International Journal of Control, Automation, and Systems 22(6) (2024) 1761-1778
http://dx.doi.org/10.1007/s12555-024-0298-1
ISSN:1598-6446 eISSN:2005-4092
http://www.springer.com/12555

Hand Tracking: Survey

Jinuk Heo, Hyelim Choi, Yongseok Lee, Hyunsu Kim, Harim Ji, Hyunreal Park, Youngseon Lee, Cheongkee Jung, Hai-Nguyen Nguyen, and Dongjun Lee*

Abstract: Hand tracking is relevant to such a variety of applications including human-robot interaction (HRI), human-computer interaction (HCI), virtual reality (VR), and augmented reality (AR). Accurate and robust hand tracking however is challenging due to the intricacies of dynamic motion within small space and the complex interactions with nearby objects, coupled with the hurdles in real-time hand mesh reconstruction. In this paper, we conduct a comprehensive examination and analysis of existing hand tracking technologies. Through the review of major works in the literature, we have discovered numerous studies employing a diverse array of sensors, leading us to propose their categorization into seven types: vision, soft wearable, encoder, magnetic, inertial measurement unit (IMU), electromyography (EMG), and the fusion of sensor modalities. Our findings indicate that no singular solution surpasses all others, attributing to the inherent limitations of using a single sensor modality. As a result, we assert that integrating multiple sensor modalities presents a viable path toward devising a superior hand tracking solution. Ultimately, this survey paper aims to bolster interdisciplinary research efforts across the spectrum of hand tracking technologies, thereby contributing to the advancement of the field.

Keywords: Augmented reality, computer vision, data gloves, exoskeleton gloves, hand tracking, human-computer interaction, human-robot interaction, mixed reality, virtual reality, wearable devices.

1. INTRODUCTION

Real-time three-dimensional (3D) hand pose estimation, or hand tracking, has emerged as a significant area of interest over the last decade, particularly in fields such as human-robot interaction (HRI), human-computer interaction (HCI), virtual reality (VR), and augmented reality (AR), due to its capacity to enhance user immersion and presence in various applications. To date, a multitude of hand tracking technologies has been developed, some of which have achieved commercial success and widespread adoption, yet the necessity of further advancements remains vast. In response, this paper aims to comprehensively study and analyze existing hand tracking technologies, with a particular focus on identifying promising directions for the development of hand tracking solutions in alignment with the evolution of related fields.

Hand motion tracking poses significant challenges. Firstly, compared to other body parts, the hand allows for dynamic, sophisticated, and dexterous motion within a small space, leading to frequent self-occlusion. Secondly, as a fundamental function of human manipulation, hands are required to frequently interact with different objects, environments, and other humans. Therefore, hand tracking must be robust against various interferences. The relatively small size of the hand makes it difficult for sensor attachment, and wearability becomes crucial if the technology requires wearing devices. Although hand tracking has advanced significantly in recent years, no existing technology has yet emerged as a satisfactory solution that resolves all these challenges and offers a universal and reliable platform for reasonably large pool of applications.

Meanwhile, the advancement in hand tracking hold great potential to transform many application domains. First, it can significantly enhance immersion and presence in VR and AR environments [6,7], surpassing traditional handheld controllers or gamepads by enabling more natural and sophisticated hand movements. Faster and more accurate hand tracking would allow for the implementation of more dexterous and agile manipulations within VR/AR contexts. Second, hand tracking technology is poised to significantly improve robot teleoperation, as manifest in the Avatar XPRIZE competition held in 2022 [8]. Enhanced hand tracking technologies are cru-

Manuscript received April 17, 2024; accepted April 22, 2024. Recommended by Editor-in-Chief Hyo-Sung Ahn. This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MSIT) (RS-2023-00208052).

Jinuk Heo, Hyelim Choi, Hyunsu Kim, Harim Ji, Hyunreal Park, Youngseon Lee, Cheongkee Jung, Hai-Nguyen Nguyen, and Dongjun Lee are with the Department of Mechanical Engineering, IAMD and IOER, Seoul National University, Seoul 08826, Korea (e-mails: jinukheo@snu.ac.kr, helmchoi@snu.ac.kr, hyunsu.kim@snu.ac.kr, jiharim0911@snu.ac.kr, hyun3556@snu.ac.kr, yslee1765@snu.ac.kr, blueflag147@snu.ac.kr, hanntonkin@snu.ac.kr, djlee@snu.ac.kr). Yongseok Lee is now with the Samsung Research, Seoul 06765, Korea, and he was with the Department of Mechanical Engineering, IAMD and IOER, Seoul National University, Seoul 08826, Korea (e-mail: yongseoklee@snu.ac.kr).
* Corresponding author.

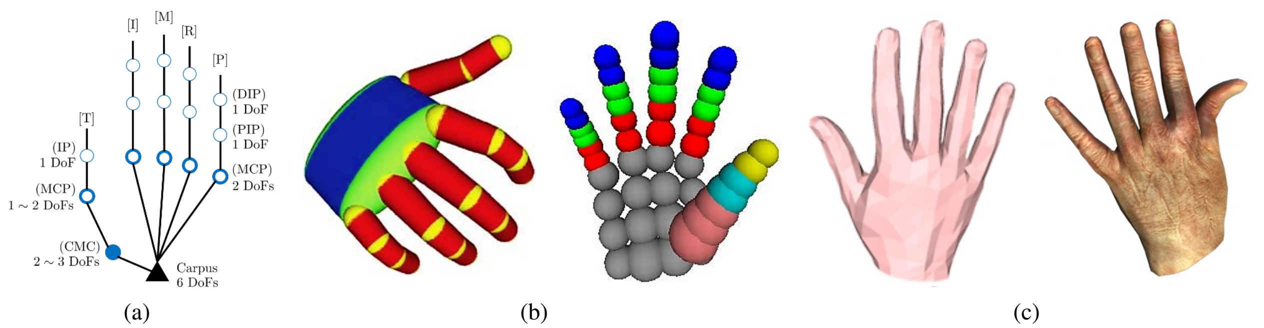


Fig. 1. The examples of hand model: (a) Segment-joint model [1], (b) geometric primitive models [2,3], and (c) soft hand models including MANO [4] and Handy [5].

cial for intuitive control and essential for the accurate and skilled teleoperation of humanoid robots. With the increasing release of commercial humanoid robots such as Tesla's Optimus [9] and Apptronik's Apollo [10], the demand for sophisticated hand tracking solutions is expected to rise significantly. Third, hand tracking can be leveraged to capture high-quality data from intricate tasks performed by human experts. This approach is increasingly popular in robot learning [11-16], where the quality of data for imitation learning is paramount, especially as the complexity of tasks escalates.

Given the growing need for hand tracking across various domains, it is being investigated in numerous fields through diverse approaches. After reviewing recent findings, we propose classifying them into seven categories based on the sensing modality: vision, soft wearable, encoder, magnetic, inertial measurement unit (IMU), electromyography (EMG), and the fusion of sensor modalities. Each sensing approach presents inherent challenges; thus, we recognize the integration of diverse technologies from various disciplines, including computer vision, soft wearable sensors, mechatronics, and mobile robotics, as a promising direction for advancing hand tracking. Notably, recent studies have begun to integrate two different sensor modalities—visual and inertial—which were primarily utilized in computer vision and mobile robotics [17-20].

To date, there has been a lack of comprehensive analysis of hand tracking technologies, despite their not being a new field. Previous survey and review papers have often focused solely on a specific type of method (e.g., vision-based [21,22], fiber bragg grating (FBG) sensor-based [23], and visual-inertial fusion-based [24]), or narrow fields of hand tracking technologies [25]. This paper aims to holistically survey and categorize hand tracking technologies and suggest the promising direction to accelerate interdisciplinary studies across various research fields related to hand tracking.

In Section 2, we provide some preliminary backgrounds for hand tracking, including hand modeling and available sensors. Then, in Section 3, hand tracking methodologies are introduced, along with a succinct overview of the research trajectory for each type of sensor. Next, some widely-available commercial hand tracking products are briefly introduced in Section 4. Finally, in Section 5, we conclude this paper with a discussion on the promising direction of hand tracking.

