Research Progress on Technologies and Applications of Human-Computer Interaction Gestures Based on Vision and Sensors

Linhao Li

Department of Computer Science, Southwestern University of Finance and Economics, Chengdu, China

huh282013847@gmail.com

Abstract. Against the backdrop of rapid Artificial Intelligence (AI) and Internet of Thing (IoT) advancement, traditional HCI modalities (keyboards, mice) show limitations, spurring attention to intuitive gesture interaction—a core part of NUIs. This technology mainly has two routes. Vision-based systems use cameras, with deep learning (e.g., YOLOv5 for hand detection, CNN for feature extraction) enabling tasks like keypoint and spatio-temporal feature analysis, and real-time recognition via stereo image fusion. Sensor-based systems include wearable (EMG sensors, smart gloves) and non-wearable (Leap Motion, Kinect, WDHS) devices, with multi-sensor fusion enhancing robustness; bioimpedance measurement is an emerging non-wearable method. These technologies apply to intelligent HCI, disability assistance, VR/AR, and healthcare but face challenges like poor environmental robustness and high computing demands, playing a key role in advancing HCI. Looking forward, gesture recognition is expected to evolve toward more natural, adaptive, and multimodal interaction modes by integrating speech, eye tracking, and emotion recognition for context-aware communication.

Keywords: Human-Computer Interaction; Gesture Interaction; Vision-based Gesture Recognition.

1. Introduction

With the rapid advancement of artificial intelligence (AI) and the Internet of Things (IoT) technologies, human-computer interaction (HCI) technology—serving as a bridge connecting humans and the digital world—is undergoing a revolutionary transformation. While traditional human-computer interaction (HCI) modalities such as keyboards, mice, and touchscreens are well-established and reliable, they exhibit inherent limitations in certain scenarios, failing to meet the increasingly diversified interaction requirements. Gesture interaction, as a natural and intuitive interaction modality, has garnered widespread attention in recent years. Gesture interaction technologies based on vision and sensors enable a more natural and efficient human-computer interaction experience by capturing and interpreting human gestural movements. As a crucial component of natural user interfaces (NUIs), these technologies have achieved remarkable technological breakthroughs in recent years, thereby demonstrating significant research value and promising application prospects [1].

The development of gesture interaction technology originates from the emulation and extension of humans' natural communication modalities. In daily communication, humans extensively utilize gestures to convey information, express emotions, and enhance communication efficacy. The introduction of this natural interaction modality into the field of human-computer interaction (HCI) can reduce learning costs and enhance interaction efficiency, and it exhibits unique advantages particularly in emerging application scenarios such as virtual reality (VR), augmented reality (AR), and smart homes. Vision-based gesture recognition technology relies primarily on cameras to capture gesture images, which are then analyzed and recognized via computer vision algorithms. In contrast, sensor-based gesture recognition leverages devices such as accelerometers, gyroscopes, and electromyographic (EMG) sensors to collect gesture motion data. These two technical routes each possess distinct characteristics, complement one another, and collectively drive the advancement of gesture interaction technology [2, 3].

Investigating vision-based and sensor-based gesture interaction technologies holds multifaceted significance. First and foremost, it expands the modalities of human-computer interaction (HCI), rendering the interaction process more natural and intuitive, which aligns with human cognitive habits. Secondly, this technology can be applied to a variety of specialized scenarios, such as situations in virtual reality (VR) environments that require free hand operations [4], or scenarios in medical settings that demand aseptic operations [5]. Furthermore, gesture interaction technology offers new interaction possibilities for people with disabilities [6], contributing to bridging the digital divide. Finally, with the popularization of the Internet of Things (IoT) and smart devices, gesture interaction—as a contactless interaction modality—has also demonstrated unique value in terms of public health. Consequently, in-depth research on vision-based and sensor-based gesture interaction technologies carries significant implications for advancing the development of the human-computer interaction (HCI) field and meeting diverse application requirements.

2. Progress in the Technical Field

As a crucial component of natural user interfaces (NUIs), gesture recognition technology is mainly categorized into two types: vision-based methods and sensor-based device methods [7]. While traditional gesture recognition methods rely primarily on rectangular windows for training—resulting in limited detection accuracy [8]—in recent years, with the advancement of deep learning and multi-modal fusion technology, significant technological breakthroughs have been brought about for gesture recognition [9, 10].

