IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 10 November 2023, accepted 1 January 2024, date of publication 8 January 2024, date of current version 12 January 2024.

*Digital Object Identifier 10.1109/ACCESS.2024.3350338*

SURVEY

# A Decade of Progress in Human Motion Recognition: A Comprehensive Survey From 2010 to 2020

**DONGHYEON NOH, HOJIN YOON, AND DONGHUN LEE**
Mechanical Engineering Department, Soongsil University, Seoul 06978, South Korea

Corresponding author: Donghun Lee (dhlee04@gmail.com)

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2023 R1F1A1074704); Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2023-0-00218); the MSIT (Ministry of Science and ICT), Korea, under the Innovative Human Resource Development for Local Intellectualization support program (IITP-2023-RS-2022-00156360) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation); Korea institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE) (P0017123, HRD Program for Industrial Innovation).

**ABSTRACT** The central perspective of this review is to categorize research in Human Motion Recognition (HMR) over the past decade into two significant categories: vision sensor-based (VS) methods and wearable sensor-based (WS) methods. Within each category, research is further assessed from the viewpoints of sensors, classification algorithms, datasets, gesture types, target body parts, and performance. This approach allows for a comprehensive assessment of the overall research trends and technological advancements in HMR. Both VS methods and WS methods present their own sets of advantages and challenges. VS methods face challenges related to limited workspace, varying lighting conditions, occlusion, and complex image processing. Conversely, WS methods, compared to VS methods, deals with challenges associated with multiple sensor calibration, intrusiveness, and magnetic field mapping due to sensor placement. As such, the choice between these methods depends on the specific application, the required level of accuracy, and user preferences. Gaining insights into the nature of various HMR methods and staying informed about recent research trends is of utmost importance. By the end of this review, readers will gain a comprehensive and systematic understanding of the latest developments in HMR techniques, which will serve as a valuable resource for researchers and practitioners alike.

**INDEX TERMS** Human gesture recognition, human motion recognition, human-robot interaction, wearable sensor, vision sensor.

## I. INTRODUCTION

With the widespread adoption of smart factories, the industrial sector is increasing the intelligence and autonomy of objects related to manufacturing, procurement, logistics, and consumers and linking them organically through the Internet of Things to accelerate the refinement of autonomous data connection, collection, and analysis systems. This acceleration of digitalization in manufacturing processes is inevitably linked to expanding automation and enhancing operability performance. Recently, the computing power of controllers has dramatically improved based on GPU technology, and perception technologies based on sensors and deep learning are also continuously developing. However, the problem of uncertainty in the control and recognition of industrial robots must be addressed. Therefore, in cluttered circumstances where it is difficult to achieve complete automation and in the case of challenging automation tasks that require high operability and a wide range of work areas, there is an increasing interest in human-robot interaction (HRI) or human-robot collaboration (HRC) technology [1], [2], [3], [4] where humans actively intervene as supervisors for the robot's tasks [5], [6], [7], [8]. Moreover, in the face of these challenges, motion recognition technology plays a crucial

The associate editor coordinating the review of this manuscript and approving it for publication was Tao Liu.

role in enhancing the control and recognition of industrial robots. HMR can be broadly categorized into two main categories: VS methods and WS methods. VS methods is further subdivided into marker-based and markerless VS methods.

VS methods utilize camera video sequences to analyze human motion. This approach can be further divided into marker-based and markerless methods. The first marker-based method was Murray et al. [9] analyzed "Walking Patterns of Normal Men" in 1964 using a photographic method based on a reflective marker, a kind of passive marker. Later, in 1982, Tayler et al. [10] proposed an automated motion measurement system capable of detecting and tracking passive markers illuminated by infrared LEDs to reduce patient discomfort and distraction caused by wearing a marker telemetry backpack system in clinical gait analysis. Later, this marker-based method was mainly used in research on human factors in the 1970s and 1980s and various fields such as computer visualization, sport science, ergonomics, and human-computer interaction using 3D motion capture technology in the 1990s and 2000s. And the markerless gesture recognition research using a vision sensor was first used in human-computer interaction (HCI), which emerged with the spread of personal computers in the late 1970s and early 1980s. Hanson et al. [33] stated that web camera interface-based gesture recognition and control research began in the early 1990s to help disabled people. In addition, Yamato [34]'s 1992 study on the proposal of a human behavior recognition method using image-based data, and HMM is the most famous paper as the origin of hand gesture recognition research. The primary sensors of markerless motion methods are camera-type vision sensors such as single cam (RGB or depth), multi (RGB+depth) cam, Time of flight (ToF) cam, and IR camera. Before 2010, research using a single RGB web camera was the main focus, and sometimes studies using multiple RGB and depth cameras were also conducted. Both methods are used for motion capture and gesture recognition, and their internal algorithms include marker tracking, motion analysis, background subtraction, human body segmentation, feature extraction, and classification.

WS methods use sensors attached to the body to collect motion data. The initial study of gesture recognition research using wearable sensors was a hand gesture recognition study based on the data glove of Foley [93] in the mid-to-late 1980s. The data glove developed by VPL Research converts the movement between the hand and the finger into an electrical signal, and it is composed of optical fibers as long as the length of the finger composed of LEDs and phototransistors, and light is output from the light emitting diode when the finger is moved. The hand motion was recognized by receiving it through the phototransistor. Based on this, a method for a worker to interact with a computer in a virtual environment was proposed. Later, in the early 1990s, along with the development of a sensor called "Wearcomp" that can recognize hand gestures based on HMD and camera by Mann [94], based on this, hand gesture recognition and control command conversion to the computer was carried out. In addition, "Glove-Talk II" was conducted, a hand gesture recognition study based on the data glove and neural networks interface of Fels and Hinton [95]. As in the previous cases, early gesture recognition studies based on wearable sensors were limited to the hand. However, as time passed, the gesture recognition range expanded to the whole body. In addition, it was confirmed that the data-glove type of sensor was mainly used, but it was expanded to various types, such as IMU, sEMG, and Myo-armband, over time. The data collected from these sensors is processed through various algorithms, such as Kalman filters, SVM, and ANN.

This paper presents a comprehensive survey of the literature on HMR based HRI published in the last ten years. Our survey focuses on the methods and techniques used in the last decade and aims to provide an overview of the state-of-the-art in this field. The paper is structured to provide a comprehensive overview of the methods used for motion recognition and their applications in HRI and highlight the challenges and opportunities for future research in this area.

## II. VISION SENSOR-BASED METHODS

The VS methods have two main approaches: marker-based and markerless. Marker-based methods use reflective markers or other visual cues to track human movements, while markerless methods rely on the human body's and the environment's intrinsic features to estimate movements. These two approaches to vision-based methods offer different advantages and limitations, each with unique applications and challenges. This review will explore the latest developments and trends in VS methods and WS methods, including a comprehensive examination of the different motion capture and analysis techniques. To ensure that this review covers a broad range of relevant studies and perspectives as well as the latest advancements and practical considerations in the field, the selection criteria for reference studies were set as follows:

1) Recent studies published within the last ten years ensure relevant and up-to-date information.
2) Studies validated or tested through practical or real-world applications have shown promising results.
3) Studies have considered the practicality and feasibility of implementing the proposed HMR system in real-world settings.
4) Studies have compared and evaluated approaches and techniques in HMR for HRI or HCI.

### A. MARKER-BASED METHODS

Table 1 presents the results of a survey of VS methods gesture recognition research using marker in robotics, computer science, human-computer interaction, and robot automation between 2010 and 2021. The table contains number of cameras, marker type, algorithm, gesture type, and performance perspectives. Only papers that met the research criteria were included, while those deemed inappropriate for the study for reasons such as missing critical information such as accuracy, algorithm, and research for simple movement measurement, not for recognizing and classifying movements, were excluded.

Chardonnens et al. conducted a study [11] to automatically measure the duration of the main phase of ski jumping using a reflective marker and an IMU sensor. As such, the marker-based method was developed to analyze the motion by measuring the continuous dynamic motion of the entire body of the subject and the purpose of generating 3D motion and has been mainly used in sports science, computer graphics, and computer visualization. The marker-based methods in this review install multiple optical motion cameras within a limited measurement range, attach markers to specific body segments, and recognize gestures of body segments by utilizing the trajectory tracking information of the markers. Whether an optical sensor is used, markers are divided into optical and non-optical markers. Among them, optical markers using an optical sensor are further classified into passive and active markers according to whether or not the marker emits light.

1) Active markers are electronic devices that emit a signal or light that a motion capture system can detect. These markers require a power source, such as a battery, and their tracking capabilities can be very accurate. However, they are also more expensive than passive markers and require more maintenance.
2) On the other hand, passive markers do not emit any signal or light; instead, they reflect light to the motion capture system. These markers are usually small, reflective spheres attached to the body, and they are less expensive and require less maintenance than active markers. However, they may be less accurate than active markers, especially with low lighting or interference from other reflective surfaces.

It was confirmed that marker-based methods had a higher proportion of dynamic gesture measurement research than static gestures due to the purpose and characteristics of sensor development. Of the 21 papers reviewed in this review, 14 (66.67%) ([12], [13], [14], [17], [20], [21], [22], [23], [25], [28], [29], [30], [31], [32]) were studies related to dynamic gestures. Eight studies (33.3%) ([15], [16], [18], [19]) were conducted to recognize static gestures based on precise position estimation technology for various still motions by attaching many markers to specific body parts ([26], [27], [29], [30]).

For performance evaluation, various performance indices such as accuracy, F1 score, and error rate were used, as the degree of improvement in accuracy compared to previous studies, comparison of accuracy using various algorithms, and normalization using independent data other than the data used to create the learning model. It is not easy to use as an objective evaluation index as it is evaluated by various criteria depending on the research purpose, such as performance and accuracy comparison in an indoor static research environment and an actual dynamic environment.

#### 1) PASSIVE MARKER-BASED STUDIES

Kulić et al. [12] used a reflective marker to recognize the subject's motion, analyzed and learned data on four types of motion in real-time, and created motion data for IRT humanoid control. Lee and Han [13] used eight motion cams and 14 reflective markers to develop unmanned monitoring technology to respond to safety problems at industrial sites and performed motion recognition of the entire body. Among the tasks, 22 were detected with an accuracy of 88%, and three were not detected due to environmental errors. Afterward, performance evaluation [14] was conducted to see whether the newly developed Kinect sensor could be used in gesture recognition research. The previously measured optical cam and marker-based subject motion data were used as ground-truth data for performance comparison. According to the results of these experiments, despite a position error of about 10.7 cm and a joint rotation angle error of 16.2° compared to the actual data, the recognition accuracy was 88%. Gardner et al. [15] created five static hand posture datasets for 12 subjects. They evaluated the performance by classifying them into three conditions (Raw, Aggregate, and Transformed): (1) Poor classification accuracy in the raw dataset, (2) Satisfying classification accuracy in the aggregate dataset, (3) Transformed dataset: The algorithm with the best average performance (MLP_BER: 0.183±0.168) and the algorithm with the best single performance (k-NN_BER: 0.158±0.152) were identified.

Li et al. [16] investigated sEMG sensor-based hand gesture recognition methodology for 18 subjects, classified 13 different hand gestures using six sEMG sensors, and classified dynamic features of fingers through 25 reflective markers. The SVM algorithm was used for data learning and classification, and an average classification accuracy of 98.45±0.83% was shown for 13 gestures. Chen et al. [17] use DGCNN (dynamic graph CNN) to perform spatiotemporal 3D event cloud recognition for gesture recognition. The performance of DGCNN was compared with the pointNet and pointNet++ algorithms used for recognizing the same data set (IMB DVS 128 gesture, DHP 19). Performance confirmed. Jiang et al. [18] proposed a new approach to recognizing hand gestures by estimating skin stimuli with multiple soft sensors, and the skin deformation pattern was first measured. The gesture recognition accuracy for static hand gestures and American sign language data (0 to 9) was verified with various algorithms (LDA, K-NN, RF) to verify the proposed method. Rahman et al. [19] presented an analysis and comparison of classifiers' efficiency when determining hand gestures using motion capture marker positions provided by Vicon cameras and a total of five algorithms (SGD, DT, LR, B, RF) and analyzed by comparing the results. As a result of the comparison, it was confirmed that the algorithm with the best performance was RF, and the algorithm with the lowest performance was DT.

### 2) ACTIVE MARKER-BASED STUDIES

Active marker comprises a light source rather than a reflector, and IRED (Infrared Ray Emitting Diode) is mainly used. Each diode has a different emission frequency. Since the camera is tuned to the frequency of the diode to be captured, it can identify a specific marker, so it has the advantage of reducing data post-processing time compared to passive markers.

Obdržálek et al. [20] proposed a correction method using a skeletal model based on the active marker's location data to overcome the skeletal model's decrease in accuracy due to the occlusion (occlusion) of the Kinect sensor. Although an error in the case of limited movement, it was confirmed that the joint estimation accuracy of the skeletal model was almost similar to the motion capture result. Chen, Mingyu, et al. [21] compares and analyzes the accuracy of the linear classifier and the HMM algorithm for two types of datasets (Implicit and Explicit) to identify an algorithm suitable for an efficient motion gesture recognition method in both user-dependent and independent data. As a result of the experiment, in the case of the Implicit 6D gesture, the HMM algorithm showed a classification accuracy of 91.9%, which was improved by 6.7%. In the case of the Explicit 6D gesture, the HMM algorithm showed a classification accuracy of 96.9%, which was improved by 3.4%, confirming that the HMM algorithm was efficient. It became. Pavllo et al. [22] applied machine learning (ML) technology to the motion capture system to solve the degradation of performance due to occlusion when performing inverse kinematics (IK) analysis based on marker data to analyze the IK of hands and fingers using only activated markers. A method was proposed, and the error performance was improved (3.82°) compared to the existing IK solver.

Andrychowicz et al. [23] proposed a hand manipulation method to use reinforcement learning to perform vision-based object orientation in a physical shadow dexterous hand. To measure the comparison data for evaluation, the movement data of the dexterous shadow hand-measured with an active LED marker was used, and learning was performed using the CNN algorithm, showing 90% accuracy. The study by Fern'ndez-Baena et al. [24] demonstrated the use of the Kinect sensor for marker-based motion capture and validated its accuracy for upper-body joint movements. However, the study also identified limitations with the accuracy of the Kinect sensor for lower body movements. The study uses a Kinect sensor to capture the motion data of participants performing upper-body and lower-body joint movements. The Kinect sensor has a depth camera that tracks the body's movement without needing markers, making it a low-cost and convenient tool for motion capture. The authors used the captured data to extract features such as joint angles, velocity, and acceleration, which were used for further analysis. The study's main goal was to validate the accuracy of the Kinect sensor's motion capture data. The authors used a gold standard motion capture system as a reference to compare the Kinect data. They found that the Kinect data was accurate for capturing the movements of the upper body but less accurate for capturing the movements of the lower body. Specifically, the authors found that the Kinect sensor tended to overestimate joint angles and that the accuracy of the lower body movements decreased with the increased complexity of the movement. The authors did not use any ML methods in their study. Instead, they used statistical analysis to compare the Kinect data to the gold standard motion capture data.

While hand gesture recognition methods that use color taping and gloves may not fall within the category of traditional optical markers, they are still classified as marker-based techniques in the context of HRI research. These methods enable the detection and tracking of a specific body part, facilitating the creation of movements in virtual reality (VR) environments, thus serving as a valuable tool for exploring the dynamics of HRI. Wang and Popović [25] proposed a real-time hand-tracking method using a color glove. The system captures color images of the user's hand wearing a color glove and processes the images to track the movement of the hand in 3D space. The features used in this method are based on the color histogram of the pixels in the hand region. The system uses a single camera to capture the images, making it a low-cost solution. The recognition accuracy of the system is not explicitly reported in the paper. Maino and Foresti [26] also proposed a real-time hand gesture recognition method using a color glove. The system captures color images of the user's hand wearing a color glove and uses a feature vector of 9 dimensions, including color and texture information. The system uses two ML algorithms, K-NN and LVQ (Learning Vector Quantization). The recognition accuracy of the LVQ is 97.79%, while the K-NN algorithm's accuracy is not explicitly reported. Ballarbi et al. [27] proposed a hand gesture interaction method using color taping on the user's fingers. The system captures color images of the user's hand and uses a color detection algorithm to identify the taped fingers. The system then performs static hand posture recognition and dynamic hand gesture recognition. The features used in this method are based on the position and orientation of the fingers. The system uses a rule-based algorithm to classify hand gestures. The recognition accuracy of the system is not explicitly reported.