2. PRELIMINARIES

2.1. Hand modeling

Hand tracking is a real-time 3D pose estimation of articulated rigid bodies, thus, hand configuration and constraints should be established first. The human hand anatomy is illustrated in Fig. 2, where the palm comprises the carpus and metacarpus. All fingers, excluding the thumb, are made up of three articulated bones; proximal, middle, and distal phalanges. These are connected to the root by the metacarpophalangeal (MCP), proximal interphalangeal (PIP), and distal interphalangeal (DIP) joints, respectively. The thumb slightly differs in structure, consisting of articulation of the metacarpal, proximal, and distal phalanges, which are connected by the carpometacarpal (CMC), MCP, and interphalangeal (IP) joints. Based on the hand anatomy, several hand modeling have been suggested and they can be classified into three types according to the detail level: the segment-joint model, the geometric primitive model [2,3], and the soft hand model [4,5].

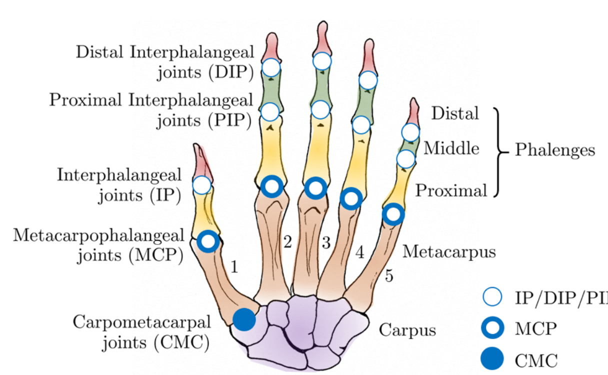


Fig. 2. The anatomy of the human hand.

Table 1. Strengths and weaknesses of each sensor type for hand tracking. The evaluation items consists of interferences (occlusion, magnetic interference, and contact), reliability (accuracy, and repeatability), comfortability (calibration, and wearability), and cost. Sensor types consists of vision (RGB, and RGB-D), soft wearable (flex, resistive strain sensor (RSS), capacitive strain sensor (CSS), fiber Bragg grating (FBG), and piezoresistive fabric), encoder, magnetic sensor with permanent magnet (PM) or electromagnet (EM), inertial measurement unit (IMU) with compass, and electromyography (EMG).

Sensor Type		Interferences			Reliability		Comfortability		Cost
		Occl.	Mag.	Contact	Accuracy	Repeat.	Calib.	Wearability	
Vision	RGB	weak	robust	robust	high	high	free	free	low
	RGB-D								medium
Soft	Flex	robust	robust	weak	high	low	complex	low-stretch	low
	RSS							good	low
	CSS							good	high
	FBG							good	high
	Piezo.							good	medium
Encoder		robust	robust	robust	medium	medium	simple	high-drag	medium
Magnetic	PM	robust	weak	robust	high	high	medium	good	low
	EM							bulky	high
IMU w/ Comp.		robust	weak	impact-weak	pos. drift	high	simple	good	low
EMG		robust	robust	robust	low	low	complex	good	medium

2.1.1 Segment-joint model

This is the most widely used model for hand tracking, which is illustrated in Fig. 1(a). The position and rotation of the palm have 6 DOFs (degree-of-freedoms), and each finger except the thumb has 4 DOFs; 2 DOFs for the MCP joint and 1 DOF for each of the PIP and DIP joints. The interpretation of thumb kinematics varies slightly for each paper. The thumb can be considered to have 4 DOFs as the same as other fingers, or its CMC and MCP joint can be interpreted to have higher DOFs, resulting in 5 DOFs or 6 DOFs. Consequently, the total DOFs of the hand varies from 26 to 28, depending on the interpretation of the thumb.

2.1.2 Geometric primitive model

Another approach for hand modeling, the geometric primitive hand model, has the same kinematics as the segment-joint model, but it includes rigid volumetric shapes, whereas the segment-joint model does not. As an example of the geometric primitive model, the hand can be approximated as a composition of elliptic cylinders, ellipsoids, spheres, and cones, as illustrated in Fig. 1(b) [2]. Similarly, this geometric primitive hand model can be simplified to use only spheres [3].

2.1.3 Soft hand model

Soft hand models are designed to represent not just the skeletal structure but also skin mesh. The most widely used example of soft hand model is MANO [4]. It is overparameterized by 15 ball joints and global orientation, and its mesh is constructed via modified linear blend skinning (LBS). Unlike standard LBS approaches, it leverages both pose and shape to prevent excessive smoothing at the joints. This model has been used in numerous real-time hand reconstruction studies [26-32]. The SMPL-X [33] is a holistic human model which combined the MANO hand model with the SMPL body model [34] and the FLAME head model [35]. Another soft hand model, Handy [5], improved several weaknesses of the MANO using much more hand scan data. Handy can even reconstruct skin textures by leveraging a generative adversarial network (GAN) [36] to capture high-frequency details.

2.2. Sensors

As shown in Table 1, hand tracking has been developed utilizing various types of sensors, and each has its advantages and disadvantages. Vision-based hand tracking, which has been extensively researched recently due to advancements in deep learning, has not yet fully overcome its vulnerability to occlusion and changing light conditions. In addition, the limited field of view (FOV) is also one of the major issues for this type of hand tracking. On the other hand, soft wearable sensor-based hand tracking cannot discriminate between deformation by hand motion and deformation by contact. Additionally, bad repeatability due to hysteresis, the need for recalibration whenever a new user wears the device, and low durability remain as challenges to be addressed. Encoder-based hand tracking adopts exoskeleton glove, and it usually facilitates force haptic feedback at the same time. However, it leads to poor wearability due to the friction between exoskeleton link-

Table 2. Summary of hand tracking technologies using vision type sensors. Papers are listed in order of time. (Terms) LBS: linear-blend-skinning, PCK: percentage-of-correct-keypoints, AUC: area-under-the-curve, PA: procrustes-alignment, MPJPE /MPVPE: mean-per-joint/vertex-position-error.

Year	Ref.	Sensor Type	Parameters	Mesh	Accuracy
2014	[3]	Depth Creative Senz3D	26	X	wrist/fingertips err. 7.3-11.7 mm custom dataset
2014	[37]	multiple RGBs	26	X	24.1 mm Dexter1 [38]
2015	[39]	Depth Kinect V2	73	X LBS	Pose: 15.0 mm Dexter1 [38]
2015	[40]	Depth	26	X	fingertips err. 19.6 mm Dexter1 [38]
2015	[41]	RGB-D	26	X	phalanx end-pt. err. < 10 mm custom synthetic dataset
2017	[42]	RGB-D RealSense SR300	26	X	32.6 mm (under occlusion) EgoDexter [42]
2017	[43]	RGB	63	X	2D-PCK AUC = 0.817 RHD [43]
2018	[44]	Depth	-	X	8.42 mm NYU [45]
2018	[46]	RGB	63	X	3D-PCK AUC = 0.887, 0.990 RHD [43], STB [47]
2018	[48]	RGB	26	X	3D-PCK AUC = 0.965 Stereo dataset [49]
2019	[50]	Depth	123	X	12.3 mm (NYU [45])
2019	[51]	RGB	-	O custom	Pose: 8.03 mm, Mesh: 7.95 mm custom synthetic dataset
2019	[52]	RGB	-	O Adam [53]	3D-PCK AUC = 0.994, 0.84 STB [47], DO [54]
2019	[26]	RGB	61	O MANO [4]	3D-PCK AUC = 0.926, 0.65, 0.995 RHD [43], DO [54], SHD [47]
2019	[55]	stereo RGB	-	X	7.18 mm (STB [47])
2020	[56]	RGB	61	O MANO [4]	Pose: 30.4, Mesh: 10.6
2020	[57]	RGB	63	X	-
2020	[58]	RGB	63	X	MPJPE 16.02, EPE 7.95, 20.89 InterH. [58], STB [47], RHD [43]
2020	[59]	IR image	63 BigHand2.2M	X	8.3 mm (slow), 8.8 mm (fast) custom dataset
2020	[60]	4 fisheye B/W cams	26	X LBS	9.4-14.7 mm (custom dataset)
2020	[27]	4 thermal cams	9	O MANO [4]	Pose: 12.0 mm (free motion) 26.8 mm (obj.interaction)
2022	[28]	RGB	61	O MANO [4]	Pose: 12.78 mm InterHand2.6M [58]
2022	[29]	RGB	61	O MANO [4]	Pose: 8.79 mm, Mesh: 9.03 mm InterHand2.6M [58]
2022	[30]	RGB	61	O MANO [4]	Pose: 9.1 mm, Mesh: 8.8 mm HO3D [56]
2023	[61]	RGB	47	O	Pose: MPJPE 11.8 mm, 11.0 mm Mesh: MPVPE 11.9 mm, 10.9 mm FreiHAND [62], HO3D [56]

Table 3. Table 2. continued.