2.1. Progress in Vision-Based Technology

A vision-based gesture recognition system acquires images or video streams via a camera and utilizes computer vision algorithms for gesture detection, tracking, and recognition. This method features the characteristics of being natural and intuitive, and it does not require users to wear additional devices—thus, it has garnered widespread attention [11].

2.1.1 Hand Detection and Tracking

Hand detection is the first step in gesture recognition, and it aims to accurately locate hand regions from images or videos [10, 12]. In real-world scenarios such as complex backgrounds, lighting variations, and hand occlusions, efficiently and accurately detecting and tracking bare hands remains a challenge [8]. Traditional methods typically use rectangular windows for hand detection, but they achieve poor performance when handling the diversity of hand appearances [8].

Modern approaches leverage deep learning models, such as YOLOv5, to achieve robust hand detection [13]. For instance, the HGR-FYOLO system utilizes a frozen YOLOv5 model to improve gesture recognition performance for both individuals with normal limb function and those with limb impairments—particularly excelling in handling adverse lighting and complex backgrounds [13]. Additionally, some studies enhance the accuracy of hand detection by analyzing keypoints of the overall human body, body parts, and bare hands in images [8] (Fig. 1).

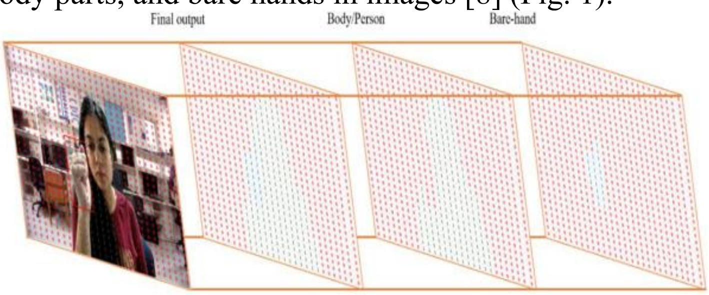

Fig. 1 It illustrates the detection process—starting from the overall image, moving to various body parts, and then to bare hand keypoints—demonstrating the critical role of accurately locating the hand in gesture recognition [8].

2.1.2 Gesture Feature Extraction

Gesture Feature Extraction is the core link in gesture recognition, and it aims to extract key information that can distinguish different gestures from the detected hand regions [9].

Pose Estimation and Keypoint Detection: The human hand has multiple joints, making its pose estimation complex. To address this, researchers have proposed various frameworks; for example, the HandyPose system implements hand pose estimation by leveraging [14] (supplementary technical details) and uses hand keypoints (e.g., cascaded fingertips, palm centers) to achieve accurate gesture recognition [15]. These keypoints can form a skeletal representation of the hand, which serves as the basis for gesture recognition.

Spatio-Temporal Feature Extraction: For dynamic gestures, in addition to spatial features, temporal features are equally important [16, 17]. For example, the American Sign Language (ASL) finger spelling recognition system combines MobileNetV3-Small for spatial feature extraction and extracts short-range and long-range temporal features via a Variable-Length Filter Temporal Convolutional Neural Network (VTCNN) to capture robust representations of dynamic gestures [16]. In addition, 3D hand trajectories and geometric features (such as the number of inflection points, quadrant sequences, and curvature) are also used for gesture recognition [17].

Multi-Modal Feature Fusion: Fusing multi-modal features from RGB images and depth images can improve the accuracy of gesture recognition [9]. For example, some studies have proposed a deep learning-based end-to-end hybrid system for hand detection and gesture recognition, which enhances recognition performance by integrating information from different modalities [12].

2.1.3 Application of Deep Learning Models

Deep learning plays a crucial role in the field of gesture recognition, particularly convolutional neural networks (CNNs) [10, 17]. CNNs can automatically learn high-level features from raw image data, eliminating the tediousness of manual feature design in traditional methods [8, 17].

CNN and SIFT: In Pakistani Sign Language (PSL) recognition, combining CNN and the SIFT (Scale-Invariant Feature Transform) algorithm can effectively improve the recognition rate [17]. SIFT keypoints play an important role in various computer vision applications such as image matching, gesture recognition, autonomous driving, face recognition, and image stitching [19] (Fig. 2).

