scanning the surface of objects. They can be used to find out if an object is in motion and to calculate the friction force. The data can be used to determine whether the dexterous hand can hold the object stably [3]. Piezoelectric sensors have good sensitivity and a high-speed response, thus many types of motion changes can be detected. They are unsuitable for the measurement of static loads. If the force is applied continuously, the generated electric charge of the sensor will gradually dissipate [6]. Therefore, piezoelectric sensors are often employed in conjunction with other sensors. Thus, they will be able to find both fixed-position and moving-position forces. The problem of adjusting the grasp based on changes in contact force is also relatively serious.

2.4. Triboelectric sensor

Triboelectric sensors are new types of sensors that have been developed in the past few years. The first type of these devices is that they can generate electrical signals independently without an external power source. They work by triboelectric charging and electrostatic induction, etc. Their main advantage is that they do not need to connect with an external power supply and have very low power consumption; therefore, they are very suitable for portable or wearable devices [5]. The other advantages of these sensors are that they have a simple structure, high sensitivity and self-powered operation. However, there are also some defects; they have a small dynamic range and are not very durable in the long run [6]. At present, this technology is still in the initial stages of application for dexterous hands. The stability of the output signal and the consistency of the result still need to be improved. More work needs to be done before these sensors can be widely used in practice.

2.5. Optical sensor

Optical tactile sensors capture deformation images of the surface contacted by an elastomer through a built-in micro-camera, and then use computer vision algorithms to reconstruct the contact force distribution and surface features. This technology is listed as one of the key enabling technologies for dexterous hand precision manipulation [3]. Representative works include the GelSight series of sensors, which achieve spatial resolutions up to the micrometer level. In recent years, optical sensors have also begun to be embedded in soft fingers. Their advantages include extremely high resolution and rich information, while their disadvantages include large size, complex system, and potential degradation in optical imaging quality during significant deformation of soft fingers [6].

2.6. Comprehensive performance comparison

Table 1 summarizes the comparison of five types of sensors in terms of key indicators. It can be seen from the table that no single sensor can simultaneously meet all the requirements of a dexterous hand, including high sensitivity, wide measurement range, fast response, high resolution, low power consumption, and ease of integration. Piezoresistive sensors are widely used because they have a

DOI: 10.54254/2755-2721/2026.38119

simple structure, low cost, and are easy to make into sensor arrays. Capacitive sensors have technological advantages in electronic skin integration. Piezoelectric sensors play an irreplaceable role in dynamic tactile perception, such as slip detection. Triboelectric sensors represent a new direction of self-powered sensing but are not yet mature. Optical sensors stand out in fine texture recognition with ultra-high resolution. Therefore, the integration of multiple types of sensors has become an inevitable trend in practical system design [5, 8].

Table 1. Comparison of the performance of five types of tactile sensors

Sensor type	Sensitivity	Range	Response speed	Spatial resolution	Integration difficulty	Self-powered	Static force detection
Piezoresistive	Medium-high	Middle	Middle	High (array)	Low	No	Yes
Capacitive	Middle	Medium-high	Middle	Medium-high	Middle	No	Yes
Piezoelectric type	Tall	Middle	Extremely fast	Middle	Middle	No	No
Frictional electricity type	Tall	Middle	Fast	Middle	Medium-high	Yes	No
Optical type	Extremely high	Wide	Middle	Extremely high	Tall	No	Yes

3. Deployment strategy of tactile sensors in dexterous hands

Despite the superior performance of sensors, their practical value will be greatly diminished if they cannot be effectively integrated into the limited space of a dexterous hand and operate reliably. Deployment involves multiple challenges such as spatial layout, mechanical compatibility, signal transmission and processing, as well as long-term stability [6].