Year	Ref.	Sensor Type	Parameters	Mesh	Accuracy
2023	[63]	RGB	-	O AMVUR	Pose: MPJPE 8.3 mm Mesh: MPVPE 8.2 mm HO3Dv2 [56]
2023	[64]	RGB	61	O MANO [4]	Pose: MPJPE 8.0 9mm Mesh: MPVPE 8.29 mm InterHand2.6M [58]
2023	[65]	RGB	61	O SMPL-X	Mesh: PA-MPVPE 5.65 mm synthetic data
2023	[31]	RGB	61	O MANO [4]	Pose: PA-MPJPE 6.0 mm, 7.7 mm Mesh: PA-MPVPE 5.7 mm, 7.9 mm FreiHAND [62], HO3D [56]
2023	[32]	RGB	61	O MANO [4]	Pose: PA-J-PE 5.3 mm, 8.5 mm Mesh: PA-V-PE 5.2 mm, 8.6 mm Dex-YCB [66], HO3D-v2 [56]

Table 4. Summary of hand tracking technologies using soft wearable type sensors. Papers are listed in order of time. (Terms) FBG: fiber Bragg grating, RSS/CSS: resistive/capacitive strain sensor, IMU: inertial measurement unit, SPAW: soft polymer acoustic waveguides.

Year	Ref.	Sensor Type	Parameters	Mesh	Accuracy
2011	[67]	FBG	14	X	-
2014	[68]	Flex	14	X	-
2014	[69]	RSS ionic liquid	2 2 fingers	X	-
2015	[70]	RSS ionic liquid	11	X	-
2016	[71]	Flex	10	X	-
2016	[72]	Flex	10	X	6 deg
2016	[73]	RSS liquid metal	6 3 fingers	X	< 9 deg
2016	[74,75]	Piezoresistive fabric	19	X	< 5 deg
2017	[76]	RSS liquid metal	21	X	1.39 deg
2017	[77]	CSS textile & silicone	10	X	-
2018	[78]	Flex (& RGB, IMU)	11	X	-
2018	[79]	Piezoresistive fabric	2 2 fingers	X	-
2019	[80]	CSS (& RGB-D) textile & silicone	31	X	7.6 deg
2020	[81]	FBG	6	X	< 1 deg
2020	[82]	FBG	14	X	1.63 deg
2020	[83]	FBG	10	X	0.80 deg
2021	[84]	Piezoresistive fabric	-	X	-
2023	[85]	CSS fabric	10	X	7.2 deg
2023	[86]	SPAW strain silicone	10	X	< 2.00 deg

Table 5. Summary of hand tracking technologies using encoder, magnetic sensor, IMU, EMG, and the fusion of sensor modalities. Papers are listed in order of time for each type. (Terms) PM: permanent magnet, EM: electromagnet, EMG: electromyography.

Year	Ref.	Sensor Type	Parameters	Mesh	Accuracy
2014	[87]*	Encoder			
2016	[88]*	Encoder magnetic encoder	13 3 fingers	X	< 1 mm index fingertip only
2023	[89]	Encoder potentiometer	13 3 fingers	X	0.95-5.61 mm index fingertip only
2011	[90]*	Magnetic sensor & PM	-	X	-
2016	[91]*	Magnetic sensor & EM	- 3 fingers	X	1.33 mm single fingertip only
2019	[92]	Magnetic sensor & EM	5 (per unit)	X	4.41 mm, 4.65 deg
2021	[93]*	Magnetic sensor & PM exoskeleton	15	X	1.9-3.2 deg
2021	[94]*	Magnetic sensor & EM Polhemus [95]	19 4 fingers	X	4.7-5.2 mm joint positions
1999	[96]*	IMU 2-axis accelerometers	-	X	- pseudo static gesture
2009	[97]*	IMU 3-axis accelerometers	9 2 fingers	X	-
2015	[98]	IMU & Compass	-	X	5.95 deg
2017	[99]	IMU & Compass	23	X	< 3.32 deg
2018	[100]	IMU & Compass & CSS	16 3 fingers	X	0.068-1.567 deg
2019	[101]	IMU & Compass	18	X	< 3 deg
2021	[102]	EMG Myo	21	X	6.24 deg
2020	[17]	IMU & depth	21	X MSRA14 [3]	wrist/fingertips err. 9.34 mm
2020	[18]	Gyroscope & depth	26	X	2D pixel err. < 10 custom dataset
2021	[19]	IMU & stereo-RGB ZED mini	20 3 fingers	X	10.69 mm (free motion) 12.68 mm (obj.interaction)
2021	[103]	gyroscope & EMG Myo, gForce [104]	-	X	25.7 mm
2023	[20]	CSS & RGB-D & IMU	-	O	0.142-0.206 rad

* Only the distal phalanx is tracked. The whole finger pose is calculated via inverse kinematics.

ages causing fast fatigue to the user. Hand tracking based on magnetic sensors or IMUs with compasses is highly susceptible to magnetic interference, making hand tracking impossible near electronic devices. EMG sensor-based hand tracking still lacks in accuracy and reliability. Representative studies for each type of hand tracking are summarized at Tables 2-5.

Despite various attempts for hand tracking, unresolved issues remain in all methods. As will be introduced in Subsection 3.7, recent research actively investigates multi-sensor fusion to overcome the weaknesses of single sensor-based hand tracking.

3. HAND TRACKING METHODOLOGIES

We categorize hand tracking technologies into seven types depending on the sensing method: 1) vision, 2) soft wearable, 3) encoder, 4) magnetic, 5) IMU with compass, 6) EMG, and 7) fusion of sensor modalities.

3.1. Vision

Despite the recent popularity and considerable progress of vision-based hand tracking methods, challenges persist. Visual occlusion is one fundamental limitation of these approaches, as hand pose becomes not fully-observable when parts of the hand are obscured. Another limitation coming from the recent trend of adopting learning-based methods is that the dataset is expensive due to the need for extensive and diverse datasets (e.g., varied skin colors and lighting conditions, and accurately annotated hand poses).

Vision-based methods can be classified according to whether they use depth or color. Depth provides 3D information vital for 3D pose estimation but is limited by range, light interference, and environmental texture. Color, captured by conventional cameras, offers rich visual detail and some cues of relative depth even without 3D data. Although including depth can provide richer information, recent trend is exploiting RGB images without depth due to the easy accessibility of sensors and abundant data on the internet.

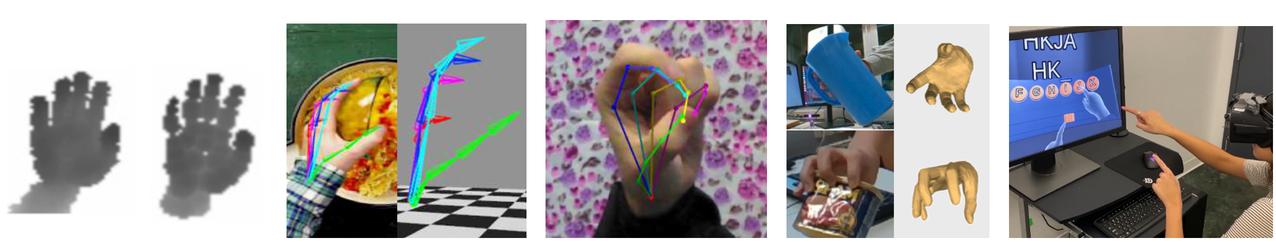
3.1.1 Depth

Given depth information, estimating hand pose was posed as a filtering/optimization problem in [3,39-41]. Makris *et al*. [41] introduced a geometric primitive model made up of cylinders and spheres and rendered the depth of the model to run a hierarchical particle filter. Optimization was conducted in [3,39,40] to minimize the cost related to the discrepancy of the depth measurement and the depth rendered from the hand model (i.e., spheres [3], mesh [39], and weighted 3D Gaussian mixtures [40]).