In summary, all three studies used color-based methods for hand gesture recognition, and they differ in the features used, the ML algorithms applied, and the recognition accuracies achieved. The study [26] achieved the highest recognition accuracy using LVQ, while the study [25] proposed a low-cost solution using a single camera. Study [27] focused on hand gesture interaction for tabletop interfaces, demonstrating the potential of color taping on fingers as an alternative to gloves.

ECCDOT C-DOT et al. [28] studied recognizing hand gestures using a 2D single cam while holding a 2D marker in the subject's hand. After attaching markers to the hands of five subjects, 'moving and clicking the virtual mouse' was performed, and the error rates were low at 0.1% and 1%, respectively, for two errors (mistaking marker movement for the click, click gesture not recognized). Amendola et al. [29] proposed recognizing the subject's movement by attaching an RFID tag to the body as an efficient patient gesture recognition method that can be applied in the medical and rehabilitation fields. Three subjects attached an RFID tag and proceeded with five periodic movements and single gestures. Motion recognition was performed through the SVM algorithm. For arm movements, it improved from 86% to 98%, and for leg movements, it improved from 60% to 78%. Ishiyama et al. [30] proposed a subject's hand motion recognition method using a single camera and gloves with AR marker patterns attached to solve the poor performance of human hand gesture recognition under non-uniform lighting conditions. A total of 122 recognition rate of about 78.68% (96 out of 122) was shown using the dog's posture and 20 image data sets for each posture. Roy, Prasun, et al. [31] proposed an air-writing framework based on a CNN algorithm using a general video camera and recognizing a color tip attached to the end for gesture recognition. For 20 subjects, each person wrote the number 50 times and showed an accuracy of 97.7% in English, 95.4% in Bengali, and 93.7% in Devanagari. Nakamura and Hoshino [32] proposed an algorithm for analyzing the subject's whole-body movement using color markers to make separate clothes and individual training possible during Tankendo training. In this study, the subject's joint angle was estimated using the HSV color detection algorithm, and its performance was demonstrated by showing standard errors of 7.7° at 90° and 4.5° at 120°.

### 3) ALGORITHMS FOR MARKER TRACKING AND MOTION ANALYSIS

In the marker-based method, the internal algorithms used for motion or gesture recognition typically involve two steps: marker tracking and motion analysis.

Marker tracking detects and tracks the reflective markers attached to the human body. The goal of marker tracking is to accurately estimate the 3D position of each marker in real-time, which provides a direct measurement of the human body's motion. It tracks the real-time position and orientation of the markers attached to the human body segments. Several popular marker tracking methods exist, including optical flow, corner detection, feature-based tracking, Kalman filter, particle filter, and optical marker tracking (OMT). The corner detection method detects the corners of the markers and tracks them based on their positions. Feature-based tracking method uses features such as SIFT (Scale-Invariant Feature Transform), SURF (Speeded Up Robust Feature), and ORB (Oriented FAST and Rotated BRIEF) to detect and track markers. The Kalman filter method uses a mathematical model to predict the movement of markers based on their past positions. The particle filter method uses a probabilistic model to track markers by representing the markers as a set of particles and updating the particles' positions based on the observations. OMT method uses a combination of optical flow and corner detection to track markers.

Motion analysis, which directly contributes the motion recognition, is the process of using the marker positions to estimate the complete human body pose. The motion analysis can be done using various algorithms, including forward kinematics (FK), IK, and ML-based methods.

1) Human skeletal kinematics such as FK and IK use a pre-defined human body model to calculate the body pose based on the marker positions and the marker positions to solve for the body joint angles, respectively.
2) ML-based methods, such as DTs, RFs, and support vector machines, can map the marker positions to the body pose, leveraging extensive training data. Once the human body pose is estimated, various features can be extracted for gesture recognition. These features may include joint angles, joint velocities, and limb lengths, and they can be used as input to ML-based classifiers to perform gesture recognition. Moreover, the algorithms used for gesture recognition in marker-based methods can be categorized into two main groups: non-neural network (non-NN) and neural network (NN) algorithms.
3) Non-neural network algorithms for gesture recognition in marker-based motion capture include K-NN, Hidden Markov models (HMM), SVM, Dynamic Time Wrapping (DTW), RF, Latent Dirichlet Allocation (LDA), Template Matching Algorithm (TMA), HSV algorithm, and LVQ. The K-NN algorithm is a simple algorithm that classifies an input gesture based on the k-closest training samples. The HSV algorithm uses a gesture's hue, saturation, and value to classify it into a pre-defined category. The HMM algorithm is a statistical model that can model a sequence of observations by a Markov process. The SVM algorithm is a supervised learning algorithm that can be used for binary classification and multi-class classification problems. DTW is an algorithm that compares two sequences by warping the time axis to minimize the Euclidean distance between them. RF is an algorithm that generates a set of DTs from the training data and outputs the class label based on the majority vote of the DTs. LDA is a linear discriminant analysis(LDA) algorithm that uses Bayes' theorem to predict the class labels of the gestures. TMA is a gesture recognition algorithm that uses a pre-defined template to recognize gestures.
4) In the case of neural network-based algorithms, Multi-Layer Perceptron (MLP), CNN, and Dynamic Graph CNN (DGCNN) have been used. MLP is a feedforward neural network consisting of multiple hidden neurons and an output layer that outputs the class labels. CNN uses convolutional layers to extract features from the input data and uses fully connected layers to classify the data into different categories. DGCNN is a dynamic graph CNN algorithm that can handle non-Euclidean structured data and process data with varying lengths and shapes.

#### 4) HARDWARE

In the case of the marker-based method, there are differences for each camera, but generally, gestures can be measured in 30 to 2000 frames, so the results most similar to actual movements are output. A. Fern'ndez-Baena et al. [20] compared the angle errors of the shoulder, hip, and knee joints measured using the marker and the Kinect sensor to evaluate the effectiveness of using the Kinect sensor for motion recognition-based rehabilitation training. As a result of the study, it was judged that the motion data measured by Kinect had a low error (0.097%) of 6.78° to 8.98° compared to the actual motion, so it was judged to be valid for use in rehabilitation exercise research. Motion data was used. As such, the marker-based motion recognition technique, as in the previous cases ([14], [18], [20], [23], [24]), enables high-quality gesture recognition most similar to actual motion; it has the advantage that it can be used as actual motion data for evaluating the motion of data. On the other hand, as in the case of previous studies ([13], [14], [20], [22]), the marker-based method does not recognize the marker to be recognized due to obstacles or body obstruction during the subject's motion. Due to chronic limitations such as marker occlusion leading to errors, there is a disadvantage in that high accuracy is only shown in an artificially limited environment. In addition, wearing the equipment is cumbersome; it is difficult to attach it by him/herself, and one cannot escape the workspace within the FOV of the motion camera. Due to noise effects, applying it in a dynamic environment and non-uniform lighting conditions is challenging. It is challenging to apply it in actual industrial sites due to various constraints, such as poor reproducibility of research because sensors cannot consistently be attached to the exact location. However, it is valid for verifying data measured by other sensors. It is most suitable for validating these sensors and verifying final data.

### B. MARKERLESS METHODS

Table 2 presents the results of a survey of VS methods markerless gesture recognition research in robotics, computer science, human-computer interaction, and robot automation between 2010 and 2021. The table contains camera type, algorithm, gesture type, and performance perspectives. Only papers that met the research criteria were included, while those deemed inappropriate for the study for reasons such as missing critical information such as accuracy, algorithm, and sensor, or those that used point recognition of hand or arm center for motion tracking were excluded.

Marker-based and markerless motion recognition methods have application domains in various fields, including sports, healthcare, robotics, and entertainment. However, the specific application domains of each method may differ due to their strengths and weaknesses.

Marker-based motion recognition methods are commonly used in biomechanics research, sports performance analysis, and clinical gait analysis. They provide highly accurate and precise measurements of joint angles, joint velocities, and forces, which are essential for understanding movement patterns and identifying areas of improvement in athletic performance or rehabilitation. However, marker-based methods require physical markers to be placed on the body, which can be uncomfortable and may interfere with natural movement.

Markerless motion recognition methods, on the other hand, have been applied in areas such as computer vision, virtual reality, and gaming. They do not require physical markers, making them more convenient and less invasive than marker-based methods. Markerless methods are also better suited for tracking full-body motion, such as in dance and choreography, where multiple markers would be needed to capture the complexity of the movements. However, markerless methods may be less accurate than marker-based methods in tracking individual joint angles and forces, which can limit their use in biomechanical research.

#### 1) RGB CAMERA

The RGB camera plays an essential role in gesture recognition by capturing the color information of the object being recognized. It provides additional visual features that can improve the recognition accuracy of the system. The data obtained from the RGB camera can be used to extract various features for gesture recognition, including:

1) Hand and arm shape: The RGB camera can capture the shape of the hand and arm, which can be used to recognize specific gestures.
2) Skin color: The RGB camera can capture the skin color of the hand, which can be used to distinguish it from the background and improve recognition accuracy.
3) Background color: The RGB camera can capture the background color, detecting when the user's hand is in front of a specific color and triggering a specific gesture.

Sigalas et al. [35] proposed a new approach to vision-based hand-gesture recognition and analyzed the motion by matching it with the upper arm model after analyzing the part corresponding to the joint parameters of the body in the image data. In that study, a hand-gesture recognition accuracy of about 86% was achieved based on the MLP+RBF algorithm. Murthy and Jadon [36] proposed an NN-based gesture recognition study and constructed a dataset with images captured in video images. About 89% hand gesture recognition accuracy was achieved by capturing the background without the user's hands and removing the background from the image data set. Raheja et al. [37] proposed a new approach to control the robot through simple hand motions in front of the camera. After extracting the hand-gesture area from the image in the video image, PCA-based pattern-matching was used to achieve about 90% accuracy. In addition, Cao et al. [38] proposed an MLP-based hand-gesture feature extraction method. To classify the edge of the hand image, we used the Laplacian of Gaussian (LoG) edge detection method and achieved an accuracy of about 97.4%. Nagi et al. [39] proposed an HRI interface with a mobile robot based on real-time hand gestures

**TABLE 1. Vision sensor-based gesture recognition with the marker.**

| Index | Year | Num of camera | Marker | | | Algorithms | | Dataset | Gesture | Target-parts | Performance (A: accuracy, E: error) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Optical | | Non Optical | NN | Non-NN | | Type | | |
| | | | Active | Passive | | | | | | | |
| [25] | 2009 | 1 | | | color glove | | NN | | dynamic | hand | E: (Avg: 5~10cm, Max: 15cm) |
| [26] | 2011 | 1 | | | color glove | | K-NN, LVQ1,2,3 | | static | hand | A(K-NN:85.67%, LVQ1: 96.36%, LVQ1+LVQ2: 97.57%, LVQ1+LVQ3: 97.79%) |
| [27] | 2011 | 1 | | | color taping | | HSV | | static & dynamic | fingertip | 100% |
| [12] | 2011 | | | reflective | | | HMM | | dynamic | body | E: 3.5% |
| [20] | 2012 | 8 | LED | | | | Motion builder | | dynamic | body | E: in 10cm |
| [21] | 2012 | 4 | LED | | | | HMM | | dynamic | hand | A: 96% |
| [13] | 2013 | 8 | | reflective | | | DTW | | dynamic | body | A: 88% |
| [28] | 2013 | 1 | | | AR marker | | HMM | | static | hand | A: 99% |
| [14] | 2013 | 6 | | reflective | | | DTW | | dynamic | body | A: 88% |
| [15] | 2014 | 1 | | reflective | | MLP | SVM, K-NN, RF | | dynamic | back of the hand | E: MLP(0.183±0.168) K-NN(0.158±0.152) |
| [29] | 2015 | | | | RFID | | SVM | | dynamic | body | A: (Avg: 82.5%, Best: 90.7%) |
| [30] | 2016 | 1 | | | AR marker | | TMA | | static | hand | A: 99.5% |
| [31] | 2018 | 1 | | | color tip | CNN | | | dynamic | | A: (Avg: 97.7%) |
| [22] | 2018 | 1 | LED | | | NN | | | dynamic | back of hand | E: 3.82° |
| [32] | 2018 | 1 | | | color | | HSV model | | dynamic | body | A: 95.5% |
| [16] | 2019 | 6 | | reflective | | | SVM | | dynamic | back of hand | A: 98.45±0.83% |
| [33] | 2019 | | | | color | CNN | | | dynamic | fingertip | A: 97.29% |
| [17] | 2020 | 4 | | reflective | | DGCNN | | | dynamic | hand, arm | A: 94.94% |
| [23] | 2020 | 16 | LED | | | CNN | | IBM DVS, DHP19 | dynamic | hand | A: 90% |
| [18] | 2020 | 1 | | reflective | | | LDA, K-NN, RF | ASL dataset | static | back of hand | A: 95.8% |
| [19] | 2021 | 4 | | reflective | | | RF | Andrew gardner [2014] | static | hand | A: 88.9% |

(1) ASL dataset: https://public.roboflow.com/object-detection/american-sign-language-letters; (2) DHP 19 dataset: https://sites.google.com/view/dhp19; (3) IBM DVS 128 gesture dataset: https://research.ibm.com/interactive/dvsgesture/; (4) Andrew gardner: https://www.slideshare.net/andrewgardner5811;

and used MPCNN for hand-gesture classification. Six gestures were recognized with 96% accuracy, and real-time performance of 0.82s/recognition was confirmed in the environment of 128MB RAM and ARM 11 533MHz processor. Celik and Kuntalp [40] proposed a manipulation control method using image processing in human-machine interaction (HMI). After converting the RGB values into HIS values, the hand region in the image was separated and binarized with a low-pass filter.

Moreover, based on TMA and SSA (sign signal algorithm), the separated hand's gesture mode was recognized with error rates of TMA (6.66%) and SSA (9.4%). Cho et al. [41] proposed an FPGA-based gesture recognition system and analyzed the accuracy according to the background conditions of the space where the subject performed the gesture. Features were extracted based on the optical flow algorithm and as a result of recognizing 25 gestures in different experimental environments, simple indoor (100%), complex indoor (100%), complex static outdoor (99.33%), complex dynamic outdoor (90.67%) %) accuracy. Jeong et al. [42] captured the motion path when drawing a sign in the air, recognized the user's hand gesture, and proposed an interaction method with the TV based on this. An image captured by a standard VGA (video graphics array) was used, and as a result of gesture recognition with a particle filter algorithm, an average recognition rate of 92.34% was achieved.

Barros et al. [43] learned five gesture commands based on the MCCNN algorithm and, as a result of testing based on the Logitech C905 USB camera, achieved an F1 score of 96.85%. Li et al. [44] learned 100 RGB image data for each gesture with ANFIS and SVM algorithms to control a virtual robot with ten hand gestures remotely and achieved recognition accuracy of 96.3% and 98.83%, respectively. Bhame et al. [45] proposed a gesture recognition method that can be used in HCI application development. RGB image data was obtained for 30 subjects, and

94.9% accuracy was achieved by training with the DHMM algorithm. Simul et al. [46] proposed a real-time facial expression and gender classification method based on RGB images and SVM and achieved a classification accuracy of face (86%), facial expression (95.33%), and gender detection (94.67%). Lamb and Madhe [47] proposed an automatic bed positioning system through the patient's feature gesture input and achieved 96% accuracy by combining a new algorithm called Symlet Wavelet with the existing Euclidean distance algorithm.