3.1. Fingertip deployment

Fingertips are the areas that come into contact with objects most frequently during operation, so most dexterous hands prioritize deploying sensors at fingertips. Human-level dexterous operation requires fingertips to possess high-density and high-resolution tactile perception capabilities, which places extreme demands on the miniaturization and integration of sensors [7]. The bionic finger developed by Tang Chaoquan et al. measures 63 mm × 20 mm × 22 mm in size and includes phalanges, a flexible shell, fingernails, liquid, pressure-sensitive components, and temperature-sensitive components [4]. The distal phalanx is 32 mm × 15 mm × 8 mm and includes both pressure and temperature sensing in a small space. Tests with this bionic finger show that it can distinguish materials with different hardness levels by measuring the rate of change in the pressure curve [4]. The biggest problem is that the space is too small when sensors are placed at the fingertips. When designing, it needs to keep a balance between the number of sensors, the strength of the structure, and the quality of the signals.

DOI: 10.54254/2755-2721/2026.38119

3.2. Sensor placement on the palm and full hand

Most previous research has focused on the fingertips. The palm can also be used to collect some data by contacting a large object. For example, it can provide information on the shape and position of the centre of mass for an object, as well as the stability of the grasp [3]. Full-hand tactile sensation is needed for process monitoring in complex assembly operations [3]. Liu Yibo and others have conducted a comprehensive study on the current progress of palm-driven humanoid dexterous hands [6]. The RBO Hand 2 is composed of two independently driven pneumatic cylinders in the palm to perform different thumb-opposition motions, and it has a total of 7 degrees of freedom and can reach 31 grasping postures in the GRASP classification [6]. The RBO Hand 3 can further perform manipulations of objects such as a rotating pen, a banana and a screwdriver in the hand [6]. BlueHand has a structure with three bellows actuators in parallel at the thenar eminence, a total of 21 degrees of freedom, and can perform all grasping poses in the GRASP classification to achieve the highest score in the Kapandji test [6]. He et al. also realised all the grasping postures in the GRASP classification by directly printing distributed pneumatic tactile sensors on the fingers and palm using multi-material 3D printing technology [6].The difficulty of palm sensing lies in its large area and complex surface, requiring a flexible array with large area and high uniformity.

3.3. Soft gripper insertion

Soft dexterous hands have become an important direction in embodied intelligence due to their inherent safety and high adaptability [5, 6]. Sensor embedding does not significantly alter the mechanical properties of soft materials. The sensor embedding strategies for soft humanoid dexterous hands include co-curing integrated molding, surface attachment, and internal embedding [6]. Zhao et al. integrated flexible optical waveguide sensors inside the fingers, with one end equipped with a light source and the other end equipped with a photonic strain sensor. When bending, the fingers of the sensors deform; consequently, there is a change in light propagation loss that can detect surface roughness and shape [6]. Gu and others have developed a soft neural prosthetic hand with flexible pressure sensors in the fingertips. The hand can be connected to electromyography for control and sense of touch for amputees [6]. Sensors are also susceptible to damage from soft hands. Maintain a stable signal in the presence of large deformation, ensure the long-term reliability of the interface, and reduce signal crosstalk.

3.4. Main engineering challenges

Tactile Sensor Placement Faces Four Main Problems. First of all, the Space available for us to use is relatively small. Secondly, there is signal interference. If the many sensors are too close together, there will be electromagnetic, thermal and mechanical interference, and the signal will be inaccurate. The third is stability in long-term operation. Many flexible sensors have not been bent and worn in any tests. The fourth is that the system is too complex. With an increasing number of sensors, it is required to collect, process and integrate the data more rapidly [8]. Now, tactile sensors are also available in other places. They are also in the palm and may spread out to cover all of the hand. The Location of the Sensors is also changing. In the past, sensors were mainly attached to rigid hands; now, some soft-hand structures have also been introduced that contain sensors. Although tactile sensors have been widely applied in dexterous hands, some problems still need to be addressed, such as how to achieve high-density integration of multiple sensors reliably and control the total system cost strictly.

DOI: 10.54254/2755-2721/2026.38119

4. Multimodal sensor fusion

Tactile Sensors are not very practical alone. They will cooperate with other kinds of sensors, such as vision and force sensors. Thus, the system will be more stable in detection and control. Multimodal fusion is the integration of data from multiple sensors in a single system for joint use. This information can be combined at the level of data, features, or decisions. Thus, it will be possible to address the shortcomings of a single-sensor system. It will also help stabilise and improve the accuracy of perception results.