More recently, learning-based methods have been widely used to estimate hand poses from depth [42,44,50,59]. In [42], Mueller *et al*. trained convolutional neural networks (CNNs) in a supervised manner to predict 2D/3D joint positions and the hand pose was fit with an optimization process. Moon *et al*. [44] cast the depth measurement into a 3D voxel map and trained CNNs via supervised learning to predict 3D heatmaps of the joint positions. This class of methods that leverages supervised learning requires datasets with ground truth annotations of what is being estimated (e.g., 3D joint positions), which is costly to obtain. Alternatively, self-supervised methods and transfer learning can be utilized to train the networks without or with little annotation. Wan *et al*. [50] parameterized hand with 3D spheres and trained a network that estimates the 3D center positions of the spheres with a differentiable depth renderer by minimizing the difference between the rendered and measured depths as shown in Fig. 3(a). Transfer learning was employed in [59] to train a network that predicts the 3D skeleton from an IR image, which is less fragile to motion blur than a depth image.

3.1.2 Color

Color images are much more accessible than depth images. In addition, hand tracking using color image is applicable to numerous videos on the internet while depth images are not easily available. Therefore research on color-based hand tracking is more active than depth-based hand tracking in recent years. In terms of algorithms, the use of deep learning [26,29-31,43,46,48,51,52,55-58,60, 61,63,64] is a dominant stream than optimization-based methods. Many studies train networks in a supervised fashion to estimate from an RGB image either a skeleton with 2.5D/3D joint positions [57,58] or a 3D mesh (e.g., MANO) [26,29,52,56,64]. These methods require datasets with ground truth labels, which are costly to obtain (e.g., motion capture in a multi-view studio). To circumvent this issue, some studies proposed to exploit synthetic data whose ground truth is already known [43,46,48,51,55]. However, synthetic images are usually not photo-realistic, which degrades the estimation when tested on real images. Mueller *et al*. [48] trained a generative adversarial network (GAN) to translate synthetic images to realistic ones such that the translated images follow the distribution of real images. Utilizing a GAN greatly enhanced robustness to small occlusions and varying camera viewpoints as shown in Fig. 3(b). Another approach to circumvent acquiring costly datasets is weakly-supervised learning. In [46,51], the networks were fine-tuned on real data that does not contain 3D ground truth labels by weakly supervising on depth measurements, which are easier to obtain than the 3D ground truths. The hand tracking performance is reasonable as shown in Fig. 3(c), even it has trained without ground truth labels. Similarly, without resorting to synthetic data, Z. Tu, *et al*. [61] utilized self-supervision



(a) (b) (c) (d) (e)

Fig. 3. Hand tracking technologies using a single RGB-D camera - (a) Self-supervised hand tracking by Wan, *et al*. in 2019 [50], where the left and right figures are real depth map and rendered depth map, respectively; a single RGB camera - (b) GANerated Hands by Mueller, *et al*. in 2018 [48], (c) Weakly-supervised hand tracking by Cai, *et al*. in 2018 [46], (d) HandOccNet by Park, *et al*. in 2022 [30]; four monochrome egocentric cameras - (e) MEgATrack by Facebook (now Meta) Reality Labs in 2020 [60].

to handle the lack of 3D annotations in datasets. More recently, Z. Jiang, *et al*. [63] proposed a probabilistic attention model-based hand tracking which does not require large numbers of 3D ground truths and relieves model dependency at the same time. On the other hand, instead of making the best of limited datasets, Pavlakos *et al*. [31] simply scraped up multi-source large-scale data to make the best performance.

Visual occlusion, which is the fundamental limitation of vision, has been challenged by recent studies [27,30,32]. Hu *et al*. [27] developed a set of hand-facing wrist-worn thermal cameras, where the silhouette images of the hand were used to estimate the 3D joint positions through a network. Feature injection/fusion was introduced in [30,32] to exploit the most out of visible parts' information to enhance robustness to occlusion. The exemplar result of [30] is in Fig. 3(d), which shows robustness with object interacting scenarios. Another limitation of 3D hand pose estimation from a single image is that the problem is inherently ill-posed because the depth is not observable. To overcome this, multiple camera images with overlapped field-of-views (FOVs) can render depth observable. For instance, multi-view RGB images were used in [37], and four monochrome fisheye cameras were employed in [60]. The latter one is also embedded to the Meta Quest's hand tracking software as shown in Fig. 3(e). In other studies, an RGB sequence was adopted in [61,65].

3.2. Soft wearable

Soft wearable sensors utilize totally different sensing modality to the vision, which are attachable to the skin or worn as gloves and measure hand poses by sensing joint angles through deformation. However, they lack the ability to provide global position and orientation, necessitating external sensors like cameras for comprehensive pose estimation. Soft wearable sensors also struggle to distinguish between deformation by bending and deformation by contact. Calibration poses another challenge, requiring meticulous setup for each sensor and additional adjustments for different users. Moreover, they exhibit low repeatability due to hysteresis and less durability compared to other sensor types.

These sensors vary in type based on the deformation measured (e.g., bending, stretching) and measuring method (e.g., electrical resistance, capacitance). Common examples include flex sensors, RSS, CSS, FBG sensors, and piezoresistive fabric sensors.

3.2.1 Flex sensor

The resistance of a flex sensor is changed based on the degree of bending, which means that the bending angle is detected simply by measuring its resistance. For hand tracking, flex sensors are used to acquire the finger joint angle [68,71,72,78]. Unfortunately, flex sensors can only detect bending in one direction and are not stretchable, which means they are not suitable for measuring the movement of complex joints such as MCP and CMC joints with multi-DOFs. Also, the wearability of flex sensor is poor due to its low stretchability, but Shen, *et al*. [71] overcame it combining an flex sensor with stretchable belts as shown in Fig. 4 (a). Meanwhile, soft polymer-based RSS and CSS can serve as alternatives for soft wearable sensor-based hand tracking.

3.2.2 Resistive strain sensor (RSS)

Stretchable polymer-based strain sensors can measure more complex joint movements, flexion-extension, and abduction-adduction, by arranging sensors in an array. Embedding microchannels within a soft elastomer and filling these channels with either liquid metal or ionic liquid produces a soft stretchable RSS. This approach offers better wearability than the flex sensor, with a cheap and simple manufacturing process. Several studies have utilized the soft and stretchable RSS for hand tracking [69,70,73,76]. The microchannel in RSS is designed in a zigzag pattern at joint areas to enhance sensitivity as Chossat, *et al*. did in [70] as shown in Fig. 4(b). The main drawback of RSS is its sensitivity to changes in temperature and humidity. Additionally, it demonstrates low durability and suffers from hysteresis. Using CSS can mitigate these drawbacks, making it also commonly used in soft wearable-based hand tracking.

(a) (b) (c) (d) (e)

Fig. 4. Hand tracking technologies using (a) flex sensors by Shen, *et al*. in 2016 [71]; (b) RSS array by Chossat, *et al*. in 2015 [70]; (c) CSS array by Glauser, *et al*. in 2019 [80]; (d) FBG sensors by Silva, *et al*. in 2011 [67]; and (e) piezoresistive fabric for multimodal sensing by Bianchi, *et al*. in 2016 [74].

3.2.3 Capacitive strain sensor (CSS)

Soft stretchable CSS is typically less sensitive to environmental changes (e.g., temperature) and more durable than RSS, hence it is frequently used in recent hand tracking research [77,80,85,100]. It is made from materials like polymer or knit, and measures the change in electric capacitance coming from strain deformation. Especially, Glauser, *et al*. [80] incorporated an extensive array of CSSs (44 in total) on the back of a glove as shown in Fig. 4(c), and utilized deep learning to train hand pose estimation network. This method breaks away from the traditional one-to-one mapping of soft wearable sensors to joints, employing a far greater number of sensors than the hand's DOF.