Mohanty et al. [48] proposed a new hand-gesture recognition method that integrates multi-image features and multi-kernel learning SVM to improve the accuracy of multi-class hand-gesture recognition and generalize the algorithm. Ghosh and Ari [49] used two ASL (American sign language) datasets for static hand-gesture recognition, and through cross-validation of various recognition algorithms, the maximum dataset-1 (82.25%) and dataset-2 (87.67%) per dataset) confirmed the recognition accuracy. Oyedotun and Khashman [50] discusses the use of deep learning in recognizing static hand gestures in vision-based systems. The authors propose a deep learning approach for recognizing hand gestures using a CNN and evaluate its performance compared to traditional ML algorithms. As a result of a comparative analysis of learning performance using the CNN algorithms and SDAE algorithm for the ASL dataset, a gesture recognition rate of 91.33% and 92.83%, respectively, was confirmed. The research by Wang et al. [51] discusses using deep learning in HMR for predictive context-aware HRC. The authors propose a deep learning-based method for recognizing human motions using CNNs and evaluate its performance in recognizing human motions in real-time. The study highlights the potential of deep learning in HMR and its applications in predictive context-aware HRC systems. RGB images with complex backgrounds were used, and as a result of learning using DCNN and Alexnet, a gesture recognition accuracy of about 96.6% was achieved. Singha et al. [52] developed a dynamic hand-gesture method combining a 3-frame difference technique and skin filtering to solve the causes of chronic performance degradation in VS methods gesture recognition research, such as complex background, lighting, and occlusion. A recognition method was proposed. After learning a database of 40 gesture classes (10 numbers, 26 alphabets, four arithmetic operators) for 20 subjects with four algorithms (ANN, SVM, K-NN, classifier fusion), the performance was compared. It was confirmed that the proposed classifier fusion (92.23%) was about 3.92% higher than the lowest-performing SVM (88.31%).

Song et al. [53] implemented mode control for mouse position control (up, down, left, and right), left click, and no action based on dynamic hand-gesture recognition using a USB single camera. We collected 30 gesture datasets for six gesture modes per subject in white background and cluttered background, respectively, and as a result, the recognition accuracy of 95.95% and 84.07% was achieved in white background and cluttered background, respectively. Sun et al. [54] proposed a segmentation method for human hands to implement real-time hand-gesture tracking in images with complex backgrounds. The authors start by creating a model of skin color to identify the regions of the video that contain a hand. They then used an AdaBoost classifier based on Haar wavelets to detect the specific hand gestures within these regions. The authors chose to use skin color as a critical feature for detecting hands in the video because human skin has a unique color that a computer can identify. To do this, they first collected a set of images containing hands and used them to create a model of the average skin color of a human hand. They then used this model to identify the regions of the video that are likely to contain a hand. Once the regions containing hands were identified, the authors used an AdaBoost classifier based on Haar wavelets to detect the specific hand gestures within these regions. The Haar wavelets were used to extract features from the video frames, then to train the classifier to recognize different hand gestures. The authors took one video frame at a time to analyze the hand gestures and cut it into smaller sections for analysis. This denaturation technique allowed them to identify each gesture's specific features and improve the classifier's accuracy. As a result, hand motion recognition was performed based on the CNN algorithm, and an average recognition rate of 98.3% was achieved.

Islam et al. [55] presented robust methodologies for an underwater robot to visually detect, follow, and interact with a diver for collaborative task execution. It introduces two autonomous diver-following algorithms based on spatial and frequency-domain features and a CNN-based model for tracking-by-detection. The paper also proposed a hand gesture-based human-robot communication framework that is more computationally efficient than existing grammar-based frameworks. The proposed framework used deep visual detectors for accurate hand gesture recognition and a finite-state machine for gesture-to-instruction mapping. The paper validates the effectiveness of the proposed methodologies through field experiments in closed and open-water environments and demonstrates the usability benefits of the proposed interaction framework compared to existing methods through a user interaction study. As a result, it was trained using a dataset of about 10,000 underwater hand gestures RGB images and achieved about 80% recognition accuracy. Chang et al. [56] proposed a method to improve gesture recognition accuracy with a Faster R-CNN algorithm using VGG16 and a Gaussian filter to remove noise from image data. As a result of using five-fold cross-validation, the experimental results show that an improved Faster R-CNN algorithm significantly improves mean average precision to 99.89%, which provides a better method for gesture recognition in HRI applications. Fang et al. [57] proposed a new gesture recognition algorithm based on CNN and DCGAN (deep convolution generative adversarial networks) to break the bottleneck. Based on the CNN+DCGAN algorithm, we achieved a hand-gesture recognition accuracy of 92.70% and experimentally proved that the proposed method is less sensitive to lighting and background interference. Sahoo et al. [58] proposed a user-independent hand gesture recognition system to solve problems caused by lighting changes, diversity of user hand shapes, and high similarity between classes in the automatic recognition of visual-based static hand-gesture images. Learning was conducted based on the MU data set and three algorithms, and the PCA-based deep CNN algorithm (87.83±1.79%) achieved higher accuracy than the control group CNN (73.86±1.04%) and FFCN (84.02±0.59%).

Jiang et al. [59] proposed a gesture detection network and AUV control algorithm for interaction with an autonomous underwater vehicle. We conducted data learning based on 8,000 training image sets and CNN, F-RCNN, and YOLOv4 algorithms and verified the superiority of the proposed F-RCNN by comparing and verifying the accuracy of 83%, 89%, and 79%, respectively. Al-Hammadi et al. [60] proposed a dynamic hand-gesture recognition system using multiple deep-learning architectures for hand segmentation, local and global feature representations, sequence feature globalization, and recognition. The proposed architecture is evaluated on a challenging dataset of 40 dynamic hand gestures performed by 40 subjects in an uncontrolled environment. The learning accuracy was compared according to whether or not the generalization work was performed on the data set, and the 3DCNN+MLP algorithm showed an initial accuracy of 98.62%, and after the generalization work, the quantitative precision deterioration was evaluated at 87.69%. In the case of the 3DCNN+AE (autoencoder) algorithm, an accuracy of 84.89% was confirmed after the generalization process compared to an initial accuracy of 98.75%. Al-Hammadi et al. [61] proposed an efficient deep CNN (3DCNN) for hand-gesture recognition, showing 100% accuracy for dependent data from training data but 84.38% for independent data. Choudhary and Tazi [62] proposed a method for segmenting and recognizing hand-gesture usable in HCI using a real-time image sequence captured by a video recording device to track the potential subject region (PSR) by itself. As a result of learning 9,000 images for eight gestures based on the VGG16-CNN algorithm, it was confirmed that eight gestures could be recognized with an accuracy of 96% or more.

### 2) RGB-DEPTH CAMERA

The following are examples of research on gesture recognition using an RGB-Depth camera. Kinect is a stereo camera sensor developed by Microsoft in 2010 for motion recognition of game players and has been used in earnest for gesture recognition research since 2012. Although Kinect is a product name, it is marked separately as RGB-D (K) in Table 2 because it shows much research data and utilization as a classification standard. Cases of gesture recognition based on RGB-Depth rather than Kinect were listed first, then cases of gesture recognition based on Kinect sensors were reviewed.

RGB-D sensors typically use infrared or structured light patterns to capture depth information. While they may not have the same wide field of view or multiple sensors as the Kinect, they still offer several advantages for gesture recognition, such as:

1) Higher resolution color information: RGB-D sensors typically have higher resolution color cameras than the Kinect sensor, which can help recognize gestures that involve color cues, such as hand signs.
2) Smaller form factor: RGB-D sensors are typically smaller and more compact than the Kinect sensor, making integrating them into different devices and environments easier.
3) Lower cost: RGB-D sensors are generally less expensive than the Kinect sensor, making them a more affordable option for gesture recognition applications.

Some of the features that can be extracted through the data obtained from an RGB-D sensor include:

1) Depth maps: RGB-D sensors capture depth maps, which provide information about the distance of objects from the sensor. These depth maps can be used to recognize gestures that involve movement towards or away from the sensor.
2) Point clouds: RGB-D sensors also capture point clouds, which provide a 3D representation of the environment. These point clouds can be used to recognize gestures that involve interaction with the environment, such as pointing or grasping.
3) Skeleton tracking: Like the Kinect sensor, some RGB-D sensors also offer skeleton tracking, which provides information about the position and movement of joints in the human body. This can be used to recognize various body movements and gestures.

Ohn-Bar and Trivedi [63] proposed an RGB and Depth-based in-vehicle gesture interface from rough hand to fine finger motion. The robustness improvement for noise factors was quantitatively evaluated by conducting research in frequent self-occlusion. Nineteen gestures performed by eight subjects were learned based on SVM, and an accuracy of about 96.97% was achieved. Coupeté et al. [64] proposed implementing hand trajectory-tracking and gesture recognition of workers based on a depth camera with a top view to implementing HRC with collaborative robots on factory assembly lines. As a result of learning the gesture data set for 20 assembling cycles of 13 workers with HMM, a training accuracy of about 85% and a test accuracy of 80% were achieved. Yu et al. [65] proposed a human-UAV interaction (HUI) method based on hand & arm gesture recognition for natural interaction between humans and multi-UAV systems. In that study, nine types of static arm gesture images of operators were acquired with Asus' Xtion Pro Live camera, similar to Kinect, and rule-based classification based on the Nite library was applied for mapping with motion control commands of AR drones.

### 3) KINECT SENSOR

The following studies focus on gesture recognition using the Kinect sensor and its applications in human-computer interaction and robotics. Kinect and RGB-D sensors use depth information to enable gesture recognition, but Kinect sensors have advantages over traditional RGB-D sensors. Here are a few advantages of the Kinect sensor for gesture recognition:

1) Wide field of view: Kinect sensors have a broader view field than traditional RGB-D sensors. This allows the Kinect sensor to capture more of the surrounding environment and provides a larger area for gesture recognition.
2) Active illumination: Kinect sensors use active illumination to capture depth information. This means that the sensor emits its infrared light, which is then reflected by objects in the environment. This active illumination allows the Kinect sensor to work in low-light environments and provides more accurate depth information.
3) Multiple sensors: The latest Kinect sensor (Kinect v2) has multiple sensors allowing more accurate and robust gesture recognition. For example, the sensor can use body tracking to track multiple people in the same environment and distinguish between them.

With the data obtained from the Kinect sensor, several features can be extracted for gesture recognition, such as:

1) Joint positions: The Kinect sensor captures the 3D position of various joints in the human body, such as the head, hands, elbows, knees, and feet. These joint positions can be used to infer the posture and movement of the body and recognize gestures.
2) Body orientation: The Kinect sensor also provides information about the body's orientation, which can help recognize more complex gestures involving rotation or twisting.
3) Hand shape and movements: The Kinect sensor can also capture the shape and movement of the hands, which helps recognize hand gestures and gestures that involve manipulating objects.
4) Facial expressions: The Kinect sensor can also capture facial expressions, which can be used to recognize emotions or commands that involve facial expressions.

Overall, the Kinect sensor provides a rich set of features that can be used for gesture recognition, making it a popular choice for researchers and developers.

Ren et al. [66] developed a hand gesture recognition system based on the Kinect sensor. They used hand position, shape, and motion features to train an ML classifier (RF) to recognize 12 hand gestures. The system achieved an average recognition accuracy of 93.9%. The study's contribution is providing a comprehensive exploration of the Kinect sensor's potential for hand gesture recognition and its applications in human-computer interaction. Gu et al. [67] developed a gesture recognition system that recognizes ten hand gestures using an RF classifier. They utilized features including hand shape, hand location, and hand motion to recognize gestures. The study's contribution shows that the Kinect sensor is adequate for recognizing hand gestures and has potential applications in HRI. Xu et al. [68] developed a real-time dynamic gesture recognition system that recognizes gestures for robot navigation. They used an SVM classifier to recognize eight hand gestures based on hand shape, motion, and orientation. The study's contribution is showing the effectiveness of the Kinect sensor in real-time gesture recognition for robotics applications. Qian et al. [69] developed a gesture-based remote HRI system using the Kinect sensor. They used a DT classifier to recognize eight different gestures. The system was able to control the movement and navigation of the robot. The study's contribution shows the Kinect sensor's potential in developing a gesture-based remote-control system for HRI.

Yeo et al. [70] proposed a hand-tracking and hand-gesture recognition method from dynamic motion in a complex background and performed a quantitative performance comparative evaluation of a low-cost webcam and Kinect sensor. Pisharady and Saerbeck [71] proposed a body gesture-based detection and recognition algorithm for human interaction and a household floor cleaner robot. Pisharady conducted model learning based on the Kinect skeletal model and DTW and achieved the highest accuracy of 97.26% when joint angle, position, and direction were used together with skeletal reconstruction. Chen et al. [72] proposed a real-time dynamic hand gesture recognition system using the Kinect sensor. The system extracts depth and color features from the hand region and uses a K-NN classifier for recognition. The system achieved an average recognition rate of 92.8% on a dataset of 5 hand gestures, and the recognition time is less than 50ms. The study's contribution is a practical and efficient gesture recognition system for human-computer interaction. Coupeté et al. [73] developed a gesture recognition system using a depth camera for HRC on the assembly line. The system uses Haar-like features and an SVM classifier to recognize nine hand gestures. The recognition rate achieved is 94.4% for a single person and 91.6% for two people. The study's contribution is the application of gesture recognition in the context of HRC, which could improve productivity and safety in the manufacturing industry. Vinh and Tri [74] proposed a hand gesture recognition system based on depth images using the Kinect sensor. The system uses a combination of depth and color features and HMM classifier to recognize six hand gestures. The system achieved an average recognition rate of 90.2% on a dataset of 10 participants. The study's contribution is using HMM for gesture recognition, which could improve recognition accuracy in dynamic gesture recognition scenarios. Alasady et al. [75] presented an online dynamic gesture recognition system for HRI. The system extracts depth and motion features and uses an incremental online learning algorithm for recognition. The system achieved a recognition rate of 92.3% for a dataset of 10 dynamic hand gestures. The study's contribution is using online learning for gesture recognition, which could improve recognition accuracy in dynamic scenarios.

Li et al. [76] proposed a dynamic gesture recognition system for the Internet of Things (IoT). The system uses the Kinect sensor to extract depth and motion features and a CNN classifier for recognition. The system achieved an accuracy of 96.5% for a dataset of 10 hand gestures. The study's contribution is using CNN for gesture recognition, which could improve recognition accuracy and adaptability to various gesture scenarios. Additionally, the study's application in IoT could enable more intuitive and efficient human-computer interaction. Jiang et al. [77] showed that the skeletonization algorithm and CNN for the recognition algorithm could reduce the impact of shooting angle and environment on the recognition effect and improve gesture recognition accuracy in complex environments. In that study, the experimental results show that compared with the SVM method, dictionary learning + sparse representation, CNN method, and other methods, the recognition rate reaches 96.01%. Dong et al. [78] converted gesture recognition into a shortest path problem by converting a gesture feature matrix into an undirected graph and proposed a new dynamic gesture recognition algorithm of DPCNN for real-time HRI. The MSRC-12 dynamic database was trained with the DPCNN algorithm, showing an average recognition rate of 93.68%, and the effectiveness was proven through performance comparison with other algorithms (DL-COPAR: 93.80%, LC-KSVD: 91.61%).

Cheng et al. [79] proposed CNN and RBM (Restricted Boltzmann Machine) joint networks to solve issues such as accuracy, real-time, and low robustness of gesture recognition. CNNs are well-suited for processing visual data, such as images or video frames, and are widely used in computer vision tasks. However, they can be limited by the amount of training data available and the complexity of the task. Restricted Boltzmann Machines (RBMs) is a generative model that can learn to represent high-dimensional data more compactly and informally. RBMs can pre-process data and reduce its dimensionality, improving subsequent processing steps' performance. By combining CNN and RBM, the strengths of both models can be leveraged to improve gesture recognition accuracy. The CNN can extract features from the raw image data, and the RBM can further process and compress these features to create a more robust and informative representation. In addition, using RBM can improve the system's robustness by allowing it to recognize similar gestures even if performed slightly differently. This can help overcome issues with variability in how people perform gestures. Finally, using RBM can also improve the real-time feasibility of the system by reducing the computational complexity of the processing steps. This is important for applications where low latency is critical, such as in real-time gesture recognition for human-computer interaction. Through simulation analysis, it is found that the joint network has a high recognition rate in simple sample gesture recognition, and its error is only 3.9%. Then on the complex sample, the joint network and other centralized networks do not perform well, mainly because RBM requires strict data distribution. Sun [80] proposed a method to enable smooth interaction with sound and lighting systems using an interactive panel controlled by hand in AR. For four gesture classes of 10 subjects, gesture learning was conducted with SVM and 2D CNN algorithm, and the 2D CNN algorithm confirmed 100% recognition performance, which was 4% higher.

Kuang et al. [81] proposed a 'One-shot' gesture recognition method that can be a highly efficient communication channel in the HRI system. The proposed approach can provide efficient, one-shot gesture recognition without elaborately designed features. The experiments on a social robot (JiaJia) demonstrate that the proposed approach can flexibly be used in a human-robot collaboration system because learning the data of 5 gestures of 5 subjects with DTW achieved an accuracy of about 92.4%. Cardenas and Chavez [82] tested various integration methods to fuse spatiotemporal features to improve recognition performance in a linear SVM classifier. The UTD-MHAD data set, which includes data from 27 tasks performed by eight subjects, was used for learning. Compared to the recognition accuracy of previous studies, it was confirmed that the algorithm combining HCM, CNN, and spherical coordinates algorithm showed the best accuracy (94.81%).