4.1. Vision-tactile fusion

Vision can provide some information about the objects and their surroundings before contact, such as the shape and position of the object. After real contact, the tactile sensor will provide more detailed information about the touch. These two kinds of sensing can help each other to obtain all-round information for the robot. At present, a new type of integration for various kinds of information is provided by vision-language-action models (VLA). VLA models are all-in-one multimodal neural networks that combine vision, language and actions to have robots follow natural language instructions for all kinds of tasks [5]. A strength of the VLA model is that it is an all-in-one system for vision, language and action. In this way, the robot can directly connect the images it sees and the text commands it receives with real-world actions. It can help the robot better grasp the current situation of a new task and complete it [9]. The end-to-end VLA framework is less prone to the errors that occur when modules are used separately in the old robot system. It can also use the knowledge learned by large vision-language models to perform new tasks with little or no additional training [9].

But most existing VLA (Vision-Language-Action) models are still based on visual information and have not incorporated tactile information effectively. The current model lacks the perception of fine details in physical interaction, and to improve the accuracy of operational perception for VLAs, new modalities such as touch are being explored [5]. The M3Fusion framework proposed by Mao Xiaohan et al. builds unified scene-level multimodal representations, M3-tokens, through close integration of multi-view first-person 2D visual semantic features and 3D geometric features [10]. It has achieved an Acc@0.25 of 74.07% and an Acc@0.5 of 49.84% in 3D visual localisation tasks and verified the effectiveness of unified multimodal representations [10]. Some recent research has started to explore the use of tactile information as an observed variable in the world model to improve the generalisation ability of operational strategies [9, 11].

4.2. Tactile-force sensing fusion

Force sensing generally measures the interaction force at the wrist or joint, and tactile sensing provides distributed information of the contact area. Together, they can be used to learn about the working process in more detail. Direction of force is better at encoding task geometric information than the magnitude of force, and it is more robust to transfer from simulation to reality. This finding provides new ideas for strategy learning in contact-rich operations [8]. In soft-bodied humanoid dexterous hands, force-sensing information can be obtained indirectly via internal pressure changes in pneumatic actuators, and in conjunction with flexible tactile sensors at the fingertips, fine control of grasping force is achieved [6]. Gu et al. developed a soft-body neural prosthetic hand with flexible pressure sensors in the fingertips for tactile feedback and bidirectional closed-loop control of the nervous system [6].

DOI: 10.54254/2755-2721/2026.38119

4.3. Classification of fusion strategies

The three kinds of Multimodal Fusion are introduced. Directly merge the raw data to prevent information loss, and then carry out spatio-temporal alignment. First, extract features from all types of input at the feature level, then combine them. This way is flexible and can store the relevant information from all over. Decision-level fusion is performed on each type of input independently, and then these results are combined. It will be relatively stable; however, some of the beneficial information that different types of input shared or used to support each other will be lost. Mao Xiaohan and others proposed the M3Fusion framework for feature-level fusion. Its M3 fusion layer can match features from different views using 3D geometric information based on spatial position. Then it will organise the information and convert it into M3-token representations in a shared semantic and geometric space. This framework has good performance in 3D visual localisation and 3D question-answering tests [10]. According to the particular requirements of the application and hardware conditions, an appropriate fusion method needs to be selected.

4.4. Preliminary exploration of haptics and VLA model

Introducing haptics into the VLA model is a highly promising frontier direction. The core contradiction of the current VLA model lies in the decoupling between cognitive architecture and the physical world. The action concepts constructed in the multimodal model within the linguistic and visual spaces cannot escape the limitations of the original modality, making it difficult to truly understand physical concepts such as mass and force [9]. The main challenges currently faced include four aspects. Firstly, there is a lack of standardized haptic representations compatible with the Transformer architecture. Secondly, large-scale haptic datasets are far fewer than visual and linguistic datasets, and embodied datasets are extremely small in scale, with the largest open-source robot dataset, OpenX-Embodiment, containing less than 2 million trajectories [9]. Thirdly, there is a contradiction between the low latency required for haptic closed-loop control and the high computational overhead of large models. Fourthly, the insufficient accuracy of haptic simulation restricts strategy transfer [12]. The future development of VLA requires breakthroughs in cross-modal sensor information fusion at the perception level, endowing it with internalized physical common sense and dynamic deduction capabilities at the cognitive level, and achieving differentiation and unification between the brain and cerebellum at the execution level [9].