3.2.4 FBG sensor

FBG sensors, developed more recently compared to the aforementioned sensors, have very high durability, accuracy, and compact size. Exposing UV light to an optical fiber, a periodic grating is created that reflects a specific wavelength, known as the Bragg wavelength. This reflection changes with strain and temperature so that we can sense the strain by measuring the shift in the Bragg wavelength and the temperature. Creating multiple gratings in one long optical fiber, the deformation of multiple joints can be measured using only one fiber, which makes it very compact. Due to these advantages, many studies use FBG sensors for hand tracking [67,81-83]. The major shortcoming of this sensor is the expensive cost of optical equipment. The FBG sensing glove proposed in [67] uses a single optical fiber to cross all the hand as shown in Fig. 4(d), and the fiber is connected to a bulky optical equipment.

Meanwhile, there is a sensor similar to the FBG sensor, named soft polymer acoustic waveguide (SPAW) which measures the time of flight (TOF) of transmitted acoustic waves and echoes to determine the strain within a sensor unit [86]. Unlike the FBG sensor, SPAW utilizes acoustic waves, not light, allowing for various frequency-based techniques.

3.2.5 Piezoresistive fabric

Several studies made the gloves that work as the sensor themselves, which extremely enhances the wearability [74,75,79,84]. The piezoresistive sensor is similar to RSS in that it measures the changes in resistance caused by the sensor deformation. The key difference is that it reacts to external pressure, while RSS reacts to the strain. It is originally developed as a tactile sensor, however, depending on the design of the sensor mechanism, it can also function as a goniometer or a strain gauge. Bianchi, *et al*. designed multimodal sensing gloves using knitted piezoresistive fabric. As shown in Fig. 4(e), it senses the contact normal force on the palmar side and finger bending on the dorsal side [74].

3.3. Encoder

A hand is an articulation of rigid-bodies and hand tracking can be solved by exploiting linkage kinematic structure. Early exoskeleton haptic gloves were studied primarily focusing on rehabilitation and haptic feedback [105-109] with capability of just distinguishing a few grips. Henceforth, exoskeleton haptic gloves are attempting accurate hand tracking through encoder sensing and linkage kinematics analysis, as well as exerting force feedback [87-89]. Performing hand tracking and exerting force feedback at the same time using the same exoskeleton structure is a significant advantage of this type of hand tracking. This contrasts with other kinds of hand tracking systems, which often struggle to integrate with force feedback devices due to their susceptibility to interferences like occlusion and contact. Furthermore, encoders ensure highly precise and reliable sensing. However, the exoskeleton linkage is relatively heavy and constraining finger movements, which causes quick fatigue and limits dexterity. The linkage also only directly measures the target segment (usually the distal phalanx), thus, the full finger pose estimation via inverse kinematics is often not so accurate or, in some cases, even impossible. The device shown in Fig. 5(a) is an exoskeleton device for finger tracking and force feedback, named HEXOTRAC [88], which seems very bulky and tracks the finger-tip segments only.

3.4. Magnetic

The 6 DOF pose magnetic tracking is a well-studied problem [110,111]. Pose tracking using electromagnetic sensors does not require a line of sight and is widely used in the field of human-computer interaction (HCI), with a magnet attached to the finger as an input [112-116]. Such technology has further evolved into hand tracking.

The magnetic hand tracking technology is divided into two methods: using permanent magnets and generating electromagnetic fields with alternating current. The former method is cheaper and easier to wear on the hand, leading to its adoption in various research for hand tracking [90,93]. It however struggles with tracking multiple 3D poses and is influenced by the Earth's magnetic field, reducing accuracy. Meanwhile, alternating current can generate multiple frequencies of oscillating voltage. It is possible to distinguish multiple frequencies and eliminate the influence of the Earth's magnetic field by filtering and amplifying the signal, accurately tracking multiple points for hand tracking [91,92,94]. However, generating strong electromagnetic fields requires bulky equipment. As an example, AuraRing [92] is a 5 DOF electromagnetic tracker which can be worn on the finger as shown in Fig. 5(b), which is quite bulky to wear multiple trackers on the same finger. Both methods have the disadvantage of a nar-

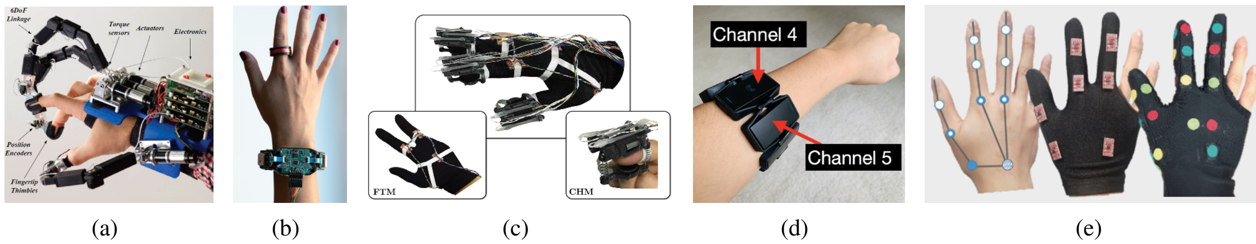


Fig. 5. Hand tracking technologies using encoder - (a) HEXOTRAC by Sarakoglou, *et al*. in 2016 [88]; magnetic sensors - (b) AuraRing by Parizi, *et al*. in 2019 [92]; IMUs and compasses - (c) Finger Tracking Modules (FTMs) by Lee, *et al*. in 2018 [100]; EMG - (d) NeuroPose by Liu, *et al*. in 2021 [102]; and visual-inertial fusion - (e) VIST glove by Lee, *et al*. in 2021 [19].

row tracking range near the source of the electromagnetic field. They are also susceptible to magnetic interference, significantly affected by nearby steel structures and electronic devices, limiting their applications.

3.5. IMU & compass

Unlike vision-based hand tracking, IMU-based hand tracking does not suffer from occlusion or changing light conditions. Furthermore, it does not require laborious calibration for each user, unlike soft wearable sensor-based hand tracking. At the early stage of research, only 3-axis accelerometers were used for hand tracking [96,97]. Assuming a pseudo-static state, gesture recognition was performed by sensing the direction of gravity. Subsequently, a 9-axis IMU, adding on a 3-axis gyroscope and a 3-axis compass, enabled dynamic hand motion tracking [98-101]. Here, the gyroscope detects angular velocity, and the compass corrects orientation errors. Although an IMU is low-cost and calibration is relatively simple, it suffers from global position drift. In addition, the compass is susceptible to magnetic interference and the accelerometer is susceptible to impacts. Thus, in [100], Lee, *et al*. utilized combination of 9-axis IMUs and CSS sensors to avoid magnetic interference caused by a finger-tip haptic device as shown in Fig. 5(c).

3.6. Electromyography

Electromyography (EMG) sensors detect electrical signals from muscle activation. Since the muscles that activate finger movements are located in the forearm, EMG sensors can detect hand gestures by being worn only around the wrist, without the need for any hand-worn sensors or external cameras. However, due to the noisy and inaccurate sensing, they have primarily been used for discrete gesture recognition [117-122]. NeuroPose [102] is one of the few studies on continuous 3D hand tracking using EMG sensors with machine learning-based techniques. It uses only a wrist-held armband as shown in Fig. 5(d). Although the accuracy of the sensors is not yet guaranteed, EMG sensor-based hand tracking is promising because it is easy to integrate with other sensors and haptic devices.

3.7. Fusion of sensor modalities

Due to the limitations inherent in each sensor modality, researchers have developed hand tracking techniques that utilize multiple sensor modalities. Combining vision and inertial measurements, several studies have achieved robust hand tracking in scenarios of occlusion and rapid motion [17-19]. To begin with, a robust initializer that fuses depth images and IMUs was suggested in [17], which is essential to handle difficulties of vision-based methods (e.g., loss of tracking due to occlusion and rapid motion). Motion blur induced by rapid motion was dealt with in [18], which overcame this problem through the fusion of depth camera and gyroscope. Occlusion and out-of-FOV problems were addressed in [19], which introduced a tightly-coupled sensor fusion of stereo RGB images and 6-axis IMUs that do not cause a magnetic interference problem. The hand tracking glove consists of two layers-inner layer for IMU array attachment, and outer layer for colored blobs attachment as shown in Fig. 5(e) shows. The stereo vision corrects IMU drifts and the IMUs help matching detected markers to the corresponding model marker set.

Vision sensor, soft wearable sensor, and IMU are integrated altogether in [20], where the global 3D hand position (i.e., wrist position) was estimated using an RGB-D camera and IMUs on the arm, and the local hand pose was estimated with a depth image and soft wearable sensors.