#### 4) TOF, DVS, DSV, AND PMD

In addition, studies on gesture recognition based on one type of sensor, such as a ToF camera, DVS (Dynamic Vision Sensor), DSV (Dynamic Stereo Vision), and PMD (Photon Mixing device), have been conducted. The subsequent studies used various types of sensors for gesture recognition, including Dynamic vision sensor cameras [83], ToF cameras [84], [86], [87], [89], Bio-inspired 3D vision sensors [85], and RGB-D cameras [88]. These sensors capture different types of information, such as temporal changes of an object's intensity, depth, and intensity information, and color and depth information of the hand gesture, which are then processed using various ML methods for gesture recognition. Here are some key features of the sensors:

1) DVS camera: A DVS camera is an event-based camera that captures motion in a frame-free manner, which means it only records changes in the scene rather than capturing a full frame at a specific frame rate. DVS cameras can be useful for fast and low-latency gesture recognition applications, but they typically have lower spatial resolution than Kinect sensors and other cameras.
2) ToF cameras: Like the Kinect sensor, ToF cameras capture depth information by measuring the time it takes for light to bounce back from objects in the scene. However, ToF cameras typically have a shorter range and lower depth resolution than the Kinect sensor. They can be useful for shorter-range gesture recognition applications but may not be ideal for larger environments.
3) Bio-inspired 3D vision sensor: Bio-inspired 3D vision sensors are cameras designed to mimic biological

visual systems' structure and function. They typically use spiking neural networks to process visual information and can be more power-efficient than traditional cameras. However, these cameras may have lower spatial and depth resolution than the Kinect sensor.

Ahn et al. [83] proposed a dynamic vision sensor camera-based bare-hand gesture recognition system. The system features a dynamic vision sensor camera that captures the temporal changes of an object's intensity, which is then preprocessed using a temporal filter. The filtered data is then processed using an SVM classifier to recognize the gestures. The system achieved recognition accuracy of 91.2% and showed promising results for recognizing dynamic gestures. Oprisescu et al. [84] presented an automatic static hand gesture recognition system using time-of-flight (ToF) cameras. The system features a ToF camera that captures the depth and intensity information of the hand gesture. The captured data is then preprocessed using a morphological filter and processed using a neural network classifier. The system achieved recognition accuracy of 95.8% for a 10-class gesture recognition task, and the results indicated the potential of using ToF cameras for static gesture recognition. Kohn et al. [85] proposed a real-time gesture recognition system using a bio-inspired 3D vision sensor. The system features a bio-inspired vision sensor that mimics the behavior of the human retina, which captures both the intensity and temporal information of the gesture. The captured data is then processed using a spiking neural network classifier. The system achieved recognition accuracy of 95.5% for a 10-class gesture recognition task, and the results demonstrated the effectiveness of using bio-inspired vision sensors for gesture recognition. Kulkarni et al. [86] presented a static gesture recognition system using a PMD ToF camera. The system features a PMD ToF camera that captures the depth information of the hand gesture. The captured data is then preprocessed using a filtering and segmentation technique and processed using an SVM classifier. The system achieved recognition accuracy of 97.5% for a 4-class gesture recognition task, and the results showed the potential of using PMD ToF cameras for static gesture recognition. Kopinski et al. [87] proposed a multi-sensor fusion strategy for hand gesture recognition using time-of-flight-based sensors.

The system features multiple ToF cameras that capture the hand gesture from different viewpoints, and the captured data is fused using a decision-level fusion technique. The fused data is then processed using an SVM classifier. The system achieved recognition accuracy of 98.5% for a 6-class gesture recognition task, and the results demonstrated the effectiveness of using multi-sensor fusion for hand gesture recognition. Sachara et al. [88] presented a free-hand gesture recognition system using 3D CNNs for in-car infotainment control. The system features a 3D CNN that extracts spatial and temporal features of the hand gesture from RGB-D data. The extracted features are then processed using an SVM classifier. The system achieved recognition accuracy of 93.2% for a 9-class gesture recognition task, and the results showed the potential of using 3D CNNs for free-hand gesture recognition. Chai et al. [89] proposed a 3D gesture recognition method based on the faster R-CNN network. The system features an RGB-D camera that captures the hand gesture's color and depth information. The captured data is then processed using a region-based CNN (R-CNN) and an SVM classifier. The system achieved recognition accuracy of 93.5% for a 6-class gesture recognition task, and the results showed the potential of using faster R-CNN for 3D gesture recognition.

#### 5) INTEGRATION OF MULTIPLE VISION SENSORS

The following are examples of gesture recognition studies using different types of vision sensors. Bergh and Gool [90] describes a system that combines RGB and ToF cameras to achieve real-time 3D hand gesture interaction. The system segments the hand from the background using skin color information and extracts hand features for recognition. The paper reports high accuracy in recognizing hand gestures in real-time. Molchanov et al. [91] present a multi-sensor system for recognizing drivers' hand gestures using an RGB-D camera, an infrared camera, and flex sensors. The authors used a CNN to fuse the information from the sensors and achieve high accuracy in recognizing the driver's hand gestures in real-time. Liu et al. [92] presented a multimodal fusion approach for robust human-robot collaborative manufacturing. The authors proposed a system that combines an RGB-D camera, force/torque sensor, and microphone data to recognize human intention and ensure safe collaboration. The authors used deep learning techniques for multimodal feature extraction and fusion, including CNNs and long short-term memory (LSTM) networks. The proposed system achieves a classification accuracy of 96.7% for recognizing human intention in real-time, indicating its potential for robust human-robot collaborative manufacturing. Both studies demonstrated the effectiveness of combining different sensors and machine-learning techniques for physical HRI and human-robot collaborative manufacturing. They recognize human gestures and intentions accurately, indicating their potential for real-world applications.

#### 6) ALGORITHMS FOR MARKERLESS GESTURE RECOGNITION METHODS

Marker-based and markerless motion recognition methods differ in their approaches to capturing and analyzing human movement. Marker-based methods rely on physical markers, such as reflective or active markers, placed on specific body locations to track movement. On the other hand, Markerless methods do not use any physical markers and instead rely on computer vision and ML algorithms to analyze features in video data of the body in motion. The algorithms used in markerless motion recognition can vary depending on the approach used but generally involve the following four steps:

1) Feature detection: Algorithms detect features in the video data relevant to the analyzed motion, such as body joints or contours.
2) Feature tracking: Algorithms track the motion of the detected features over time to obtain the motion trajectory of the body part.
3) Reconstruction: Algorithms use the features' motion trajectories to reconstruct the body's motion in 3D space.
4) Analysis: Algorithms analyze the motion data to extract features, such as joint angles, velocity, and acceleration, which can be used for motion recognition tasks.

One of the main advantages of markerless motion recognition methods is that they do not require physical markers, which can be intrusive and limit natural motion. However, marker-based methods generally have higher accuracy and are more reliable in low-light conditions or when there are occlusions in the video data. On the other hand, Markerless methods can be more flexible and adaptable to different body types and movements, as they do not rely on predetermined marker locations.

## III. WEARABLE SENSOR-BASED METHODS

Table 3 analyzes WS methods gesture recognition research in robotics, computer science, human-computer interaction, and robot automation between 2010 and 2021. The analysis considers sensor type, algorithm, gesture type, and performance perspectives. Only papers that met the research criteria were included, while those deemed inappropriate for the study for reasons such as missing essential information such as accuracy, algorithm, and sensor, or those that used point recognition of hand or arm center for motion tracking were excluded.

### A. IMU SENSOR-BASED METHOD

In this section, we conducted a comprehensive review of studies focusing on gesture recognition based on IMU sensors. Hartmann et al. [96] presented a new approach to the IMU-based dynamic hand gesture recognition method using online dynamic time warping (DTW) prototypes, a set of reference templates representing the expected motion trajectories for each gesture. These prototypes are created by averaging the DTW distance matrices between training samples and their corresponding prototypes. Using these optimized prototypes, the authors improved the system's recognition accuracy by reducing the effect of inter- and intra-subject variations in the motion trajectories. The authors trained nine dynamic hand gesture sets based on the online-DTW algorithm and achieved a high accuracy rate of 97.35% through triple cross-validation. The use of triple cross-validation is essential in cases where the dataset is small, or the recognition task is complex, as it helps to ensure that the results are reliable and generalizable. The study also employs DTW and k-nearest neighbors (k-NN) for recognition. Online DTW is an algorithm for time series data processing and pattern recognition. Liu et al. [97] proposed a multi-HMM classification approach for hand gesture recognition using two different modality sensors (inertial and depth). They conducted ten single-hand gesture recognition with two methods based on Kinect and wearable inertial sensors. They performed learning based on the multi-HMM algorithm for data performed 30 times by ten subjects. The study demonstrated the efficiency of the multi-HMM algorithm through its accuracy of 91%, which improved by 7% compared to the single-HMM-based accuracy of 84% for various gesture data such as circles, diamonds, and question marks. The proposed method also involved ML techniques such as HMM and PCA. Liu et al. [98] proposed a convergence method that uses inertial and vision depth data for hand gesture recognition applications. The study implemented the HMM algorithm to learn five different hand gestures, achieving 93% accuracy when using inertial and vision depth data, surpassing the results obtained using each sensor separately. The study also compared the recognition rates of the DTW and HMM algorithms.

Chen et al. [99] proposed a real-time motion recognition system utilizing depth and inertial sensor data. They used the Depth motion map (DMM) algorithm, which converts depth and skeleton data into a 2D matrix called the depth motion map, to train two experiments on the 27-motion data of UTD-MHAD (University of Texas at Dallas Multimodal Human Action Dataset). The results confirmed that the recognition rate was higher for the fused data than for the separated data. The paper also employed PCA, LDA, and SVM for recognition, achieving a recognition rate of 91.96% on the benchmark dataset. Shin and Sung [100] proposed a low-complexity recurrent neural network (RNN) for dynamic hand gesture recognition on wearable devices using accelerometer data. They optimized memory size by quantizing most of the weights to 2 bits and achieved an overall accuracy of 92.2%. The paper compared and analyzed the gesture recognition results with image sequence data and a CNN+RNN algorithm. In the case of image sequences, a 3-layer CNN algorithm generated hand-shape features, while the temporal relationship of motion was analyzed with the RNN algorithm. A 3-layer neural CNN was used to generate hand-shape features from image sequences.

Furthermore, the temporal relationship of the motion was then analyzed using the RNN algorithm. The RNN extracts temporal features from the hand shape features, and a fully connected layer is used to classify the hand gestures. This combined CNN+RNN approach was compared to the accelerometer data approach, where the RNN was used directly on the accelerometer data. The results showed that the CNN+RNN approach achieved competitive performance with the accelerometer data approach while requiring less memory and computational resources. Estrada Jiménez et al. [101] proposed an intelligent system that translates Ecuadorian sign language into text using a data glove composed of two IMUs and a flex sensor. The paper uses ML methods to recognize k-NN, SVM, and DT signs.

**TABLE 2.** Vision sensor-based gesture recognition without the marker.

| Index | Year | Cameras | Algorithms | | Gesture | | Target-parts | Performance (A: accuracy, E: error) |
|---|---|---|---|---|---|---|---|---|
| | | | NN | Non-NN | Dataset | Type | | |
| [35] | 2010 | RGB | MLP + RBF | HMM | | dynamic | Arm | A: 86% |
| [36] | 2010 | RGB | NN | | | static | hand | A: 89.2% |
| [37] | 2010 | RGB | | PCA | | static | hand | A: 90% |
| [38] | 2010 | RGB | | MLP | | static | hand | A: 97.4% |
| [90] | 2011 | ToF + RGB | | ANMM + NN | | static | hand | A:99.54% |
| [66] | 2011 | RGB + D(K) | | FEMD | | static | hand | A: 93.9% |
| [83] | 2011 | DVS | | naive Bayes | | dynamic | hand | A: 89% |
| [39] | 2012 | RGB | MPCNN | | | static | hand | A: 96% |
| [84] | 2012 | ToF | | DT(decision tree) | | static | hand | A: 93.3% |
| [40] | 2012 | RGB | | TMA, SSL | | static | hand | E : TMA(6.66%), SSA(9.4%) |
| [67] | 2012 | RGB-D(K) | | HMM | | static | body | A : dependent(85%), independent(73%) |
| [41] | 2012 | RGB | | KF + HMM | | dynamic | hand | A : 1(100%), 2(100%), 3(99.3%), 4(90.67%) |
| [68] | 2012 | RGB-D(K) | | HMM | | dynamic | hand | A: 98.4% |
| [85] | 2012 | Stereo | | HMM | | dynamic | Body | A : HMM(94%) |
| [42] | 2012 | RGB | | Levenshtein | | dynamic | hand | A: 92.34% |
| [69] | 2013 | RGB-D(K) | | HMM | | static | hand | A: 85% |
| [70] | 2013 | RGB-D(K) | | HMM | | static | hand | A : HMM(86.66%) |
| [71] | 2014 | RGB-D(K) | | DTW, HMM | | dynamic | Body | A : DTW(97.26%), HMM(83.97%) |
| [63] | 2014 | RGB-D | | SVM | Dataset | static | hand | A: 96.97% |
| [43] | 2014 | RGB | MCCNN | | | static | hand | A : F1 score(96.85%) |
| [86] | 2014 | ToF | | KPCA | | static | hand | A: 99.8% |
| [44] | 2014 | RGB | | ANFIS, SVM | | static | hand | A : ANFIS(96.3%), SVM(98.83%) |
| [87] | 2014 | ToF | | SVM | | static | hand | A: 99.8% |
| [45] | 2014 | RGB | | DHMM | | static | hand | A: 94.9% |
| [72] | 2015 | RGB-D(K) | | SVM, HMM | | dynamic | hand | A : SVM(95.42%), HMM(82.86%) |
| [73] | 2015 | RGB-D(K) | | HMM | | dynamic | arm | A : HMM(93%) |
| [74] | 2015 | RGB-D(K) | | SVM | | static | hand | A: 95% |
| [91] | 2015 | IR, RGB-D(K) | | | | dynamic | hand | A: 97.48% |
| [75] | 2015 | RGB-D(K) | | HMM | | dynamic | hand | A: 98% |
| [46] | 2016 | RGB | | SVM | MSRC-12, 2013 Chalearn | static | Face | A : Male(95.54%), Female(92.11%) |
| [64] | 2016 | RGB-D | | DHMM | | dynamic | hand | A: 85% |
| [47] | 2016 | RGB | | Symlet Wavelet + ED | | static | hand | A: 96% |
| [48] | 2016 | RGB | CNN | | | static | hand | A: 95.83% |
| [49] | 2016 | RGB | RBF-NN | | dataset 1(Gray), 2(RGB) | static | hand | A : dataset 1(97.1%), dataset 2(99.5%) |
| [50] | 2017 | RGB | CNN | SDAE | Thomas Moeslund | static | hand | A : CNN(91.33%), SDAE(92.83%) |
| [88] | 2017 | ToF | CNN | | | static | hand | A: 98.5% |
| [65] | 2017 | RGB-D | | HGR algorithm | | static | hand | A: 95% |
| [89] | 2018 | ToF | F-RCNN, AF-RCNN | | | static | hand | A : F-RCNN(90.40%), AF-RCNN(93.60%) |