Fig. 2. It illustrates the wide application of keypoint technology in fields such as 3D reconstruction, image matching, and gesture recognition, and other fields—highlighting the importance of feature extraction methods like SIFT [19]

Multi-Task Learning: For complex sign language recognition tasks, Multi-Task Learning (MTL) can handle multiple related tasks simultaneously, thereby enhancing overall performance [16]

Real-Time Recognition: To achieve real-time human-computer interaction, researchers are committed to developing efficient deep learning models [10, 20]. For example, a real-time and robust visual gesture recognition method using stereo images improves recognition accuracy by fusing information from left and right images and adopting rule-based methods [20].

2.2. Progress in Sensor-Based Technology

A sensor-based gesture recognition system captures gesture data via wearable or non-wearable sensors, and then analyzes and recognizes the data [21].

2.2.1 Wearable sensors

Wearable sensors are typically worn directly on the user's body, such as data gloves and EMG sensors, among others [22].

Electromyographic (EMG) Sensors: Electromyography (EMG) measures the electrical activity of muscles, and gesture recognition is performed via machine learning algorithms [21]. Despite the significant progress made by multimodal theories and machine learning algorithms in the field of EMG interfaces, the collection and annotation of large amounts of data remain labor-intensive and time-consuming [21].

Smart Gloves: Smart gloves can detect a wide range of gestures and synchronize signals to smartphones or other devices via Bluetooth and other means to enable human-computer interaction [23] (Fig. 3).

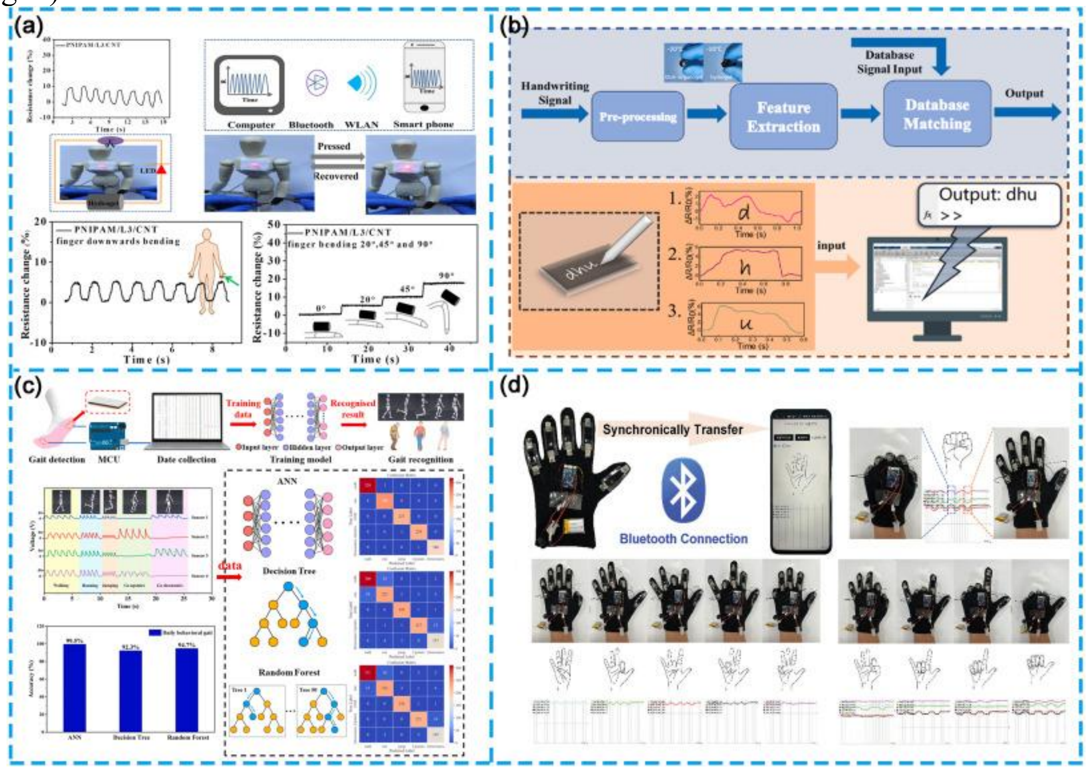

Fig. 3 It illustrates various gesture poses detected by smart gloves and the process of synchronized signal transmission—highlighting their application potential in gesture recognition [23].