Without the vision modality, the combination of EMG and gyroscope measurements were tried in [103]. The hand motion is tracked using an armband device worn on the wrist (Myo and gForce [104]) with a recursive neural network (RNN), but the accuracy is not good due to the noise of EMG sensors.

The biggest merit of the fusion of sensor modalities is that it enables tracking in challenging scenarios such as interacting with electric devices while a significant portion of the hand is occluded, which was originally impossible by exploiting a single modality. However, complementarily combining different sensor modalities without disturbing each other is difficult in terms of hardware and software because each sensor modality is different in sensing method, sensing period, and mapping method. Therefore, more research on the hand tracking by fusion of sensor modalities is needed.

4. COMMERCIAL PRODUCTS

Hand tracking technologies have improved significantly over the past few decades, leading to their widespread release as commercial products. While various sensor-based hand tracking technologies have reached a usable level of sophistication, there is no solution yet that works robustly in a wide variety of scenarios and offers high accuracy at the same time. In this section, we will survey some of the commercial products for each type of hand tracking. More products are discussed in detail in another paper [130] focusing on commercial data gloves.

4.1. Vision

Leap Motion Controller 2, Stereo IR 170, and Ultraleap 3Di are Leap Motion's stereo IR cameras, and the hand tracking software is available on the website [123]. Fig. 6(a) is the Ultraleap 3Di. The price[1] of these products are US$139, US$262, and US$275, respectively. Meta's head mounted displays (HMD), Quest series [131], are also equipped with vision-based hand tracking. The latest product within the series, Quest 3, can track hands with four IR cameras and two RGB cameras embedded in the HMD, and the price is US$499. Apple Vision Pro [132], which is the newest HMD released on February 2024, are also equipped with hand tracking with its own stereoscopic 3D main camera system, and the price is US$3,499. HMDs are more expensive than camera-only products because they include display, battery, applications, and other additional functions. Vision-based hand tracking systems tend to be cheaper than other types of hand tracking systems to be described next, because they do not need to fabricate data gloves. However, they suffer from self-occlusion, object occlusion, and narrow FOV because they rely on cameras.

[1] All the prices in this paper are as of April 2024.

4.2. Soft wearable

The CyberGlove series by CyberGlove Systems [124] are soft wearable sensor-based hand tracking gloves. The latest version, CyberGlove 3, is equipped with either 18 or 22 soft sensors embedded in the glove to estimate hand pose, and it costs around US$13,000. The appearance of this glove is in Fig. 6(b). StretchSense [133] also produces motion capture gloves, which equips 16 or 26 soft sensors, and the prices varies from US$795 to US$6,995 according to the product lines. Soft wearable sensor-based hand tracking gloves cost more than vision-based hand tracking systems on average due to their difficulty in fabrication. These gloves need an additional external motion capture system to track their global position and orientation, and they are vulnerable to contact on the sensor location.

Fig. 6. (a) Ultraleap 3Di by Leap Motion [123]. (b) CyberGlove 3 by CyberGlove Systems [124]. (c) Dexmo by Dextarobotics [125]. (d) Sense Glove DK1 by Sense Glove [126]. (e) Manus Quantum Gloves by Manus VR [127]. (f) HaptX Gloves G1 by HaptX [128]. (g) Manus Prime 3 by ManusVR [127]. (h) Motion Glove by Quester [129].

4.3. Encoder

Dextarobotics' Dexmo haptic glove [125] and Senseglove's Sense Glove DK1 [126] track hand motion using encoders. These gloves have the advantage of providing force feedback, but the exoskeleton structures may cause the feeling of high resistance that disturbs the fast movement of fingers. In addition, as you can check it from their appearances in Figs. 6(b) and 6(c), they can only track finger-tip segments and calculate finger poses via inverse kinematics which leads to inaccuracy of tracking. Dexmo haptic glove is currently not available to consumers, and Senseglove discontinues Sense Glove DK1, but the price was about €3,000 when it was served.

4.4. Magnetic

Manus Quantum Metagloves [127] and HaptX Gloves G1 [128] use electromagnetic field sensors for hand tracking. The Manus Quantum Metagloves costs US$4,999, and HaptX Gloves G1 costs US$5,495. They appears in Figs. 6(e) and 6(f), respectively. These systems are fragile to magnetic interference and must be kept away from electric devices and metals. Meanwhile, HaptX Gloves G1 includes a huge pneumatic system to provide realistic tactile haptic feedback.

4.5. IMU & compass

Manus Prime 3 [127] is equipped with IMUs, compasses, and flex sensors embedded in the glove for hand tracking, and also provides vibrotactile haptic feedback. Manus Prime 3 needs an external motion capture system to track global position and orientation, and they are vulnerable to magnetic interference due to the use of compasses. The appearance of this glove is shown in Fig. 6(g). It costs US$2,999.

4.6. Visual-inertial fusion

Quester Motion Glove [129] is equipped with multiple 6-axis IMUs without compasses and utilizes a stereo RGB camera to capture colored markers on the gloves. This product is not currently available, but it will be released soon. The price is expected to be low because it utilizes only low-cost sensors, and it does not require calibration thanks to the auto-calibration algorithm making it comfortable to use. The core technology used in Quester Motion Glove is the same with [19], and the concept picture of this product is in Fig. 6(h).

5. CONCLUSION

In this survey paper, we analyze the existing hand tracking methodologies and organize them into seven categories with respect to the sensing modality. Hand tracking technology has kept advancing from recognizing pseudo-static gestures to dynamic gestures, from tracking free motion to object-interacting motion, and from reconstructing hand skeletal structure to hand surface mesh, gradually solving more complex problems. However, single modality-based hand tracking has fundamental limitations due to its sensing method. Consequently, we found out that the fusion of sensor modalities is essential to break the limit of current hand tracking technology despite of its difficulty in terms of hardware and software. In this regard, we hope that this paper will help accelerate the interdisciplinary study among the research fields (e.g., computer graphics, soft wearable sensors, mechatronics, and mobile robotics) to develop an ultimate solution of hand tracking. With more advanced hand tracking technology, we expect that various tasks and interactions in HRI/HCI and VR/AR can become as realistic as those in the real world.

CONFLICT OF INTEREST

Dongjun Lee is a Senior Editor of International Journal of Control, Automation, and Systems. Senior Editor status has no bearing on editorial consideration. The authors declare that there is no competing financial interest or personal relationship that could have appeared to influence the work reported in this paper.

REFERENCES

[1] J. Lee and T. L. Kunii, "Constraint-based hand animation," *Models and Techniques in Computer Animation*, pp. 110-127, 1993.

[2] I. Oikonomidis, N. Kyriazis, and A. A. Argyros, "Markerless and efficient 26-DOF hand pose recovery," *Proc. of Asian Conference on Computer Vision*, Berlin, pp. 744-757, 2010.

[3] C. Qian, X. Sun, Y. Wei, X. Tang, and J. Sun, "Realtime and robust hand tracking from depth," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1106-1113, 2014.

[4] J. Romero, D. Tzionas, and M. J. Black, "Embodied hands: Modeling and capturing hands and bodies together," *arXiv preprint arXiv:2201.02610*, 2022.

[5] R. A. Potamias, S. Ploumpis, S. Moschoglou, V. Triantafyllou, and S. Zafeiriou, "Handy: Towards a high fidelity 3D hand shape and appearance model," *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4670-4680, 2023.

[6] G. Buckingham, "Hand tracking for immersive virtual reality: opportunities and challenges," *Frontiers in Virtual Reality*, vol. 2, 728461, 2021.

[7] B. Cha, Y. Bae, C. Lee, D. Jeong, and J. Ryu, "Design and user evaluation of haptic augmented virtuality system for immersive virtual training," *International Journal of Control, Automation, and Systems*, vol. 20, no. 9, pp. 3032-3044, 2022.

[8] ANA Avatar XPRIZE, https://www.xprize.org/prizes/avatar/, Accessed March 2024.

[9] Tesla Optimus, https://www.youtube.com/@tesla/, Accessed March 2024.

[10] Apptronik Apollo, https://apptronik.com/apollo/, Accessed March 2024.