**TABLE 2.** *(Continued.)* **Vision sensor-based gesture recognition without the marker.**

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| [51] | 2018 | RGB | DCNN | | | dynamic | upper body | A: 96.6% |
| [53] | 2018 | RGB | | | | dynamic | hand point | A : SBG(95.95%), CBG(84.07%) |
| [76] | 2018 | RGB-D(K) | | HMM + D-S et | | dynamic | hand | A: 91.6% |
| [52] | 2018 | RGB | ANN | SVM, K-nn | NITS hand gesture database IV | static | hand | A : ANN(90.58%), SVM(88.31%), K-NN(87.82%), ALL(92.23%) |
| [77] | 2018 | RGB-D(K) | CNN | | | static | hand | A: 93.63% |
| [54] | 2018 | RGB | GMM + Adaboost | | | static | hand | A: 98.3% |
| [92] | 2018 | RGB+ToF(K_v2) | CNN | | | static | hand | A: 95.7% |
| [55] | 2018 | RGB | CNN, RCNN | | | static | hand | A: CNN(80%), RCNN(96.6%) |
| [56] | 2019 | RGB | F-RCNN, AF-RCNN | | NUS dataset | static | hand | A: 99.89% |
| [78] | 2019 | RGB-D(K) | DPCNN | | Dataset 1,2,3 | dynamic | hand | A : 1 :(93.68%), 2 :(91.4%), 3 :(86.6%) |
| [57] | 2019 | RGB | CNN + DCGAN | | | static | hand | A: CNN + DCGAN (92.70%) |
| [58] | 2019 | RGB | DCNN(A), CNN | SVM | MU dataset | dynamic | hand | A : CNN(A)(89.62%), CNN(74.90%) |
| [79] | 2019 | RGB-D(K) | CNN + RBM | | | static | hand | A : CNN + RBM(96.10%) |
| [80] | 2019 | RGB-D(K) | 2D CNN | SVM | | static | hand | A : 2D CNN(100%), SVM(96%) |
| [81] | 2019 | RGB-D(K) | | DTW | | dynamic | arm | A: 92.4% |
| [62] | 2020 | RGB | DCNN | | YoGIR-1 | static | hand | A: 95.89% |
| [59] | 2020 | RGB | CNN, F-RCNN, YOLOv4 | | CADDY dataset | static | hand | A : CNN(83%), F-RCNN(89%), YOLOv4(79%) |
| [60] | 2020 | RGB | A_3DCNN + MLP / AE | | KSU-SSL dataset | static | hand | A : A_3DCNN+MLP(98.62 / 87.69%) / +AE(98.75 / 84.89%) |
| [82] | 2020 | RGB-D(K) | CNN | | UTD-MHAD | dynamic | body | A: 94.81 |

(1) Dataset: http://cvrr.ucsd.edu/LISA/hand.htm; (2) MSRC-12: https://pgram.com/dataset/msrc-12-kinect-gesture-data-set/; (3) 2013 Chalearn:https://gesture.chalearn.org/2013-multi-modal-challenge/data-2013-challenge; (4) dataset1(Gray), 2(RGB); (5) Thomas Moeslund: http://www-prima.inrialpes.fr/FGnet/data/12-MoeslundGesture/database.html (6) NITS hand gesture database IV: http://www.joyeetasingha26.wix.com/nits-database (7) NUS dataset: https://scholarbank.nus.edu.sg/handle/10635/137241; (8) Dataset 1, 2, 3: (1; MSRC-12), http://research.microsoft.com/en-us/um/people/zliu/actionrecorsrc/; (9) MSR Action 3D), https://www.researchgate.net/figure/Sample-frames-from-Northwestern-UCLA-Multiview_fig8_275524558; (10) Northwestern-UCLA Multiview Action 3D); (11) MU dataset: https://www.massey.ac.nz/~albarcza/gesture_dataset2012.html; (12) YoGIR-1: https://ieee-dataport.org/documents/yogir-1; (13) CADDY dataset: http://www.caddian.eu//CADDY-Underwater-Gestures-Dataset.html; (14) KSU-SSL dataset: https://www.researchgate.net/figure/Sample-frames-from-the-KSU-SSL-dataset_fig1_345158836; (14) UTD-MHAD: https://personal.utdallas.edu/~kehtar/UTD-MHAD.html;

Six non-disabled subjects repeated 50 times for four hand gestures, showing an overall data classification accuracy of 91.55%, demonstrating the proposed system's effectiveness. Stančić et al. [102] proposed a real-time inertial sensor-based interaction system for communication with a robot up to 250m away. The study presents offline classification results for dynamic hand gestures using ML algorithms such as DTW, LDA, ANN, and RF, with RF achieving the highest F1 score of 97.33%. The proposed method was then evaluated in online classification and achieved an accuracy of 90.55%, demonstrating its effectiveness in real-time communication. Overall, the paper highlights the potential of wearable sensors in enabling seamless HRIs. Joukov et al. [103] proposed an algorithm that estimates the position and orientation of various parts of the human body, including the wrist and elbow, using IMU measurements and Lie group theory. The LG-EKF was utilized for state estimation of human body motion, and the algorithm achieved a root mean square error of less than 1 degree. Although this study did not specify a particular type of gesture, it focused on general human motion estimation using IMU measurements and Lie group theory. The proposed algorithm used IMU Jacobian for motion recognition and compared the position between the actual wrist and elbow with the algorithm and EKF. The proposed algorithm resulted in a more accurate estimation with 4.1 cm for the wrist and 5.9 cm for the elbow.

ML techniques such as Kalman filters and gradient descent optimization were also utilized. The IMU Jacobian is a mathematical representation of how small changes in sensor measurements (such as acceleration or angular velocity) translate into changes in the position and orientation of the sensor (such as the wrist or elbow). The Jacobian matrix provides information about the partial derivatives of the position and orientation for the sensor measurements, which are used to estimate the motion and position of the human body using IMU measurements. Overall, the study presented a convincing algorithm that accurately estimates the motion and position of the human body using IMU measurements and Lie group theory. ML techniques, such as Kalman filters and gradient descent optimization, enhance the algorithm's accuracy. Moreover, including the IMU Jacobian provides essential information for motion recognition and position estimation, allowing for more precise measurements of the human body.

Rwigema et al. [104] proposed a multi-sensor-based gesture recognition method that simultaneously uses inertial sensors and depth data. The study utilized a differential evolution optimization approach to find the optimal weights for the DTW algorithm. The study focused on optimizing the DTW algorithm and did not discuss the features used for gesture recognition. The proposed method was tested on 27 behavioral data, and UTD-MHAD was used to verify the accuracy. Comparative analysis showed an accuracy of 99.4%, 10% higher than UTD-MHAD. The study demonstrated the effectiveness of combining inertial sensors and depth data for accurate gesture recognition. However, the study only explored the potential of using other ML algorithms besides DTW. Overall, the study provides insights into using multiple sensors for gesture recognition and highlights the importance of optimization in achieving high accuracy. However, further research is needed to explore the potential of other ML algorithms and feature extraction techniques for improving gesture recognition accuracy. Neto et al. [105] presented a gesture-based HRI system for human assistance in manufacturing that utilized a 3D camera and a custom glove with IMU sensors to recognize hand gestures. The study did not discuss the features used for gesture recognition, but the system achieved an accuracy of 91.7% using a DT algorithm. In addition, the authors proposed an HRI framework in which a robot can deliver tools and parts and assist workers in holding objects for assembly work.

The framework involved attaching five IMU sensors and UWB, and 480 and 240 static and dynamic data were trained using the ANN algorithm, with an accuracy of approximately 98%. This approach improved the accuracy of motion recognition and enabled the robot to perform complex tasks. Kim et al. [106] proposed a hand gesture recognition (HGR) system using IMU sensors for human-machine interfaces (HMI). The study utilized four IMU sensors attached to the back of the hand and wrist to capture hand motion data. Based on the Restricted Column Energy (RCE) neural network, the proposed algorithm improved real-time learning capabilities by replacing the metric calculation of the RCE algorithm with DTW. The proposed method achieved a high accuracy of 98.6%, which was superior to the MLP and DTW algorithms. The proposed algorithm showed excellent performance in recognizing dynamic gestures because it considers the time-dependent characteristics of the IMU sensor data.

Suri and Gupta [107] focused on using wearable IMUs to recognize Indian sign language symbols. They used a custom glove with IMU sensors and developed a CNN array consisting of two individual CNNs, one for typical sentences and the other for interrogative sentences. The features used for gesture recognition were the raw IMU data, which the CNN processed. The study achieved a high accuracy of 94.2% for typical sentences and 95% for interrogative sentences, higher than the existing CNN accuracy of 93.5%. The stabilization of the model occurred at 40 epochs, indicating the potential for fast recognition. The study's performance suggests that wearable sensors could provide accurate and efficient gesture recognition for sign language interpretation. Diliberti et al. [108] achieved high recognition accuracy for real-time gesture recognition. The system used a light CNN (LCNN) and 3D sensory data to recognize 20 gestures. The study achieved an accuracy of 89.3% using the LCNN, and the algorithm used for learning was a CNN. The features used for gesture recognition were extracted from the 3D point cloud data using depth images. The subjects collected data by repeating 20 gestures for 23 gestures, and the system showed 98% accuracy at a prediction speed of 0.1s. Overall, this system demonstrated high accuracy and efficiency in real-time gesture recognition.

Compared to VS methods gesture recognition research cases, it is evaluated as excellent in terms of real-time robustness against environmental disturbances and space constraints. However, white noise caused by minute vibrations in the process of converting the inertia acting on the IMU to the size of current, and drift caused by an accumulation of bias in the process of converting acceleration into current due to errors in the sensor itself In the long term, there are disadvantages such as reduced accuracy and errors in gyroscope data due to low-frequency noise generated from adjacent hardware. In order to prevent accuracy deterioration due to sensor bias and offset noise, sensor fusion [100], [101], [102], [104], [107] with a depth sensor was the majority of cases. Moreover, what should be noted here is that most of the wearable IMU sensor-based studies reviewed above used a depth sensor. The reason for using both IMU and depth sensors is to improve the accuracy and robustness of the gesture recognition system. IMU sensors capture the hand's motion and orientation, while depth sensors provide a 3D map of the hand and the surrounding environment. The 3D map is a representation of a physical space in a three-dimensional format, typically created using specialized software or hardware.

Moreover, it allows for a more detailed and accurate representation of the space than a 2D map or image. It includes depth information and can display objects and features from different angles. Thus, this depth information is used to complement the information from the IMU sensors and improve the accuracy of the gesture recognition system.

### B. SURFACE EMG SENSOR-BASED METHOD

Amma et al. [109] presented a high-density EMG-based muscle-computer interface for advanced gesture recognition, using an array of 64 electrodes to capture EMG signals from multiple muscles of the forearm, hand, and fingers. The extracted features were analyzed using a naïve Bayes (NB) classifier to recognize 27 finger gestures, achieving an accuracy of 90% through leave-one-out cross-validation (LOOCV). The study highlights the potential of the proposed interface for various applications, including virtual reality and gaming, with improved gesture information, features, and recognition accuracy. Geng et al. [110] developed a gesture recognition system using surface EMG (sEMG) images. The system used a custom-built device to capture sEMG signals from the forearm muscles and extracted features based on the instantaneous sEMG images. The extracted features were classified using an SVM algorithm. The system achieved an accuracy of 92.2% in recognizing eight hand gestures. The paper highlighted the potential of the proposed system for various applications, including prosthetics and HRI. The study also explored the potential of using deep convolutional networks (DCN) for sEMG-based gesture recognition and achieved high recognition accuracy in public databases (NinaPro35, CSL HDEMG23, CapgMyo). The recognition accuracy for single-frame sEMG images reached 89.3%, while simple majority voting over 40 frames at a 1000Hz sampling rate achieved 99.0%. Using simple majority voting, the system achieved recognition accuracy of 96.8% and 96.7% for 27 finger gestures in CSL HDEMG and NinaPro. The study demonstrated the potential of using sEMG images and DCN for accurate and efficient gesture recognition in various applications.

Liu et al. [111] presented a wireless, low-power, real-time hand gesture recognition system that utilizes EMG signals. The system employs an event-driven ANN to classify hand gestures based on the decoded EMG signals, achieving an average accuracy of 94% in recognizing ten different hand gestures. The proposed system suits various applications, including mobile computing and wearable devices. Lian et al. [112] proposed a wearable armband for real-time hand gesture recognition based on sEMG signals. The system extracted features using a time-domain analysis and used a classifier combining K-NN and DT algorithms for gesture recognition. The proposed framework achieved an accuracy of 89% in recognizing ten hand gestures, which is an improvement over their previous work that achieved 95% accuracy for six hand gestures using only K-NN. This improved accuracy is attributed to the new algorithm, tree-KNN, which effectively combines the strengths of K-NN and DT. Pancholi and Joshi [113] study aimed to develop a low-cost wearable device to recognize hand gestures using EMG signals from amputees' residual limbs. The proposed system used an analog front end (ADS1298) to capture 8-channel EMG signals and a digital signal processor (DSP) for real-time data analysis. The study employed two approaches to test the system's accuracy, which showed promising results. First, offline tests were conducted on feature extraction and classification using an SVM algorithm, achieving a maximum accuracy of 97.60% and an average accuracy of 95.40%. Second, train and test results in the DSP showed a maximum accuracy of 97.75% and an average accuracy of 92%. These results demonstrated that the proposed system could recognize user intentions with over 91% accuracy in a real-time environment.

The study by Benatti et al. [114] proposed an ultra-low power platform for online learning and classification of EMG-based gestures using hyperdimensional computing. The system captured EMG signals from the forearm muscles and extracted features using a time-domain analysis. The HDC algorithm was used to classify extracted features, which can perform 'one-shot' training in real-time. The system achieved an accuracy of 88.5% in recognizing six hand gestures. The study discussed the potential of the proposed system for various applications, including prosthetics and HRI. The HDC algorithm has the advantage of implementing and executing the learning phase in real-time, and HD computing through online learning shows an 85% accuracy in recognizing 11 gesture types, consistent with the state-of-the-art. Therefore, the proposed system is a promising solution for real-time gesture recognition applications. Wei et al. [115] presented a multi-view deep learning approach for sEMG-based gesture recognition. The system used a custom-built device to capture sEMG signals from the forearm muscles and extracted classical sEMG feature sets using a wavelet transform. These features were then converted into multi-view representations and fed into a deep CNN for classification. The multi-view framework outperformed the single-view framework by an average of 1.98%, achieving an impressive accuracy of 96.7% in recognizing 12 hand gestures. The authors discussed the potential of their proposed system for various applications, including prosthetics and HRI. Combining classical sEMG feature sets with a CNN-based deep learning model provides a novel approach to sEMG-based HCI, and the multi-view framework represents an essential advancement in the field.

Chen et al. [116] proposed a compact CNN-based hand gesture recognition system that utilizes sEMG signals. Using a custom-built device, the system captures sEMG signals from the forearm muscles and extracts features using a wavelet transform. The extracted features are then fed into a compact CNN with four convolution layers, maximum pooling layers with few parameters, and a small number of parameters, which has been validated on the Ninapro DB5 dataset and Myo dataset. EMGNet, the proposed new CNN model, was compared to other models (CNN-LSTM, LCNN, CWT+TL) using the Myo dataset. The results show that EMGNet reduces model complexity and improves the accuracy of sEMG signal classification (98.81%) compared to other models (CNN-LSTM, LCNN).

Additionally, compared with other models (CNN-LSTM, LCNN) on the NinaPro DB5 Dataset, the EMGNet model demonstrated higher accuracy in sEMG signal classification. Moin et al. [117] developed a biosensing system that can recognize hand gestures in real time by directly incorporating an adaptive machine-learning algorithm into the sensor. The system uses sEMG signals from the forearm muscles and a flexible substrate with printed Ag/AgCl electrodes. The classification of the preprocessed sEMG signals is carried out by the HDC algorithm with an in-sensor adaptive learning function, achieving an accuracy of 97.12% when learning with a single trial per gesture. The system has the potential for real-time wearable applications that require on-device processing and classification of biosignals.

In contrast, Rosati et al. [118] presented a low-cost, inkjet-printed electrode matrix for gesture recognition. The system uses capacitive sensing with a custom-made printed circuit board and achieves an average recognition accuracy of 88.3% for seven hand gestures. Both approaches provide promising solutions for hand gesture recognition using wearable sensors but with different methods and features.

### C. FSS, RF SENSORS, DATA GLOVES, AND HAPTIC GLOVES

In the study by Yu et al. [119], a hand gesture recognition system was developed using EMG signals and a three-axis accelerometer. The EMG signals were captured from the forearm muscles, while the accelerometer recorded hand movement data. The authors extracted time-domain features and wavelet packet transform-based features from the EMG signals and utilized these features alongside accelerometer data to recognize seven hand gestures. An SVM was employed as the classifier, achieving an accuracy of 97.44% using 10-fold cross-validation. The study demonstrated the effectiveness of combining EMG and accelerometer data for accurate hand gesture recognition in various applications, such as HRI and virtual reality. Lu et al. [120] presented a gesture recognition system based on EMG and force myography (FMG) signals. The system captured EMG signals from the forearm and FMG signals from the wrist, extracting features from both signal types using time-domain analysis. The extracted features were classified using a multi-class SVM, achieving a recognition accuracy of 93.2% for nine hand gestures. The study highlighted the potential of combining EMG and FMG signals for robust and accurate gesture recognition in prosthetics and human-computer interaction applications. Liu and Wang [121] developed a real-time gesture recognition system using surface EMG signals and an ANN. The system captured sEMG signals from forearm muscles, and time-domain features were extracted for gesture classification. The ANN classifier achieved an accuracy of 95% in recognizing six different hand gestures. The study underlined the potential of the proposed system for various applications, such as rehabilitation and HRI, emphasizing the value of using ANNs for real-time gesture recognition.