Multi-Channel Sensor Fusion: To improve recognition robustness, researchers fuse and use multiple types of sensors, such as EMG sensors, 3-axis accelerometers, and 3-axis gyroscopes, along with other sensors [22]. For example, a system designed for Indian Sign Language (ISL) recognition acquires data using multi-channel surface electromyography (sEMG), 3-axis accelerometers, and 3-axis gyroscopes placed on both forearms [22].

2.2.2 Non-Wearable Sensors

Non-wearable sensors include depth cameras, radars, and Wi-Fi-based sensors (utilizing Wi-Fi signals), among others—these sensors do not require users to wear any devices to capture gesture information [24].

Leap Motion: Leap Motion is a high-precision hand-tracking device that can provide detailed data such as fingertip direction and position, and is commonly used for gesture recognition and pose tracking [21, 25]. It achieves precise capture of hand poses through the conversion between its own coordinate system (RF_L) and the hand coordinate system (RF_H) [21] (Fig. 4).

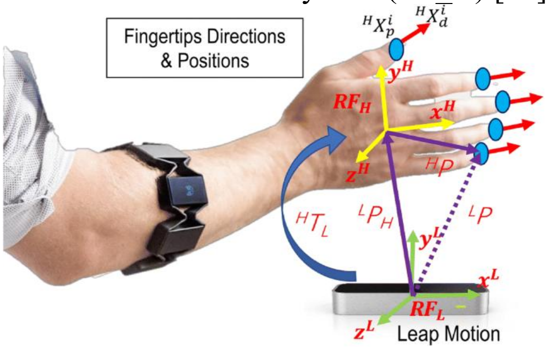

Fig. 4 It illustrates how the Leap Motion device tracks fingertip position and direction through coordinate system conversion—highlighting that this process is key to its precise hand gesture capture [21]

However, despite the excellent performance of devices such as the Leap Motion Controller 2 in hand pose recognition, capturing diverse hand poses—especially under multi-view and complex backgrounds—still requires the support of high-quality datasets [15]. To address this, some studies have proposed a multi-view Leap2 Hand Pose Dataset (ML2HP Dataset), which uses two Leap Motion Controller 2 devices to capture multi-angle and diverse hand poses [15].

Kinect Sensors: Kinect sensors are widely used in motion-sensing games. They track and analyze human movements using computer vision technology to achieve gesture recognition [26].

Wireless Device-Independent Human Sensing (WDHS): Wireless Device-Independent Human Sensing (WDHS) technology has made significant progress over the past decade, covering a wide range of applications from activity recognition to vital sign monitoring [24]. It detects human presence and movement through wireless signals (such as Wi-Fi and radar), enabling three major categories of sensing tasks: behavior recognition, motion tracking, and user identification [24]. Specifically, WDHS can recognize various limb movements (including gestures) and perform full-body motion tracking as well as vital sign monitoring [24].

2.3. Bioimpedance Measurement

Bioimpedance measurement is an emerging non-wearable gesture recognition method [27]. For example, a novel dual-electrode frequency-sweeping gesture recognition system based on bioimpedance measurement increases the diversity of impedance data by collecting data from the back of the hand and adopting a frequency-sweeping method—thus achieving high-precision recognition of common gestures and pinch gestures while reducing measurement complexity and the number of electrodes [27]..

3. Progress in Application Fields

Vision-based and sensor-based gesture recognition technologies have demonstrated great application potential in multiple fields, driving the innovation and development of human-computer interaction.

3.1. Intelligent Human-Computer Interaction

Gesture recognition is an important approach to realizing natural human-computer interaction (NHCI) [27-29].

Traditional PC Replacement: Compared with the traditional keyboard-and-mouse interaction mode, gesture recognition can provide a more natural, flexible, and intuitive way of information transmission [26].

Industrial Control: In industrial cyber-physical systems (ICPS), gesture recognition technology can enhance human-robot collaboration and improve production efficiency [30]. For example, gesture recognition can be used to control the speed or task mode of collaborative robots (cobots), such as "collab. in Task 1" or "speed control". During the interaction between humans and cobots, gesture detection is a key link to achieving safe and intuitive human-robot collaboration [30].