5. Discussion and outlook

5.1. Core research gaps

Through a literature review, this paper identifies the following specific gaps. Firstly, the long-term stability of sensors is insufficient. The performance degradation of most flexible sensors under cyclic loading and changes in temperature and humidity lacks systematic evaluation, hindering their practical application. Soft materials are more likely to be damaged by cutting and perforation from sharp objects, thus reducing the reliability of soft humanoid dexterous hands [6]. Although Terryn et al. have proposed that self-healing polymer materials can achieve self-repair of damaged areas via mild thermal stimulation, this technology has not yet been commercialised [6]. Secondly, there is no standardisation of the tactile characterisation. The format of the tactile data and the method of feature extraction are not unified, so it is difficult to adapt to basic model architectures, and the effective use of tactile information in large-scale pre-training is limited. The OmniCorpus is a text-

DOI: 10.54254/2755-2721/2026.38119

image dataset with tens of billions of entries, and the largest open-source robot dataset, OpenX-Embodiment, has fewer than 2 million trajectories; thus, there is a considerable difference in the volume of data [9]. Thirdly, there is no mature solution for the tactile-VLA interface. There is no mature solution for efficiently injecting tactile signals into the VLA model. Fourthly, there is a problem of real-time closed-loop control. Tactile feedback needs to respond in the order of milliseconds for stable grasping, but the inference delay of large models is still not low enough. The embodied ability of VLA is extremely lacking, and the spatial perception ability of multimodal large models is very weak. Even for the simplest problems of left and right distance, they often answer incorrectly [9]. Fifthly, there is difficulty in simulating to reality transfer. The accuracy of tactile simulators is much lower than that of visual simulators, making it difficult to directly deploy strategies trained based on simulation to real systems [8].

5.2. Future research directions

Based on the aforementioned gaps, this paper proposes the following directions worthy of further exploration. The first is self-powered and long-lived sensors. Develop self-powered sensors such as triboelectric sensors and enhance their stability and lifespan in real-world environments to reduce system power consumption and maintenance costs. Self-healing polymer materials provide new ideas for improving the reliability of soft dexterous hands [6]. The second is the fundamental tactile model. Drawing on the pre-training paradigm in the fields of vision and language, a large-scale tactile representation model is constructed to provide general feature extraction capabilities for various downstream tasks. The third is the VLA model for tactile perception. Designing novel network architectures and training paradigms to effectively integrate the tactile modality into the VLA framework enables joint reasoning of language-vision-action with the perception ability of physical interaction details. In the future, breakthroughs in cross-modal sensor information fusion at the perception level, such as the popularization of tactile sensors and bionic skin, are needed [9]. The fourth is the transfer of tactile perception from simulation to reality. Enhancing the physical accuracy of tactile simulation engines and developing domain adaptation and domain randomisation methods to reduce the perception gap between simulation and reality. The fifth is how to build a system with good perception-action coupling. Tactile information is directly related to dexterous hand movements in this system. It will be convenient to reduce the delay of different processing steps and prevent information loss during transmission.

6. Conclusion

This article mainly introduces three aspects of tactile perception for dexterous hands: how tactile sensors work, where the sensors should be placed, and how different kinds of information can be connected and used together. The five main types of tactile sensors have their own advantages and limits, so no single sensor can meet all application needs. Because of this, combining different sensors is becoming more and more important. The placement of sensors has also expanded from the fingertips to the whole hand, and from rigid structures to soft structures. But limited space and stability after long-term use are still major problems in practical applications. The vision-language-action model, or VLA model, provides a new way to connect tactile information with high-level decision-making. But at present, tactile information is still not widely integrated into this model. The main problems include poor long-term sensor stability, the lack of a standard way to represent tactile data, and delays in tactile control. In the future, research can continue in areas such as self-powered

DOI: 10.54254/2755-2721/2026.38119

sensors, basic models and simulation for tactile perception, and how to transfer methods from simulation to real systems.