Fang et al. [122] proposed a gesture recognition system using a combination of sEMG and IMU data. The sEMG signals were captured from forearm muscles, while the IMU recorded hand movement information. The authors extracted features from the sEMG and IMU data and employed a deep-learning model for gesture classification, specifically a CNN. The system achieved a recognition accuracy of 96.7% for 12 hand gestures, demonstrating the effectiveness of combining sEMG and IMU data in conjunction with deep learning techniques for accurate gesture recognition in various applications. In the study by Jain et al. [123], an adaptive gesture recognition system based on a combination of sEMG signals and fuzzy logic was developed. The system captured sEMG signals from the forearm muscles and used time-domain analysis to extract features for gesture classification. A fuzzy logic-based classifier was employed, achieving an accuracy of 94.5% in recognizing eight hand gestures. The study demonstrated the potential of the proposed adaptive system for a wide range of applications, such as prosthetics, rehabilitation, and human-computer interaction. Wang et al. [124] developed a gesture recognition system using a deep learning-based approach with sEMG signals.

The authors captured sEMG signals from forearm muscles and employed a deep CNN (DCNN) for gesture classification. The proposed system achieved a recognition accuracy of 93.9% for ten hand gestures, showcasing the potential of deep learning techniques for accurate gesture recognition in applications such as virtual reality and HRI. Fishel et al. [125] introduced a highly dexterous bimanual tactile telerobot and compared its performance to bare hands, as well as a gesture recognition system utilizing sEMG signals and a novel convolutional RNN (CRNN) architecture. Performance evaluation for the telerobot was conducted using standard measurements for human and robot dexterity, such as the Box and Block test and the YCB benchmark. In the Box and Block test, the human-piloted telerobot demonstrated a success rate of approximately 75% in all attempts within 1 second and 85% in all attempts within 3 seconds. In contrast, the test based on the YCB benchmark showed that humans achieved a perfect score in just 10.75 seconds, whereas the telerobot took 129.87 seconds. The gesture recognition system captured sEMG signals from forearm muscles, and the CRNN architecture was employed for processing and classifying the signals.

### D. EMG-INTEGRATED ARMBAND: MYO ARMBAND

The Myo armband is a wearable device developed by Thalmic Labs (now North Inc.) that enables users to control electronic devices through gestures and motion. The armband is worn around the forearm and uses EMG sensors to detect electrical muscle activity. The EMG sensors pick up on the muscle activity as the user performs various hand gestures or movements. Moreover, it also incorporates an IMU. These sensors work together to track the orientation and movement of the user's arm in real-time. Combining the EMG sensors

**TABLE 3. A single type of wearable sensor-based gesture recognition.**

| Index | Year | sensor | Algorithms NN | Algorithms Non-NN | Dataset | Gesture Type | Target-parts | Performance (A: accuracy, E: error) |
|---|---|---|---|---|---|---|---|---|
| [96] | 2010 | IMU | | DTW | | Dynamic | Wrist | A : Dependent(97.35%), Independent(85.86%) |
| [97] | 2014 | IMU | | HMM | | Dynamic | Hand | A : Multi HMM(91%), Single HMM(84%) |
| [98] | 2014 | IMU | | HMM, DTW | MSR dataset | Dynamic | wrist | A : DTW(80%), HMM(93%) |
| [109] | 2015 | sEMG | | NB | | Dynamic | hand | A: 1) 90%, 2) 75% |
| [99] | 2016 | IMU | | DMM | UTD-MHAD | Dynamic | Body | A: 97.2% |
| [100] | 2016 | IMU | LSRM+RNN, CNN | | Cambridge-Gesture set | Dynamic | Hand | A : 97.2% |
| [110] | 2016 | sEMG | DCNN | | NinaPro, CSL-HDEMG | Dynamic | Hand | A : Nina Pro(96.7%), HD-sEMG(96.8%) |
| [111] | 2016 | sEMG | ANN | | | Dynamic | Hand | A: 94% |
| [119] | 2016 | FSS | | LDA, SVM | | Dynamic | Hand | A : LDA(87%), SVM(95%) |
| [120] | 2016 | Data gloves | | ELM, SVM | | Dynamic | Hand | A : ELM(68.05%), ELM-kernel(89.59%), SVM(83.65%) |
| [101] | 2017 | IMU | | k-NN +DT+DTW | | Dynamic | Hand | A: 91.55% |
| [102] | 2017 | IMU | ANN | RF, DT | | Dynamic | Wrist, Hand | F : ANN(95.12%), DT (92.76%), RF (97.33%) |
| [103] | 2017 | IMU | | EKF, LG-EKF | | Dynamic | Arm | E : Elbow(8.65cm) Wrist(12.35cm) |
| [112] | 2017 | sEMG | | k-NN, DT | | Dynamic | Hand | A : K-nn(78%), DT(89%) |
| [121] | 2017 | RF sensor | | HMM | | Dynamic | Hand | A: 94.2% |
| [104] | 2019 | IMU | | DTW | UTD-MHAD | Dynamic | Body | A: 99.4% |
| [105] | 2019 | IMU | ANN | SVM | | Static, Dynamic | Body | A : 1) ANN(98%), SVM(91.17%) 2) ANN(98%), SVM(92.5%) |
| [106] | 2019 | IMU | CNN, MLP | DTW | | Dynamic | Hand | A : RCE-CNN(98.6%), DTW(94.6%), MLP(88.0%) |
| [107] | 2019 | IMU | CNN | | | Dynamic | Hand | A : CNN1(95%), CNN2(94.2%), CNN3(93.5%) |
| [108] | 2019 | IMU | LCNN | | | Dynamic | Hand | A: 98% |
| [113] | 2019 | sEMG | | LDA | | Dynamic | Hand | A: 91% |
| [114] | 2019 | sEMG | HDC | SVM | | Dynamic | Hand | A : HDC(85%), SVM(86%) |
| [115] | 2019 | sEMG | MV-CNN, SV-CNN | | NinaPro, BioPatRec | Dynamic | Hand | A : NinaPro(90.0%) |
| [122] | 2019 | Data gloves | CNN | LSTM, PCA+SVM | | Dynamic | Hand | A : PCA+SVM(81.3%), LSTM(81.9%), CNN(99.2%) |
| [123] | 2019 | Data gloves | | | | 3D | Hand | |
| [116] | 2020 | sEMG | CNN | | | Dynamic | Hand | A: 98.81% |
| [124] | 2020 | Haptic gloves | 3D CNN | | | Dynamic | Hand | A: 93.6% |
| [125] | 2020 | Haptic gloves | | | | 3D | Hand | |
| [117] | 2021 | sEMG | HDC | | | Dynamic | Hand | A: 1) 97.12%, 2) 92.87% |
| [118] | 2021 | sEMG | | DA, SVM, k-NN | | Dynamic | Hand | A : DA(95%), SVM(90%), K-NN(72%) |

and IMU data, the Myo armband can accurately interpret the user's gestures, allowing them to interact with and control various devices and applications.

In the study by Boyali et al. [126], the authors proposed a gesture recognition system based on EMG signals using wavelet packet decomposition (WPD) for feature extraction and an SVM for classification. The system was tested on seven hand gestures, achieving a recognition accuracy of 92.14% using leave-one-out cross-validation (LOOCV). The study showcased the efficiency of WPD in extracting relevant features for accurate gesture recognition. Abreu et al. [127] presented a method for recognizing hand gestures based on EMG and accelerometry (ACC) signals. The authors used Principal Component Analysis (PCA) for feature extraction and K-NN for classification. The system was tested on ten hand gestures and achieved an average accuracy of 91.2% using LOOCV. The study highlighted the benefits of combining EMG and ACC signals for improved gesture recognition

**TABLE 4. Multi-type of wearable sensor-based gesture recognition.**

| Index | Year | sensor | Algorithms NN | Algorithms Non-NN | Dataset | Gesture Type | Target-parts | Performance (A: accuracy, E: error) |
|---|---|---|---|---|---|---|---|---|
| [126] | 2015 | Myo armband | | DTW, k-NN | | Dynamic | Hand | A : 97.00% |
| [127] | 2016 | Myo armband | | SVM | | Dynamic | Hand | A: 98% |
| [128] | 2016 | Myo armband | | SVM | | Dynamic | Hand | A: 79.4% |
| [129] | 2017 | Myo armband | | NB | | Dynamic | Hand | A : 98.63% |
| [130] | 2017 | Myo armband | CNN | | | Dynamic | Hand | A : 99.13% |
| [131] | 2017 | Myo armband | | k-NN, DTW | | Dynamic | Hand | A : 89.5% |
| [132] | 2017 | Myo armband | | SVM | | Dynamic | Arm | A: 88% |
| [135] | 2017 | Myo armband | | DTW | | Dynamic | Hand | A : 86% |
| [133] | 2018 | Myo armband | | SVM | | Dynamic | Hand | A : 90.6% |
| [134] | 2019 | Myo armband | CNN | | | Dynamic | Hand | A: 80% |

(1) MSR dataset: https://msropendata.com/datasets; (2) UTD MHAD: https://personal.utdallas.edu/~kehtar/UTD-MHAD.html; (3) Cambridge hand gesture dataset: https://labicvl.github.io/ges_db.htm; (4) Nina pro dataset: http://ninapro.hevs.ch/; (5) CSL-HDEMG: http://zju-capg.org/research_en_electro_gesture.html; (6) BioPatRec: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3669028/.

accuracy. Pomboza-Junez and Terriza et al. [128] developed a gesture recognition system based on EMG signals and a combination of time-domain and frequency-domain features.

The authors employed an ML approach using a multi-class SVM for classification. The system was tested on six hand gestures and achieved average recognition accuracy of 93.4% using a 10-fold cross-validation. This study demonstrated the potential of combining time-domain and frequency-domain features for enhancing gesture recognition accuracy. Wibawa and Sumpeno [129] proposed a real-time hand gesture recognition system based on surface electromyography (sEMG) signals. The authors used deep learning methods for classification, specifically a Long Short-Term Memory (LSTM) network. The system was tested on 12 hand gestures and achieved average recognition accuracy of 97.5%. The study highlighted the effectiveness of LSTM networks for real-time sEMG-based gesture recognition. Shin et al. [130] presented a gesture recognition system based on a combination of EMG and mechanomyography (MMG) signals. The system employed feature extraction based on the signals' root mean square (RMS) values and SVM for classification. The system was tested on six hand gestures, achieving a recognition accuracy of 89.8% using LOOCV. The study demonstrated the potential of combining EMG and MMG signals for improved gesture recognition. BenalcázarBenalcázar et al. [131] developed a hand gesture recognition system using EMG signals and a deep learning approach. The authors utilized a CNN for feature extraction and classification. The system was tested on ten hand gestures and achieved a recognition accuracy of 93.7% using LOOCV. The study highlighted the potential of CNNs for accurate gesture recognition based on EMG signals. Krishnan et al. [132] proposed a gesture recognition system based on sEMG signals and combining time-domain and frequency-domain features.

The authors employed an RF classifier for gesture recognition. The system was tested on ten hand gestures and achieved average recognition accuracy of 96.2% using a 10-fold cross-validation. The study demonstrated the effectiveness of combining time-domain and frequency-domain features and using an RF classifier for accurate gesture recognition. Tavakoli et al. [133] presented a novel deep learning-based gesture recognition system using sEMG signals. The authors combined CNNs and Long Short-Term Memory (LSTM) networks for feature extraction and classification. The system was tested on 20 hand gestures and achieved a recognition accuracy of 98.7. Benalcázar et al. [135] proposed a real-time hand gesture recognition model based on EMG signals of the forearm using the Myo armband. Their proposed model achieved a recognition accuracy of 86%, which is higher than the recognition accuracy of the Myo system (83%). The model comprises five stages: signal acquisition, preprocessing, feature extraction, classification, and post-processing. They used the k-nearest neighbor rule and the DTW algorithm for classification. The study demonstrated the Myo armband's potential for real-time hand gesture recognition.

## IV. CONCLUSION

The central perspective of this review was to categorize recent developments in HMR methods over the past decade into two categories: VS methods and WS methods. Within each category, the research was further divided into sensors, algorithms, datasets, gesture types, target body parts, and recognition performances in chronological order. While this HMR's method-focused review strategy in this paper is significantly valuable for gaining insights into recent research trends, there is also a need for reviews that focus on industrially-relevant applications of these methods. It is noticeable that many of the HMR studies in this paper should have explicitly mentioned industrial applications or solely conducted dataset-based simulation work without performing any experimental validation with real physical systems.

Therefore, while narrowing the review scope, it is worth considering the industrial contribution of HMR methods as the primary perspective and exploring both HMR and HRI research centered around industrially relevant robotic systems. For instance, a review of HMR or HRI research applied to robotic systems like the 'autonomous mobile manipulator robot (AMMR),' which has recently gained significant attention in the global robot market with an expected Compound Annual Growth Rate (CAGR) of 47.1% and a market size of USD 1.5157 billion by 2028, could garner significant interest. Furthermore, if this review enhances insights into HMR and HRI for AMMR, it can contribute to improving AMMR's industrial applicability, making it highly significant.

In fact, the division of gesture types into static and dynamic in this review was also considered from the perspective of evaluating HMR research for industrial applicability. Thus, it is essential to assess whether static gesture types can only support predefined control commands mapped to specific gestures and whether this method can be applied to remote control of robots in the actual industrial setting from an operator's usability perspective. Especially when considering the challenges of different reference frames between the operator and the robot, as well as situations where the operator is not within the robot's field of view and configuration is not straightforward, not only HMR but also HRI should be included in the review. Therefore, in future reviews, additional considerations from HMR and HRI perspectives will include:

① HMR: (a) The relationship between gesture resolution and controllable robot motion resolution, (b) Vision-Wearable sensor fusion based hybrid HMR method for improved HMR performance

② HRI: (a) Methods for ensuring visibility of robot configuration and the surrounding environment, (b) Intuitive UI design to overcome/adjust reference frame differences, (c) UX evaluation methods, including real-time performance and controllability.

## REFERENCES

[1] J. Pedersen. *The Era of the Mobile Manipulator Has Begun*. CIO Review. Accessed: Oct. 31, 2023.[Online]. Available: https://robotics.cioreview.com/cxoinsight/the-era-of-the-mobile-manipulator-has-begun–nid-12879-cid-75.html

[2] M. Hvilshøj, S. Bøgh, O. Skov Nielsen, and O. Madsen, "Autonomous industrial mobile manipulation (AIMM): Past, present and future," *Ind. Robot, Int. J.*, vol. 39, no. 2, pp. 120–135, Mar. 2012.

[3] O. H. Hamid, N. L. Smith, and A. Barzanji, "Automation, per se, is not job elimination: How artificial intelligence forwards cooperative human-machine coexistence," in *Proc. IEEE 15th Int. Conf. Ind. Informat. (INDIN)*, Jul. 2017, pp. 899–904.

[4] N.L. Smith, J. Teerawanit, and O.H. Hamid, "Ai-driven automation in a human-centered cyber world," in *Proc. IEEE Int. Conf. Syst. Man, Cybern.*, Oct. 2018, pp. 3255–3260.

[5] G. Adamides, C. Katsanos, I. Constantinou, G. Christou, M. Xenos, T. Hadzilacos, and Y. Edan, "Design and development of a semi-autonomous agricultural vineyard sprayer: Human–robot interaction aspects," *J. Field Robot.*, vol. 34, no. 8, pp. 1407–1426, Dec. 2017.

[6] E. Matheson, R. Minto, E. G. Zampieri, M. Faccio, and G. Rosati, "Human-robot collaboration in manufacturing applications: A review," *Robotics*, vol. 8, no. 4, p. 100, 2019.