Smart Homes and Wearable Devices: In smart homes, gesture recognition technology can be used to control various smart devices, realizing a more convenient life experience [31]. In the field of wearable devices, gesture recognition also provides users with new interaction methods—for instance, controlling smartwatches or augmented reality (AR) glasses via gestures [31].

3.2. Assisting People with Disabilities

Sign language recognition is the most direct application of gesture recognition technology in assisting people with disabilities [17, 22].

Sign Language Recognition Systems: Efficient sign language recognition systems can help deaf and hard-of-hearing people communicate with hearing people [22]. For example, the Indian Sign Language (ISL) recognition system realizes the recognition of new users' sign language through multi-channel sensor data and transfer learning algorithms [22]. Meanwhile, the Pakistani Sign Language (PSL) recognition system uses CNN and SIFT algorithms, which significantly improves the performance of sign language recognition [17].

Hand Rehabilitation Training: Gesture recognition technology can be used to monitor and evaluate the hand movements of patients with limb impairments, assisting in rehabilitation training [24].

Interaction for People with Limb Impairments: For people with limb impairments who have a small number of fingers, robust gesture recognition systems such as HGR-FYOLO enable them to interact with desktop applications more easily [13].

3.3. Virtual Reality (VR) and Augmented Reality (AR)

Gesture recognition is a key technology for achieving immersive interaction in the VR/AR field [32].

Gaming and Entertainment: In VR games, gesture recognition enables players to interact with the virtual world through natural gestures, enhancing the game's immersion and authenticity [26]. For example, ARGo—a mobile Go collision game based on augmented reality—simulates the interactive experience of real Go through gesture recognition33].

Virtual Object Manipulation: In VR/AR environments, users can grab, move, and rotate virtual objects through gestures, enabling intuitive manipulation [22]. For instance, a gesture of moving an open palm to the right can represent horizontal scrolling, hands pressed together can represent pausing, and both hands pointing downward can indicate vertical scrolling [22] (Fig. 5).

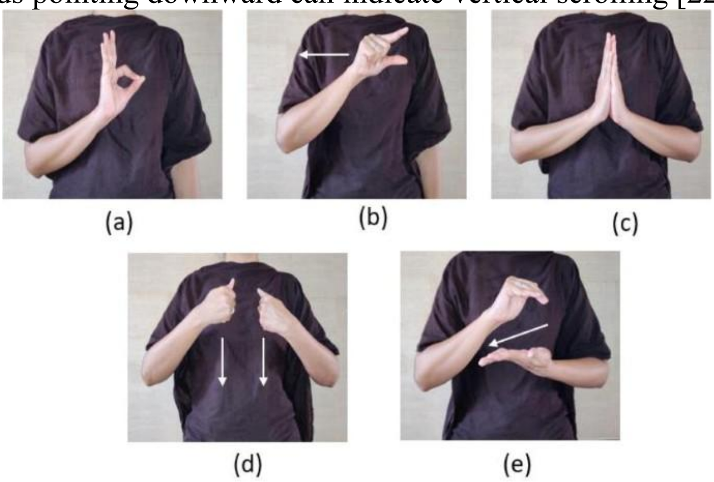

Fig. 5 illustrates various gestures and the meanings they may represent in different application scenarios—highlighting the diversity of gestures in virtual interaction [22]

3.4. Medical Healthcare

The application of gesture recognition in the medical field is mainly reflected in aspects such as aided diagnosis, rehabilitation training, and telemedicine [34]. For example, in Magnetic Resonance Imaging (MRI) image enhancement: Despite the fact that gesture recognition itself does not directly enhance MRI images, it can help doctors operate image display systems more efficiently in medical environments to conduct rapid diagnosis [34]. Specifically, doctors can adjust contrast or fuse images via gestures to improve the visual effect and quality of medical images [34].

3.5. Other Applications

Gesture recognition technology is also widely applied in fields such as autonomous driving, security monitoring, face recognition, vehicle tracking, and object recognition, among others [22]. Furthermore, in the construction of smart cities, the combination of self-powered sensor technologies (e.g., triboelectric nanogenerators, TENGs) with gesture recognition can support the digitalization and decentralized development of smart cities [31].