[11] A. Rajeswaran, V. Kumar, A. Gupta, G. Vezzani, J. Schulman, E. Todorov, and S. Levine, "Learning complex dexterous manipulation with deep reinforcement learning and demonstrations," *arXiv preprint arXiv:1709.10087*, 2017.

[12] Z. Q. Chen, K. van Wyk, Y. Chao, W. Yang, A. Mousavian, A. Gupta, and D. Fox, "Dextransfer: Real world multi-fingered dexterous grasping with minimal human demonstrations," *arXiv preprint arXiv:2209.14284*, 2022.

[13] Y. Qin, Y. Wu, S. Liu, H. Jiang, R. Yang, Y. Fu, and X. Wang, "Dexmv: Imitation learning for dexterous manipulation from human videos," *Proc. of European Conference on Computer Vision*, pp. 570-587, 2022.

[14] S. Dasari, A. Gupta, and V. Kumar, "Learning dexterous manipulation from exemplar object trajectories and pre-grasps," *Proc. of IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3889-3896, 2023.

[15] S. P. Arunachalam, I. Güzey, S. Chintala, and L. Pinto, "Holo-dex: Teaching dexterity with immersive mixed reality," *Proc. of 2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5962-5969, 2023.

[16] S. P. Arunachalam, S. Silwal, B. Evans, and L. Pinto, "Dexterous imitation made easy: A learning-based framework for efficient dexterous manipulation," *Proc. of 2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5954-5961, 2023.

[17] T. Zhang, H. Xia, C. Zhang, and Z. Zeng, "MultiModal, robust and accurate hand tracking," *Proc. of IEEE 6th International Conference on Computer and Communications (ICCC)*, pp. 1886-1890, 2020.

[18] G. Park, A. Argyros, J. Lee, and W. Woo, "3D hand tracking in the presence of excessive motion blur," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 5, pp. 1891-1901, 2020.

[19] Y. Lee, W. Do, H. Yoon, J. Heo, W. Lee, and D. Lee, "Visual-inertial hand motion tracking with robustness against occlusion, interference, and contact," *Science Robotics*, vol. 6, no. 58, 2021.

[20] N. Gosala, F. Wang, Z. Cui, H. Liang, O. Glauser, S. Wu, and O. Sorkine-Hornung, "Self-Calibrated Multi-Sensor Wearable for Hand Tracking and Modeling," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 3, pp. 1769-1784, 2023.

[21] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly, "Vision-based hand pose estimation: A review," *Computer Vision and Image Understanding*, vol. 108, no. 1-2, pp. 52-73, 2007.

[22] B. Doosti, "Hand pose estimation: A survey," *arXiv preprint arXiv:1903.01013*, 2019.

[23] X. Hu, Y. Xu, H. Zhang, J. Xie, D. Niu, Z. Zhao, and X. Qu, "The fiber bragg grating (FBG) sensing glove: A review," *IEEE Sensors Journal*, vol. 23, no. 11, pp. 11374-11382, 2023.

[24] T. Li and H. Yu, "Visual-Inertial Fusion-Based Human Pose Estimation: A Review," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, 4007816, 2023.

[25] W. Chen, C. Yu, C. Tu, Z. Lyu, J. Tang, S. Ou, Y. Fu, and Z. Xue, "A survey on hand pose estimation with wearable sensors and computer-vision-based methods," *Sensors*, vol. 20, no. 4, 1074, 2020.

[26] S. Baek, K. Kim, and T. Kim, "Pushing the envelope for rgb-based dense 3d hand pose estimation via neural rendering," *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1067-1076, 2019.

[27] F. Hu, P. He, S. Xu, Y. Li, and C. Zhang, "FingerTrak: Continuous 3D hand pose tracking by deep learning hand silhouettes captured by miniature thermal cameras on wrist," *Proc. of ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 2, pp. 1-24, 2020.

[28] S. Hampali, S. D. Sarkar, M. Rad, and V. Lepetit, "Keypoint transformer: Solving joint identification in challenging hands and object interactions for accurate 3D pose estimation," *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11090-11100, 2022.

[29] M. Li, L. An, H. Zhang, L. Wu, F. Chen, T. Yu, and Y. Liu, "Interacting attention graph for single image two-hand reconstruction," *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2761-2770, 2022.

[30] J. Park, Y. Oh, G. Moon, H. Choi, and K. M. Lee, "Handoccnet: Occlusion-robust 3D hand mesh estimation network," *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1486-1495, 2022.

[31] G. Pavlakos, D. Shan, I. Radosavovic, A. Kanazawa, D. Fouhey, and J. Malik, "Reconstructing Hands in 3D with Transformers," *arXiv preprint arXiv:2312.05251*, 2023.

[32] H. Xu, T. Wang, X. Tang, and C. Fu, "H2ONet: Hand-occlusion-and-orientation-aware network for real-time 3D hand mesh reconstruction," *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17048-17058, 2023.

[33] G. Pavlakos, V. Choutas, N. Ghorbani, T. Bolkart, A. A. Osman, D. Tzionas, and M. J. Black, "Expressive body capture: 3D hands, face, and body from a single image," *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10967-10977, 2019.

[34] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, "SMPL: A skinned multi-person linear model," *Seminal Graphics Papers: Pushing the Boundaries*, vol. 2, pp. 851-866, 2023.

[35] T. Li, T. Bolkart, M. J. Black, H. Li, and J. Romero, "Learning a model of facial shape and expression from 4D scans," *ACM Transactions on Graphics (ToG)*, vol. 36, no. 6, 194, 2017.

[36] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in Neural Information Processing Systems*, vol. 27, 2014.

[37] S. Sridhar, H. Rhodin, H. Seidel, A. Oulasvirta, and C. Theobalt, "Real-time hand tracking using a sum of anisotropic gaussians model," *Proc. of The 2nd International Conference on 3D Vision*, vol. 1, pp. 319-326, 2014.

[38] S. Sridhar, A. Oulasvirta, and C. Theobalt, "Interactive markerless articulated hand motion tracking using RGB and depth data," *Proc. of IEEE International Conference on Computer Vision*, pp. 2456-2463, 2013.

[39] T. Sharp, C. Keskin, D. Robertson, J. Taylor, J. Shotton, D. Kim, C. Rhemann, I. Leichter, A. Vinnikov, Y. Wei, *et al.*, "Accurate, robust, and flexible real-time hand tracking," *Proc. of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 3633-3642, 2015.

[40] S. Sridhar, F. Mueller, A. Oulasvirta, and C. Theobalt, "Fast and robust hand tracking using detection-guided optimization," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3213-3221, 2015.

[41] A. Makris, N. Kyriazis, and A. A. Argyros, "Hierarchical particle filtering for 3d hand tracking," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 8-17, 2015.

[42] F. Mueller, D. Mehta, O. Sotnychenko, S. Sridhar, D. Casas, and C. Theobalt, "Real-time hand tracking under occlusion from an egocentric RGB-D sensor," *Proc. of IEEE International Conference on Computer Vision*, pp. 1154-1163, 2017.

[43] C. Zimmermann and T. Brox, "Learning to estimate 3D hand pose from single RGB images," *Proc. of IEEE International Conference on Computer Vision*, pp. 4903-4911, 2017.

[44] J. Y. Chang, G. Moon, and K. M. Lee, "V2V-PoseNet: Voxel-to-voxel prediction network for accurate 3D hand and human pose estimation from a single depth map," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5079-5088, 2018.

[45] J. Tompson, M. Stein, Y. Lecun, and K. Perlin, "Real-time continuous pose recovery of human hands using convolutional networks," *ACM Transactions on Graphics (ToG)*, vol. 33, no. 5, pp. 1-10, 2014.

[46] Y. Cai, L. Ge, J. Cai, and J. Yuan, "Weakly-supervised 3D hand pose estimation from monocular RGB images," *Proc. of European Conference on Computer Vision (ECCV)*, pp. 666-682, 2018.

[47] J. Zhang, J. Jiao, M. Chen, L. Qu, X. Xu, and Q. Yang, "A hand pose tracking benchmark from stereo matching," *Proc. of IEEE International Conference on Image Processing (ICIP)*, pp. 982-986, 2017.

[48] F. Mueller, F. Bernard, O. Sotnychenko, D. Mehta, S. Sridhar, D. Casas, and C. Theobalt, "GANerated hands for real-time 3D hand tracking from monocular RGB," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 49-59, 2018.