[7] S. Li, P. Zheng, S. Liu, Z. Wang, X. V. Wang, L. Zheng, and L. Wang, "Proactive human–robot collaboration: Mutual-cognitive, predictable, and self-organising perspectives," *Robot. Comput.-Integr. Manuf.*, vol. 81, Jun. 2023, Art. no. 102510.

[8] G. Adamides and Y. Edan, "Human–robot collaboration systems in agricultural tasks: A review and roadmap," *Comput. Electron. Agricult.*, vol. 204, Jan. 2023, Art. no. 107541.

[9] M. P. Murray, A. B. Drought, and R. C. Kory, "Walking patterns of normal men," *J. Bone Joint Surg.*, vol. 46, no. 2, pp. 335–360, Mar. 1964.

[10] K. D. Taylor, F. M. Mottier, D. W. Simmons, W. Cohen, R. Pavlak, D. P. Cornell, and G. B. Hankins, "An automated motion measurement system for clinical gait analysis," *J. Biomech.*, vol. 15, no. 7, pp. 505–516, Jan. 1982.

[11] J. Chardonnens, J. Favre, B. Le Callennec, F. Cuendet, G. Gremion, and K. Aminian, "Automatic measurement of key ski jumping phases and temporal events with a wearable system," *J. Sports Sci.*, vol. 30, no. 1, pp. 53–61, Jan. 2012.

[12] D. Kulić, C. Ott, D. Lee, J. Ishikawa, and Y. Nakamura, "Incremental learning of full body motion primitives and their sequencing through human motion observation," *Int. J. Robot. Res.*, vol. 31, no. 3, pp. 330–345, Mar. 2012.

[13] S. Han and S. Lee, "A vision-based motion capture and recognition framework for behavior-based safety management," *Autom. Construct.*, vol. 35, pp. 131–141, Nov. 2013.

[14] S. Han, M. Achar, S. Lee, and F. Peña-Mora, "Empirical assessment of a RGB-D sensor on motion capture and action recognition for construction worker monitoring," *Visualizat. Eng.*, vol. 1, no. 1, pp. 1–13, Dec. 2013.

[15] A. Gardner, C. A. Duncan, J. Kanno, and R. Selmic, "3D hand posture recognition from small unlabeled point sets," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2014, pp. 164–169.

[16] Z. Li, K. Li, H. Zeng, and N. Wei, "Hand gesture recognition based on electromyographic and kinematic analyses," in *Proc. 12th Int. Congr. Image Signal Process., Biomed. Eng. Informat. (CISP-BMEI)*, Oct. 2019, pp. 1–5.

[17] J. Chen, J. Meng, X. Wang, and J. Yuan, "Dynamic graph CNN for event-camera based gesture recognition," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Oct. 2020, pp. 1–5.

[18] S. Jiang, L. Li, H. Xu, J. Xu, G. Gu, and P. B. Shull, "Stretchable e-skin patch for gesture recognition on the back of the hand," *IEEE Trans. Ind. Electron.*, vol. 67, no. 1, pp. 647–657, Jan. 2020.

[19] A. Rahman, M. Whitlock, and E. Abdelfattah, "Effectiveness of classifiers to identify hand gestures with motion capture coordinate markers," in *Proc. IEEE 11th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2021, pp. 779–784.

[20] Š. Obdržálek, G. Kurillo, F. Ofli, R. Bajcsy, E. Seto, H. Jimison, and M. Pavel, "Accuracy and robustness of Kinect pose estimation in the context of coaching of elderly population," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 1188–1193.

[21] M. Chen, G. AlRegib, and B.-H. Juang, "Feature processing and modeling for 6D motion gesture recognition," *IEEE Trans. Multimedia*, vol. 15, no. 3, pp. 561–571, Apr. 2013.

[22] D. Pavllo, T. Porssut, B. Herbelin, and R. Boulic, "Real-time marker-based finger tracking with neural networks," in *Proc. IEEE Conf. Virtual Reality 3D User Interface (VR)*, Mar. 2018, pp. 651–652.

[23] M. Andrychowicz, "Learning dexterous in-hand manipulation," *Int. J. Robot. Res.*, vol. 39, no. 1, pp. 3–20, 2020.

[24] A. Fern'ndez-Baena, A. Susín, and X. Lligadas, "Biomechanical validation of upper-body and lower-body joint movements of Kinect motion capture data for rehabilitation treatments," in *Proc. 4th Int. Conf. Intell. Netw. Collaborative Syst.*, Sep. 2012, pp. 656–661.

[25] R. Y. Wang and J. Popović, "Real-time hand-tracking with a color glove," *ACM Trans. Graph.*, vol. 28, no. 3, pp. 1–8, Jul. 2009.

[26] G. Maino and G. L. Foresti, *Image Analysis and Processing–ICIAP 2011: 16th International Conference, Ravenna, Italy, September 14-16, 2011, Proceedings, Part I*. Cham, Switzerland: Springer, 2011.

[27] A. Bellarbi, S. Benbelkacem, N. Zenati-Henda, and M. Belhocine, "Hand gesture interaction using color-based method for tabletop interfaces," in *Proc. IEEE 7th Int. Symp. Intell. Signal Process.*, Sep. 2011, pp. 1–6.

[28] P. Dutta, B. Aparna, B. Sridharan, and V. Tyagi, "Efficient marker based gesture recognition technique," *Int. J. Comput. Appl.*, vol. 81, no. 1, pp. 11–15, Nov. 2013.

[29] S. Amendola, L. Bianchi, and G. Marrocco, "Movement detection of human body segments: Passive radio-frequency identification and machine-learning technologies," *IEEE Antennas Propag. Mag.*, vol. 57, no. 3, pp. 23–37, Jun. 2015.

[30] H. Ishiyama and S. Kurabayashi, "Monochrome glove: A robust real-time hand gesture recognition method by using a fabric glove with design of structured markers," in *Proc. IEEE Virtual Reality (VR)*, Mar. 2016, pp. 187–188.

[31] P. Roy, S. Ghosh, and U. Pal, "A CNN based framework for unistroke numeral recognition in air-writing," in *Proc. 16th Int. Conf. Frontiers Handwriting Recognit. (ICFHR)*, Aug. 2018, pp. 404–409.

[32] M. Nakamura and K. Hoshino, "Tankendo motion estimation system with robustness against differences in color and size between users' clothes using 4-Color markers with elastic belts," in *Proc. 5th Int. Conf. Biomed. Bioinf. Eng.*, Nov. 2018, pp. 110–117.

[33] V. L. Hanson, J. P. Brezin, S. Crayne, S. Keates, R. Kjeldsen, J. T. Richards, C. Swart, and S. Trewin, "Improving web accessibility through an enhanced open-source browser," *IBM Syst. J.*, vol. 44, no. 3, pp. 573–588, 2005.

[34] J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in time-sequential images using hidden Markov model," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1992, pp. 379–385.

[35] M. Sigalas, H. Baltzakis, and P. Trahanias, "Gesture recognition based on arm tracking for human-robot interaction," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2010, pp. 5424–5429.

[36] G. R. S. Murthy and R. S. Jadon, "Hand gesture recognition using neural networks," in *Proc. IEEE 2nd Int. Advance Comput. Conf. (IACC)*, Feb. 2010, pp. 134–138.

[37] J. L. Raheja, R. Shyam, U. Kumar, and P. B. Prasad, "Real-time robotic hand control using hand gestures," in *Proc. 2nd Int. Conf. Mach. Learn. Comput.*, Feb. 2010, pp. 12–16.

[38] Z. Cao, X. Xu, B. Hu, M. Zhou, and Q. Li, "Real-time gesture recognition based on feature recalibration network with multi-scale information," *Neurocomputing*, vol. 347, pp. 119–130, Jun. 2019.

[39] J. Nagi, F. Ducatelle, G. A. Di Caro, D. Ciresan, U. Meier, A. Giusti, F. Nagi, J. Schmidhuber, and L. M. Gambardella, "Max-pooling convolutional neural networks for vision-based hand gesture recognition," in *Proc. IEEE Int. Conf. Signal Image Process. Appl. (ICSIPA)*, Nov. 2011, pp. 342–347.

[40] I. B. Çelik and M. Kuntalp, "Development of a robotic-arm controller by using hand gesture recognition," in *Proc. Int. Symp. Innov. Intell. Syst. Appl.*, Jul. 2012, pp. 1–5.

[41] P.-C. Cho, C.-T. Li, and W.-H. Chen, "Implementation of low-cost vision-based gesture recognition systems based on FPGA approach," in *Proc. Int. Symp. Comput., Consum. Control*, Jun. 2012, pp. 329–332.

[42] S. Jeong, J. Jin, T. Song, K. Kwon, and J. W. Jeon, "Single-camera dedicated television control system using gesture drawing," *IEEE Trans. Consum. Electron.*, vol. 58, no. 4, pp. 1129–1137, Nov. 2012.

[43] P. Barros, G. I. Parisi, D. Jirak, and S. Wermter, "Real-time gesture recognition using a humanoid robot with a deep neural architecture," in *Proc. IEEE-RAS Int. Conf. Humanoid Robots*, Nov. 2014, pp. 646–651.

[44] C. Li, H. Ma, C. Yang, and M. Fu, "Teleoperation of a virtual iCub robot under framework of parallel system via hand gesture recognition," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Jul. 2014, pp. 1469–1474.

[45] V. Bhame, R. Sreemathy, and H. Dhumal, "Vision based hand gesture recognition using eccentric approach for human computer interaction," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2014, pp. 949–953.

[46] N. S. Simul, N. M. Ara, and M. S. Islam, "A support vector machine approach for real time vision based human robot interaction," in *Proc. 19th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2016, pp. 496–500.

[47] K. Lamb and S. Madhe, "Automatic bed position control based on hand gesture recognition for disabled patients," in *Proc. Int. Conf. Autom. Control Dyn. Optim. Techn. (ICACDOT)*, Sep. 2016, pp. 148–153.

[48] A. Mohanty, S. S. Rambhatla, and R. R. Sahay, "Deep gesture: Static hand gesture recognition using CNN," in *Proc. Int. Conf. Comput. Vis. Image Process. (CVIP)*, vol. 2. Cham, Switzerland: Springer, 2017, pp. 449–461.

[49] D. K. Ghosh and S. Ari, "On an algorithm for vision-based hand gesture recognition," *Signal, Image Video Process.*, vol. 10, no. 4, pp. 655–662, Apr. 2016.

[50] O. K. Oyedotun and A. Khashman, "Deep learning in vision-based static hand gesture recognition," *Neural Comput. Appl.*, vol. 28, no. 12, pp. 3941–3951, Dec. 2017.

[51] P. Wang, H. Liu, L. Wang, and R. X. Gao, "Deep learning-based human motion recognition for predictive context-aware human-robot collaboration," *CIRP Ann.*, vol. 67, no. 1, pp. 17–20, 2018.

[52] J. Singha, A. Roy, and R. H. Laskar, "Dynamic hand gesture recognition using vision-based approach for human-computer interaction," *Neural Comput. Appl.*, vol. 29, no. 4, pp. 1129–1141, 2018.

[53] S. Song, D. Yan, and Y. Xie, "Design of control system based on hand gesture recognition," in *Proc. IEEE 15th Int. Conf. Netw., Sens. Control (ICNSC)*, Mar. 2018, pp. 1–4.

[54] J.-H. Sun, T.-T. Ji, S.-B. Zhang, J.-K. Yang, and G.-R. Ji, "Research on the hand gesture recognition based on deep learning," in *Proc. 12th Int. Symp. Antennas, Propag. EM Theory (ISAPE)*, Dec. 2018, pp. 1–4.

[55] M. J. Islam, M. Ho, and J. Sattar, "Understanding human motion and gestures for underwater human–robot collaboration," *J. Field Robot.*, vol. 36, no. 5, pp. 851–873, Aug. 2019.

[56] J. Chang, J. Xiao, J. Chai, and Z. Zhou, "An improved faster R-CNN algorithm for gesture recognition in human-robot interaction," in *Proc. Chin. Autom. Congr. (CAC)*, Nov. 2019, pp. 5761–5764.

[57] W. Fang, Y. Ding, F. Zhang, and J. Sheng, "Gesture recognition based on CNN and DCGAN for calculation and text output," *IEEE Access*, vol. 7, pp. 28230–28237, 2019.

[58] J. P. Sahoo, S. Ari, and S. K. Patra, "Hand gesture recognition using PCA based deep CNN reduced features and SVM classifier," in *Proc. IEEE Int. Symp. Smart Electron. Syst. (iSES) (Formerly iNiS)*, Dec. 2019, pp. 221–224.

[59] Y. Jiang, X. Peng, M. Xue, C. Wang, and H. Qi, "An underwater human–robot interaction using hand gestures for fuzzy control," *Int. J. Fuzzy Syst.*, vol. 23, no. 6, pp. 1879–1889, Sep. 2021.

[60] M. Al-Hammadi, G. Muhammad, W. Abdul, M. Alsulaiman, M. A. Bencherif, T. S. Alrayes, H. Mathkour, and M. A. Mekhtiche, "Deep learning-based approach for sign language gesture recognition with efficient hand gesture representation," *IEEE Access*, vol. 8, pp. 192527–192542, 2020.

[61] M. Al-Hammadi, G. Muhammad, W. Abdul, M. Alsulaiman, M. A. Bencherif, and M. A. Mekhtiche, "Hand gesture recognition for sign language using 3DCNN," *IEEE Access*, vol. 8, pp. 79491–79509, 2020.

[62] P. Choudhary and S. N. Tazi, "An adaptive system of Yogic gesture recognition for human computer interaction," in *Proc. IEEE 15th Int. Conf. Ind. Inf. Syst. (ICIIS)*, Nov. 2020, pp. 399–402.

[63] E. Ohn-Bar and M. M. Trivedi, "Hand gesture recognition in real time for automotive interfaces: A multimodal vision-based approach and evaluations," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 6, pp. 2368–2377, Dec. 2014.

[64] E. Coupeté, F. Moutarde, and S. Manitsaris, "A user-adaptive gesture recognition system applied to human-robot collaboration in factories," in *Proc. 3rd Int. Symp. Movement Comput.*, Jul. 2016, pp. 1–7.

[65] Y. Yu, X. Wang, Z. Zhong, and Y. Zhang, "ROS-based UAV control using hand gesture recognition," in *Proc. 29th Chin. Control Decis. Conf. (CCDC)*, May 2017, pp. 6795–6799.

[66] Z. Ren, J. Meng, and J. Yuan, "Depth camera based hand gesture recognition and its applications in human-computer-interaction," in *Proc. 8th Int. Conf. Inf., Commun. Signal Process.*, Dec. 2011, pp. 1–5.

[67] Y. Gu, H. Do, Y. Ou, and W. Sheng, "Human gesture recognition through a Kinect sensor," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2012, pp. 1379–1384.

[68] D. Xu, Y.-L. Chen, C. Lin, X. Kong, and X. Wu, "Real-time dynamic gesture recognition system based on depth perception for robot navigation," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2012, pp. 689–694.