4. Future Outlook and Issues to Be Addressed

Despite the significant progress made in gesture recognition technology, it still faces numerous challenges and development opportunities.

Robustness Issues: In complex and variable real-world environments, gesture recognition systems still suffer from insufficient robustness. For example, factors such as lighting variations, complex backgrounds, hand occlusions, individual differences (e.g., skin tone, hand shape), different viewing angles, and motion blur can all significantly affect recognition accuracy [8, 13, 36]. Addressing these issues requires more robust feature extraction and model generalization capabilities.

Diversity and Ambiguity of Gestures: Gestures exhibit a wide variety, and they may carry different meanings in different cultural backgrounds; even the same gesture can be ambiguous in different contexts [35]. How to effectively recognize and understand the semantics of gestures remains a long-term challenge for gesture recognition technology.

Real-Time Performance and Computational Resources: For many application scenarios (e.g., VR/AR, industrial control), gesture recognition requires high real-time performance [10, 20]. However, complex deep learning models often demand substantial computational resources, which poses a challenge in embedded devices or resource-constrained environments. Future research needs to explore the design of lightweight and efficient models [13].

Data Annotation and Datasets: High-quality, diverse gesture datasets are crucial for training deep learning models [15, 21]. Nevertheless, the collection and annotation of gesture data are labor-intensive and time-consuming, which limits the further development of models [21]. It is necessary to develop more efficient data augmentation techniques and unsupervised/semi-supervised learning methods.

User Experience and Personalization: How to design gesture interaction methods that are more in line with user habits, more intuitive, and more personalized is key to improving user experience [39]. For example, heuristic research can be used to identify user-preferred gesture sets, while considering the physical and cognitive characteristics of different users [39].

Multimodal Fusion Optimization: Although multimodal fusion has been proven to improve gesture recognition performance, in-depth research is still needed on how to better fuse data from different modalities (such as visual, depth, inertial sensors, and electromyography) and address the heterogeneity between modalities [9, 21].

Privacy and Security; In public places or sensitive applications, gesture recognition may involve issues related to user privacy and data security. For instance, machine learning algorithms on touchscreen devices can enhance security, but gesture recognition may also bring new security challenges—thus requiring a balance between convenience and security [36].

More Natural Interaction Modes: With technological development, gesture recognition will further integrate with multiple modalities such as speech, eye tracking, and haptic feedback to achieve a more natural and seamless human-computer interaction experience. For example, combining with emotion recognition, the system can better understand the user's intent and emotions [18].

Ubiquitous Gesture Recognition: Future gesture recognition will no longer be limited to specific devices or environments, but will enable "anytime, anywhere" ubiquitous sensing. Wireless Device-Independent Human Sensing (WDHS) technology—such as sensing based on Wi-Fi or radar—will play a greater role in achieving device-independent fine-grained gesture recognition [24].

Adaptive and Personalized Learning: Systems will be able to dynamically adjust recognition models based on users' habits, preferences, and learning curves, providing personalized gesture interaction experiences [39]. For instance, one-shot gesture learning methods allow the system to quickly adapt to gestures from new users or in new scenarios [39].

Deep Integration with Artificial Intelligence: Gesture recognition will be deeply integrated with more advanced artificial intelligence (AI) technologies, such as Natural Language Processing (NLP) and context awareness. This will enable systems to understand the intent and context behind gestures, achieving more intelligent decision-making and responses [37].

Advanced Materials and Sensor Technologies; Advanced materials and sensor technologies—such as new conductive hydrogels and triboelectric nanogenerators (TENGs)—will bring more compact, more sensitive, and lower-power-consumption solutions to gesture recognition [23, 31]

Expansion of Cross-Domain Applications: Gesture recognition will be applied in more vertical fields, such as surgical assistance, remote collaboration, artistic creation, and cultural heritage preservation [38].

5. Conclusion

In summary, vision-based and sensor-based gesture interaction has made great progress and shown strong potential in fields like VR, smart homes, healthcare, and industrial control. With the help of deep learning and advanced sensors, recognition accuracy and robustness have improved, supporting more natural and efficient human-computer interaction. However, challenges such as environmental adaptability, real-time performance, gesture ambiguity, and data privacy still remain. Future research will focus on lightweight models, multimodal fusion, and user-centered design. As technologies like edge computing and smart sensors mature, gesture interaction will become more natural, intelligent, and widely used in everyday life.