References

[1] Duan, J., Yu, S., Tan, H. L., Zhu, H., & Tan, C. (2022). A survey of embodied AI: From simulators to research tasks (arXiv:2103.04918). *arXiv*. https://doi.org/10.48550/arXiv.2103.04918

[2] Gao, S., Liang, W., Zheng, K., Malik, A., Ye, S., Yu, S., Tseng, W.-C., Dong, Y., Mo, K., Lin, C.-H., Ma, Q., Nah, S., Magne, L., Xiang, J., Xie, Y., Zheng, R., Niu, D., Tan, Y. L., Zentner, K. R., … Fan, L. J. (2026). DreamDojo: A generalist robot world model from large-scale human videos (arXiv:2602.06949). *arXiv*. https://doi.org/10.48550/arXiv.2602.06949

[3] Hassani, S., & Dackermann, U. (2023). A systematic review of advanced sensor technologies for non-destructive testing and structural health monitoring. *Sensors, 23*(4), 2204. https://doi.org/10.3390/s23042204

[4] Kadalagere Sampath, S., Wang, N., Wu, H., & Yang, C. (2023). Review on human-like robot manipulation using dexterous hands. *Cognitive Computation and Systems, 5*(1), 14–29. https://doi.org/10.1049/ccs2.12073

[5] Liu, Y., Chen, W., Bai, Y., Liang, X., Li, G., Gao, W., & Lin, L. (2025). Aligning cyber space with physical world: A comprehensive survey on embodied AI (arXiv:2407.06886). *arXiv*. https://doi.org/10.48550/arXiv.2407.06886

[6] Ma, Y., Song, Z., Zhuang, Y., Hao, J., & King, I. (2026). A survey on vision-language-action models for embodied AI. *IEEE Transactions on Neural Networks and Learning Systems, 1*–21. https://doi.org/10.1109/TNNLS.2025.3650584

[7] Pai, J., Achenbach, L., Montesinos, V., Forrai, B., Mees, O., & Nava, E. (2025). Mimic-video: Video-action models for generalizable robot control beyond VLAs (arXiv:2512.15692). *arXiv*. https://doi.org/10.48550/arXiv.2512.15692

[8] Qiu, K., Zhang, H., Zhang, J., Xiong, R., Lu, H., & Wang, Y. (2025). An actuator space optimal kinematic path tracking framework for tendon-driven continuum robots: Theory, algorithm and validation. *The International Journal of Robotics Research, 44*(6), 1006–1034. https://doi.org/10.1177/02783649241290525

[9] Mao, X., Zhu, C., Wang, T., & Pang, J. (2026). M3Fusion: A unified multi-view and multi-modal fusion framework for embodied 3D perception. *Chinese Science: Information Sciences, 56*(2), 312–326. https://www.sciengine.com/sci-open/api/v1/open/file/pdf/24258B6FBB1E44E286991BC0C6A8F7B1

[10] Tang, C., Tang, W., Li, C., & Yu, W. (2024). Design and performance of the tactile perception system for bionic fingers. *Journal of Tsinghua University (Natural Science Edition), 64*(3), 421–431. https://www.sciopen.com/local/article_pdf/10.16511/j.cnki.qhdxxb.2023.26.054.pdf

[11] Wang, J., Zhang, W., & Liu, T. (2025). The past and present of the visual-language-action (VLA) model. *Computations, 1*(4), 22–27. https://cccf.hrbeu.edu.cn/cn/article/pdf/preview/10.11991/cccf.202508005.pdf

[12] Liu, Y., Xiao, H., Liu, C., Sun, Z., Hao, T., & Liu, S. (2025). Research progress on soft humanoid dexterous hand. *Journal of Mechanical Transmission, 49*(6), 159–176. https://www.jxcd.net.cn/zh/article/78782384/