[69] K. Qian, J. Niu, and H. Yang, "Developing a gesture based remote human-robot interaction system using Kinect," *Int. J. Smart Home*, vol. 7, no. 4, pp. 203–208, 2013.
[70] H.-S. Yeo, B.-G. Lee, and H. Lim, "Hand tracking and gesture recognition system for human-computer interaction using low-cost hardware," *Multimedia Tools Appl.*, vol. 74, no. 8, pp. 2687–2715, Apr. 2015.
[71] P. K. Pisharady and M. Saerbeck, "A robust gesture detection and recognition algorithm for domestic robot interactions," in *Proc. 13th Int. Conf. Control Autom. Robot. Vis. (ICARCV)*, Dec. 2014, pp. 775–780.
[72] Y. Chen, B. Luo, Y.-L. Chen, G. Liang, and X. Wu, "A real-time dynamic hand gesture recognition system using Kinect sensor," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2015, pp. 2026–2030.
[73] E. Coupeté, F. Moutarde, and S. Manitsaris, "Gesture recognition using a depth camera for human robot collaboration on assembly line," *Proc. Manuf.*, vol. 3, pp. 518–525, Jan. 2015.
[74] T. Q. Vinh and N. T. Tri, "Hand gesture recognition based on depth image using Kinect sensor," in *Proc. 2nd Nat. Found. Sci. Technol. Develop. Conf. Inf. Comput. Sci. (NICS)*, Sep. 2015, pp. 34–39.
[75] E. T. Alasady, F. Khalid, P. S. Sulaiman, and R. Yaakob, "Hand gesture trajectory estimation using keypoints combination of brisk and minimum eigenvalue techniques for human computer interaction applications," in *Proc. 1st EAI Int. Conf. Comput. Sci. Eng.*, 2017, pp. 204–217.
[76] G. Li, H. Wu, G. Jiang, S. Xu, and H. Liu, "Dynamic gesture recognition in the Internet of Things," *IEEE Access*, vol. 7, pp. 23713–23724, 2019.
[77] D. Jiang, G. Li, Y. Sun, J. Kong, and B. Tao, "Gesture recognition based on skeletonization algorithm and CNN with ASL database," *Multimedia Tools Appl.*, vol. 78, no. 21, pp. 29953–29970, Nov. 2019.
[78] J. Dong, Z. Xia, W. Yan, and Q. Zhao, "Dynamic gesture recognition by directional pulse coupled neural networks for human-robot interaction in real time," *J. Vis. Commun. Image Represent.*, vol. 63, Aug. 2019, Art. no. 102583.
[79] W. Cheng, Y. Sun, G. Li, G. Jiang, and H. Liu, "Jointly network: A network based on CNN and RBM for gesture recognition," *Neural Comput. Appl.*, vol. 31, no. S1, pp. 309–323, Jan. 2019.
[80] Y. Sun, A. Armengol-Urpi, S. N. Reddy Kantareddy, J. Siegel, and S. Sarma, "MagicHand: Interact with IoT devices in augmented reality environment," in *Proc. IEEE Conf. Virtual Reality 3D User Interface (VR)*, Mar. 2019, pp. 1738–1743.
[81] Y. Kuang, H. Cheng, Y. Zheng, F. Cui, and R. Huang, "One-shot gesture recognition with attention-based DTW for human-robot collaboration," *Assem. Autom.*, vol. 40, no. 1, pp. 40–47, Aug. 2019.
[82] E. J. Escobedo Cardenas and G. C. Chavez, "Multimodal hand gesture recognition combining temporal and pose information based on CNN descriptors and histogram of cumulative magnitudes," *J. Vis. Commun. Image Represent.*, vol. 71, Aug. 2020, Art. no. 102772.
[83] E. Y. Ahn, J. H. Lee, T. Mullen, and J. Yen, "Dynamic vision sensor camera based bare hand gesture recognition," in *Proc. IEEE Symp. Comput. Intell. Multimedia, Signal Vis. Process.*, Apr. 2011, pp. 52–59.
[84] S. Oprisescu, C. Rasche, and B. Su, "Automatic static hand gesture recognition using ToF cameras," in *Proc. 20th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2012, pp. 2748–2751.
[85] B. Kohn, A. N. Belbachir, and A. Nowakowska, "Real-time gesture recognition using bio inspired 3D vision sensor," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2012, pp. 37–42.
[86] M. Kulkarni, J. Butala, and V. Udpikar, "Static gesture recognition using PMD ToF camera," in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Sep. 2014, pp. 2712–2718.
[87] T. Kopinski, D. Malysiak, A. Gepperth, and U. Handmann, "Time-of-flight based multi-sensor fusion strategies for hand gesture recognition," in *Proc. IEEE 15th Int. Symp. Comput. Intell. Informat. (CINTI)*, Nov. 2014, pp. 243–248.
[88] F. Sachara, T. Kopinski, A. Gepperth, and U. Handmann, "Free-hand gesture recognition with 3D-CNNs for in-car infotainment control in real-time," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 959–964.
[89] H. Chai, M. Fei, A. Yang, and L. Chen, "3D gesture recognition method based on faster R-CNN network," in *Proc. Austral. New Zealand Control Conf. (ANZCC)*, Dec. 2018, pp. 291–296.
[90] M. Van den Bergh and L. Van Gool, "Combining RGB and ToF cameras for real-time 3D hand gesture interaction," in *Proc. IEEE Workshop Appl. Comput. Vis. (WACV)*, Jan. 2011, pp. 66–72.
[91] P. Molchanov, S. Gupta, K. Kim, and K. Pulli, "Multi-sensor system for driver's hand-gesture recognition," in *Proc. 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG)*, vol. 1, May 2015, pp. 1–8.
[92] H. Liu, T. Fang, T. Zhou, and L. Wang, "Towards robust human-robot collaborative manufacturing: Multimodal fusion," *IEEE Access*, vol. 6, pp. 74762–74771, 2018.
[93] J. D. Foley, "Interfaces for advanced computing," *Sci. Amer.*, vol. 257, no. 4, pp. 126–135, Oct. 1987.
[94] S. Mann, "Humanistic computing: 'WearComp' as a new framework and application for intelligent signal processing," *Proc. IEEE*, vol. 86, no. 11, pp. 2123–2151, Nov. 1998.
[95] S. S. Fels and G. E. Hinton, "Glove-TalkII—A neural-network interface which maps gestures to parallel formant speech synthesizer controls," *IEEE Trans. Neural Netw.*, vol. 9, no. 1, pp. 205–212, Jan. 1998.
[96] B. Hartmann and N. Link, "Gesture recognition with inertial sensors and optimized DTW prototypes," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Oct. 2010, pp. 2102–2109.
[97] K. Liu, C. Chen, R. Jafari, and N. Kehtarnavaz, "Multi-HMM classification for hand gesture recognition using two differing modality sensors," in *Proc. IEEE Dallas Circuits Syst. Conf. (DCAS)*, Oct. 2014, pp. 1–4.
[98] K. Liu, C. Chen, R. Jafari, and N. Kehtarnavaz, "Fusion of inertial and depth sensor data for robust hand gesture recognition," *IEEE Sensors J.*, vol. 14, no. 6, pp. 1898–1903, Jun. 2014.
[99] C. Chen, R. Jafari, and N. Kehtarnavaz, "A real-time human action recognition system using depth and inertial sensor fusion," *IEEE Sensors J.*, vol. 16, no. 3, pp. 773–781, Feb. 2016.
[100] S. Shin and W. Sung, "Dynamic hand gesture recognition for wearable devices with low complexity recurrent neural networks," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2016, pp. 2274–2277.
[101] I. Torres, J. Bustamante, and D. A. Sierra, *VII Latin American Congress on Biomedical Engineering CLAIB 2016, Bucaramanga, Santander, Colombia, October 26th–28th, 2016*. Cham, Switzerland: Springer, 2017.
[102] I. Stančić, J. Musić, and T. Grujić, "Gesture recognition system for real-time mobile robot control based on inertial sensors and motion strings," *Eng. Appl. Artif. Intell.*, vol. 66, pp. 33–48, Nov. 2017.
[103] V. Joukov, J. Cesic, K. Westermann, I. Markovic, D. Kulic, and I. Petrovic, "Human motion estimation on lie groups using IMU measurements," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 1965–1972.
[104] J. Rwigema, H.-R. Choi, and T. Kim, "A differential evolution approach to optimize weights of dynamic time warping for multi-sensor based gesture recognition," *Sensors*, vol. 19, no. 5, p. 1007, Feb. 2019.
[105] P. Neto, M. Simão, N. Mendes, and M. Safeea, "Gesture-based human-robot interaction for human assistance in manufacturing," *Int. J. Adv. Manuf. Technol.*, vol. 101, nos. 1–4, pp. 119–135, Mar. 2019.
[106] M. Kim, J. Cho, S. Lee, and Y. Jung, "IMU sensor-based hand gesture recognition for human-machine interfaces," *Sensors*, vol. 19, no. 18, p. 3827, Sep. 2019.
[107] K. Suri and R. Gupta, "Convolutional neural network array for sign language recognition using wearable IMUs," in *Proc. 6th Int. Conf. Signal Process. Integr. Netw. (SPIN)*, Mar. 2019, pp. 483–488.
[108] N. Diliberti, C. Peng, C. Kaufman, Y. Dong, and J. T. Hansberger, "Real-time gesture recognition using 3D sensory data and a light convolutional neural network," in *Proc. 27th ACM Int. Conf. Multimedia*, Oct. 2019, pp. 401–410.
[109] C. Amma, T. Krings, J. Böer, and T. Schultz, "Advancing muscle-computer interfaces with high-density electromyography," in *Proc. 33rd Annu. ACM Conf. Human Factors Comput. Syst.*, Apr. 2015, pp. 929–938.
[110] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, and J. Li, "Gesture recognition by instantaneous surface EMG images," *Sci. Rep.*, vol. 6, no. 1, Nov. 2016, Art. no. 36571.
[111] X. Liu, J. Sacks, M. Zhang, A. G. Richardson, T. H. Lucas, and J. Van der Spiegel, "The virtual trackpad: An electromyography-based, wireless, real-time, low-power, embedded hand-gesture-recognition system using an event-driven artificial neural network," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 64, no. 11, pp. 1257–1261, Nov. 2017.
[112] K.-Y. Lian, C.-C. Chiu, Y.-J. Hong, and W.-T. Sung, "Wearable armband for real time hand gesture recognition," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2017, pp. 2992–2995.
[113] S. Pancholi and A. M. Joshi, "Electromyography-based hand gesture recognition system for upper limb amputees," *IEEE Sensors Lett.*, vol. 3, no. 3, pp. 1–4, Mar. 2019.

[114] S. Benatti, F. Montagna, V. Kartsch, A. Rahimi, D. Rossi, and L. Benini, "Online learning and classification of EMG-based gestures on a parallel ultra-low power platform using hyperdimensional computing," *IEEE Trans. Biomed. Circuits Syst.*, vol. 13, no. 3, pp. 516–528, Jun. 2019.

[115] W. Wei, Q. Dai, Y. Wong, Y. Hu, M. Kankanhalli, and W. Geng, "Surface-electromyography-based gesture recognition by multi-view deep learning," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 10, pp. 2964–2973, Oct. 2019.

[116] L. Chen, J. Fu, Y. Wu, H. Li, and B. Zheng, "Hand gesture recognition using compact CNN via surface electromyography signals," *Sensors*, vol. 20, no. 3, p. 672, Jan. 2020.

[117] A. Moin, A. Zhou, A. Rahimi, A. Menon, S. Benatti, G. Alexandrov, S. Tamakloe, J. Ting, N. Yamamoto, Y. Khan, F. Burghardt, L. Benini, A. C. Arias, and J. M. Rabaey, "A wearable biosensing system with in-sensor adaptive machine learning for hand gesture recognition," *Nature Electron.*, vol. 4, no. 1, pp. 54–63, Dec. 2020.

[118] G. Rosati, G. Cisotto, D. Sili, L. Compagnucci, C. De Giorgi, E. F. Pavone, A. Paccagnella, and V. Betti, "Inkjet-printed fully customizable and low-cost electrodes matrix for gesture recognition," *Sci. Rep.*, vol. 11, no. 1, pp. 1–12, Jul. 2021.

[119] A. Ferrone, X. Jiang, L. Maiolo, A. Pecora, L. Colace, and C. Menon, "A fabric-based wearable band for hand gesture recognition based on filament strain sensors: A preliminary investigation," in *Proc. IEEE Healthcare Innov. Point Care Technol. Conf. (HI-POCT)*, Nov. 2016, pp. 113–116.

[120] D. Lu, Y. Yu, and H. Liu, "Gesture recognition using data glove: An extreme learning machine method," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2016, pp. 1349–1354.

[121] H. Liu and L. Wang, "Human motion prediction for human-robot collaboration," *J. Manuf. Syst.*, vol. 44, pp. 287–294, Jul. 2017.

[122] B. Fang, Q. Lv, J. Shan, F. Sun, H. Liu, D. Guo, and Y. Zhao, "Dynamic gesture recognition using inertial sensors-based data gloves," in *Proc. IEEE 4th Int. Conf. Adv. Robot. Mechatronics (ICARM)*, Jul. 2019, pp. 390–395.

[123] M. Jain, M. Sehgal, Y. Kalra, P. Jain, N. Aggarwal, R. Singh, D. Kumar, T. K. Bera, and R. Singh, "Object detection and gesture control of four-wheel mobile robot," in *Proc. Int. Conf. Commun. Electron. Syst. (ICCES)*, Jul. 2019, pp. 303–308.

[124] X. Wang, Y. Gu, Y. Lao, and B. Fang, "Haptic glove for object recognition and hand motion detection," in *Proc. Chin. Autom. Congr. (CAC)*, Nov. 2020, pp. 7192–7196.

[125] J. A. Fishel, T. Oliver, M. Eichermueller, G. Barbieri, E. Fowler, T. Hartikainen, L. Moss, and R. Walker, "Tactile telerobots for dull, dirty, dangerous, and inaccessible tasks," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2020, pp. 11305–11310.

[126] A. Boyali, N. Hashimoto, and O. Matsumoto, "Hand posture and gesture recognition using MYO armband and spectral collaborative representation based classification," in *Proc. IEEE 4th Global Conf. Consum. Electron. (GCCE)*, Oct. 2015, pp. 200–201.

[127] J. G. Abreu, J. M. Teixeira, L. S. Figueiredo, and V. Teichrieb, "Evaluating sign language recognition using the Myo armband," in *Proc. 13th Symp. Virtual Augmented Reality (SVR)*, Jun. 2016, pp. 64–70.

[128] G. Pomboza-Junez and J. H. Terriza, "Hand gesture recognition based on sEMG signals using support vector machines," in *Proc. IEEE 6th Int. Conf. Consum. Electron. Berlin (ICCE-Berlin)*, Sep. 2016, pp. 174–178.

[129] A. D. Wibawa and S. Sumpeno, "Gesture recognition for Indonesian sign language systems (ISLS) using multimodal sensor leap motion and Myo armband controllers based-on Naïve Bayes classifier," in *Proc. Int. Conf. Soft Comput., Intell. Syst. Inf. Technol. (ICSIIT)*, Sep. 2017, pp. 1–6.

[130] S. Shin, Y. Baek, J. Lee, Y. Eun, and S. H. Son, "Korean sign language recognition using EMG and IMU sensors based on group-dependent NN models," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Nov. 2017, pp. 1–7.

[131] M. E. Benalcázar, C. Motoche, J. A. Zea, A. G. Jaramillo, C. E. Anchundia, P. Zambrano, M. Segura, F. Benalcázar Palacios, and M. Pérez, "Real-time hand gesture recognition using the Myo armband and muscle activity detection," in *Proc. IEEE 2nd Ecuador Tech. Chapters Meeting (ETCM)*, Oct. 2017, pp. 1–6.

[132] K. S. Krishnan, A. Saha, S. Ramachandran, and S. Kumar, "Recognition of human arm gestures using myo armband for the game of hand cricket," in *Proc. IEEE Int. Symp. Robot. Intell. Sensors (IRIS)*, Oct. 2017, pp. 389–394.

[133] M. Tavakoli, C. Benussi, P. A. Lopes, L. B. Osorio, and A. T. de Almeida, "Robust hand gesture recognition with a double channel surface EMG wearable armband and SVM classifier," *Biomed. Signal Process. Control*, vol. 46, pp. 121–130, Sep. 2018.

[134] N. Nasri, S. Orts-Escolano, F. Gomez-Donoso, and M. Cazorla, "Inferring static hand poses from a low-cost non-intrusive sEMG sensor," *Sensors*, vol. 19, no. 2, p. 371, Jan. 2019.

[135] M. E. Benalcázar, A. G. Jaramillo, Jonathan, A. Zea, A. Páez, and V. H. Andaluz, "Hand gesture recognition using machine learning and the Myo armband," in *Proc. 25th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2017, pp. 1040–1044.


**DONGHYEON NOH** received the B.S. degree from the School of Mechanical Engineering, Soongsil University, Seoul, South Korea, in 2021, where he is currently pursuing the master's degree. His current research interests include human motion recognition and human–robot collaboration.

**HOJIN YOON** received the B.S. degree in mechanical engineering from Soongsil University, South Korea, in 2022, where he is currently pursuing the M.S. degree in mechanical engineering. His research interests include deep learning, redundant mechanism, and mobile manipulation robot.

**DONGHUN LEE** received the B.S. degree in mechanical engineering from Soongsil University, South Korea, in 2004, and the Ph.D. degree in mechanical engineering from Seoul National University, South Korea, in 2009. He is currently an Associate Professor with the School of Mechanical Engineering, Soongsil University. His research interests include mobile manipulation, human–robot interactions, deep learning, and redundancy control.