References

[1] Ga W, Qi M, Ma M, Wang L, Yang C, Liu J, Bian Y, de Melo G, Liu S, Meng X. Employing shadows for multi-person tracking based on a single RGB-D camera. Sensors. 2020;20(4):1056.

[2] Zhang Y, Tian Y, Kong Y, Zhong B, Fu Y. Residual dense network for image restoration. IEEE Trans. Pattern Anal. Mach. Intell. 2021;43(7):2480–2495.

[3] Lu W, Wan G, Zhou Y, Fu X, Yuan P, Song S. DeepVCP: An end-to-end deep neural network for point cloud registration. In: Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV). 2019. p. 12–21.

[4] Tashakori A, Jiang Z, Servati A, Soltanian S, Narayana H, Le K, Nakayama C, Yang C, Wang ZJ, Eng JJ, Servati P. Capturing complex hand movements and object interactions using machine learning-powered stretchable smart textile gloves. Nat. Mach. Intell. 2024;6(1):106–118.

[5] Mangalam M, Oruganti S, Buckingham G, Borst CW. Enhancing hand-object interactions in virtual reality for precision manual tasks. Virtual Real. 2024;28(4).

[6] de Souza LR, Francisco R, da Rosa Tavares JE, Barbosa JLV. Intelligent environments and assistive technologies for assisting visually impaired people: a systematic literature review. Universal Access Inf. Soc. 2024.

[7] Oudah M, Al-Naji A, Chahl J. Hand gesture recognition based on computer vision: a review of techniques. J. Imaging. 2020;6(8):73.

[8] Yadav KS, KM, Laskar RH, Ahmad N. Exploration of deep learning models for localizing bare-hand in the practical environment. Eng. Appl. Artif. Intell. 2023;123:106253.

[9] Duan H, Sun Y, Cheng W, Jiang D, Yun J, Liu Y, Liu Y, Zhou D. Gesture recognition based on multi-modal feature weight. Concurrency Comput. Pract. Exp. 2020;33(5).

[10] Sen A, Mishra TK, Dash R. Deep learning-based hand gesture recognition system and design of a human–machine interface. Neural Process. Lett. 2023;55(9):12569–12596.

[11] Wang X, Jiang J, Wei Y, Kang L, Gao Y. Research on gesture recognition method based on computer vision. MATEC Web Conf. 2018;232:03042.

[12] Mohammed AAQ, Lv J, Islam MDS. A deep learning-based end-to-end composite system for hand detection and gesture recognition. Sensors. 2019;19(23):5282.

[13] Farid FA, Hashim N, Abdullah JB, Bhuiyan MR, Kairanbay M, Yusoff Z, Karim HA, Mansor S, Sarker MT, Ramasamy G. Single shot detector CNN and deep dilated masks for vision-based hand gesture recognition from video sequences. IEEE Access. 2024;12:28564–28574.

[14] Gupta D, Artacho B, Savakis A. HandyPose: Multi-level framework for hand pose estimation. Pattern Recognit. 2022;128:108674.

[15] Gil-Martín M, Marini MR, San-Segundo R, Cinque L. Dual Leap Motion Controller 2: A robust dataset for multi-view hand pose recognition. Sci. Data. 2024;11(1).

[16] Pannattee P, Kumwilaisak W, Hansakunbuntheung C, Thatphithakkul N, Kuo CCJ. American sign language fingerspelling recognition in the wild with spatio-temporal feature extraction and multi-task learning. Expert Syst. Appl. 2024;243:122901.

[17] Arooj S, Altaf S, Ahmad S, Mahmoud H, Mohamed ASN. Enhancing sign language recognition using CNN and SIFT: a case study on Pakistan sign language. J. King Saud Univ. Comput. Inf. Sci. 2024;36(2):101934.

[18] Clark A, Ahmad I. Touchless and nonverbal human-robot interfaces: an overview of the state-of-the-art. Smart Health. 2023;27:100365.

[19] Xie C, Liu Q, Chen B, Hao Z. Evaluation and analysis of feature point detection methods based on vSLAM systems. Image Vis. Comput. 2024;146:105015.

[20] Liu K, Kehtarnavaz N. Real-time robust vision-based hand gesture recognition using stereo images. J. Real-Time Image Process. 2013;11(1):201–209.

[21] Zhou X, Qi W, Ovur SE, Zhang L, Hu Y, Su H, Ferrigno G, De Momi E. A novel muscle-computer interface for hand gesture recognition using depth vision. J. Ambient Intell. Humaniz. Comput. 2020;11(11):5569–5580.

[22] Sharma S, Gupta R, Kumar A. Trbaggboost: an ensemble-based transfer learning method applied to Indian Sign Language recognition. J. Ambient Intell. Humaniz. Comput. 2020;13(7):3527–3537.

[23] Zhang S, Zhao B, Zhang D, Yang M, Huang X, Han L, Chen K, Li X, Pang R, Shang Y, Cao A. Conductive hydrogels incorporating carbon nanoparticles: a review of synthesis, performance and applications. Particuology. 2023;83:212–231.

[24] Xiao J, Li H, Wu M, Jin H, Deen MJ, Cao J. A survey on wireless device-free human sensing: application scenarios, current solutions, and open issues. ACM Comput. Surv. 2022;55(5):1–35.

[25] Yongda D, Fang L, Huang X. Research on multimodal human-robot interaction based on speech and gesture. Comput. Electr. Eng. 2018;72:443–454.

[26] Lv W. Gesture recognition in somatosensory game via Kinect sensor. Internet Technol. Lett. 2021;6(5).

[27] Ma G, Chen H, Wang P, Dong S, Wang X. A two-electrode frequency-scan system for gesture recognition. Mechatronics. 2023;94:103039.

[28] Negi PS, Pawar R, Lal R. Vision-based real-time human–computer interaction on hand gesture recognition. In: Lect. Notes Netw. Syst. Singapore: Springer; 2020. p. 499–507.

[29] Jiao J, Zhang Q, You J. Research on human-computer interaction technology based on visual user gesture recognition. In: Proc. 3rd Int. Conf. Mechatron. Eng. Inf. Technol. (ICMEIT 2019). 2019.

[30] Piardi L, Leitão P, Queiroz J, Pontes J. Role of digital technologies to enhance the human integration in industrial cyber–physical systems. Annu. Rev. Control. 2024;57:100934.

[31] Tang Y, Fu H, Xu B. Advanced design of triboelectric nanogenerators for future eco-smart cities. Adv. Compos. Hybrid Mater. 2024;7(3).

[32] Mat Sanusi KA, Majonica D, Iren D, Fanchamps N, Klemke R. MILSDeM: Guiding immersive learning system development and taxonomy evaluation. Educ. Inf. Technol. 2024;29(13):16283–16316.

[33] Lee D, Won S, Kim J, Kwon HY. ARGo: augmented reality-based mobile Go stone collision game. Virtual Real. 2024;28(1).

[34] Zhu R, Li X, Zhang X, Xu X. MRI enhancement based on visual-attention by adaptive contrast adjustment and image fusion. Multimed. Tools Appl. 2020;80(9):12991–13017.

[35] Chakraborty BK, Sarma D, Bhuyan MK, MacDorman KF. Review of constraints on vision-based gesture recognition for human–computer interaction. IET Comput. Vis. 2017;12(1):3–15.

[36] Bello AA, Chiroma H, Gital AY, Gabralla LA, Abdulhamid SM, Shuib L. Machine learning algorithms for improving security on touch screen devices: a survey, challenges and new perspectives. Neural Comput. Appl. 2020;32(17):13651–13678.

[37] Jirak D, Biertimpel D, Kerzel M, Wermter S. Solving visual object ambiguities when pointing: an unsupervised learning approach. Neural Comput. Appl. 2020;33(7):2297–2319.

[38] Kerdvibulvech C. A review of computer-based gesture interaction methods for supporting disabled people with special needs. In: Lect. Notes Comput. Sci. Cham: Springer; 2016. p. 503–506.

[39] Cabrera ME, Wachs JP. A human-centered approach to one-shot gesture learning. Front. Robot. AI. 2017;4.