[49] J. Zhang, J. Jiao, M. Chen, L. Qu, X. Xu, and Q. Yang, "3D hand pose tracking and estimation using stereo matching," *arXiv preprint arXiv:1610.07214*, 2016.

[50] C. Wan, T. Probst, L. V. Gool, and A. Yao, "Self-supervised 3D hand pose estimation through training by fitting," *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10853-10862, 2019.

[51] L. Ge, Z. Ren, Y. Li, Z. Xue, Y. Wang, J. Cai, and J. Yuan, "3D hand shape and pose estimation from a single RGB image," *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10825-10834, 2019.

[52] D. Xiang, H. Joo, and Y. Sheikh, "Monocular total capture: Posing face, body, and hands in the wild," *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10957-10966, 2019.

[53] H. Joo, T. Simon, and Y. Sheikh, "Total capture: A 3D deformation model for tracking faces, hands, and bodies," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8320-8329, 2018.

[54] S. Sridhar, F. Mueller, M. Zollhöfer, D. Casas, A. Oulasvirta, and C. Theobalt, "Real-time joint tracking of a hand manipulating an object from RGB-D input," *Proc. of European Conference on Computer Vision (ECCV)*, pp. 294-310, 2016.

[55] Y. Li, Z. Xue, Y. Wang, L. Ge, Z. Ren, and J. Rodriguez, "End-to-end 3D hand pose estimation from stereo cameras," *Proc. of The British Machine Vision Conference (BMVC)*, 2019.

[56] S. Hampali, M. Rad, M. Oberweger, and V. Lepetit, "Honnotate: A method for 3D annotation of hand and object poses," *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3193-3203, 2020.

[57] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C. L. Chang, and M. Grundmann, "Mediapipe hands: On-device real-time hand tracking," *arXiv preprint arXiv:2006.10214*, 2020.

[58] G. Moon, S. I. Yu, H. Wen, T. Shiratori, and K. Mu. Lee, "InterHand2.6M: A dataset and baseline for 3D interacting hand pose estimation from a single RGB image," *Proc. of European Conference on Computer Vision (ECCV)*, pp. 548-564, 2020.

[59] G. Park, T. Kim, and W. Woo, "3D hand pose estimation with a single infrared camera via domain transfer learning," *Proc. of IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 588-599, 2020.

[60] S. Han, B. Liu, R. Cabezas, C. D. Twigg, P. Zhang, J. Petkau, T. H. Yu, C. J. Tai, M. Akbay, Z. Wang, *et al.*, "MEgATrack: monochrome egocentric articulated hand-tracking for virtual reality," *ACM Transactions on Graphics (ToG)*, vol. 39, no. 4, 2020.

[61] Z. Tu, Z. Huang, Y. Chen, D. Kang, L. Bao, B. Yang, and J. Yuan, "Consistent 3d hand reconstruction in video via self-supervised learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[62] C. Zimmermann, D. Ceylan, J. Yang, B. Russell, M. Argus, and T. Brox, "Freihand: A dataset for markerless capture of hand pose and shape from single RGB images," *Proc. of IEEE/CVF International Conference on Computer Vision*, pp. 813-822, 2019.

[63] Z. Jiang, H. Rahmani, S. Black, and B. M. Williams, "A probabilistic attention model with occlusion-aware texture regression for 3D hand reconstruction from a single RGB image," *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 758-767, 2023.

[64] Z. Yu, S. Huang, C. Fang, T. P. Breckon, and J. Wang, "ACR: Attention collaboration-based regressor for arbitrary two-hand reconstruction," *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12955-12964, 2023.

[65] K. Karunratanakul, S. Prokudin, O. Hilliges, and S. Tang, "HARP: Personalized hand reconstruction from a monocular RGB video," *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12802-12813, 2023.

[66] Y. W. Chao, W. Yang, Y. Xiang, P. Molchanov, A. Handa, J. Tremblay, Y. S. Narang, K. Van Wyk, U. Iqbal, S. Birchfield, *et al*., "DexYCB: A benchmark for capturing hand grasping of objects," *Proc. of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9040-9049, 2021.

[67] Alexandre F. Da S., A. F. Gonçalves, P. M. Mendes, and J. H. Correia, "FBG sensing glove for monitoring hand posture," *IEEE Sensors Journal*, vol. 11, no. 10, pp. 2442-2448, 2011.

[68] G. Saggio, "A novel array of flex sensors for a goniometric glove," *Sensors and Actuators A: Physical*, 205, 2014.

[69] J. Lee, S. Kim, J. Lee, D. Yang, B. C. Park, S. Ryu, and I. Park, "A stretchable strain sensor based on a metal nanoparticle thin film for human motion detection," *Nanoscale*, vol. 6, no. 20, 2014.

[70] J. B. Chossat, Y. Tao, V. Duchaine, and Y. L. Park, "Wearable soft artificial skin for hand motion detection with embedded microfluidic strain sensing," *Proc. of IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2568-2573, 2015.

[71] Z. Shen, J. Yi, X. Li, M. H. P. Lo, M. ZQ. Chen, Y. Hu, and Z. Wang, "A soft stretchable bending sensor and data glove applications," *Robotics and Biomimetics*, vol. 3, no. 1, 22, 2016.

[72] Y. Zheng, Y. Peng, G. Wang, X. Liu, X. Dong, and J. Wang, "Development and evaluation of a sensor glove for hand function assessment and preliminary attempts at assessing hand coordination," *Measurement*, vol. 93, pp. 1-12, 2016.

[73] H. O. Michaud, L. Dejace, S. De Mulatier, and S. P. Lacour, "Design and functional evaluation of an epidermal strain sensing system for hand tracking," *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3186-3191, 2016.

[74] M. Bianchi, R. Haschke, G. Büscher, S. Ciotti, N. Carbonaro, and A. Tognetti, "A multi-modal sensing glove for human manual-interaction studies," *Electronics*, vol. 5, no. 3, 42, 2016.

[75] S. Ciotti, E. Battaglia, N. Carbonaro, A. Bicchi, and M. Tognetti, A.and Bianchi, "A synergy-based optimally designed sensing glove for functional grasp recognition," *Sensors*, vol. 16, no. 6, 811, 2016.

[76] W. Park, K. Ro, S. Kim, and J. Bae, "A soft sensor-based three-dimensional (3-D) finger motion measurement system," *Sensors*, vol. 17, no. 2, 420, 2017.

[77] A. Atalay, V. Sanchez, O. Atalay, D. M. Vogt, F. Haufe, R. J. Wood, and C. J. Walsh, "Batch fabrication of customizable silicone-textile composite capacitive strain sensors for human motion tracking," *Advanced Materials Technologies*, vol. 2, no. 9, 1700136, 2017.

[78] T. K. Chan, Y. K. Yu, H. C. Kam, and K. H. Wong, "Robust hand gesture input using computer vision, inertial measurement unit (IMU) and flex sensors," *Proc. of IEEE International Conference on Mechatronics, Robotics and Automation (ICMRA)*, pp. 95-99, 2018.

[79] H. Ryu, S. Park, J. Park, and J. Bae, "A knitted glove sensing system with compression strain for finger movements," *Smart Materials and Structures*, vol. 27, no. 5, 055016, 2018.

[80] O. Glauser, S. Wu, D. Panozzo, O. Hilliges, and O. Sorkine-Hornung, "Interactive hand pose estimation using a stretch-sensing soft glove," *ACM Transactions on Graphics (ToG)*, vol. 38, no. 4, pp. 1-15, 2019.

[81] Y. Jiang, V. Reimer, T. Schossig, M. Angelmahr, and W. Schade, "Fiber optical multifunctional human-machine interface for motion capture, temperature, and contact force monitoring," *Optics and Lasers in Engineering*, vol. 128, 106018, 2020.

[82] J. S. Kim, B. K. Kim, M. Jang, K. Kang, D. E. Kim, B. K. Ju, and J. Kim, "Wearable hand module and real-time tracking algorithms for measuring finger joint angles of different hand sizes with high accuracy using FBG strain sensor," *Sensors*, vol. 20, no. 7, 1921, 2020.

[83] C. K. Jha, K. Gajapure, and A. L. Chakraborty, "Design and evaluation of an FBG sensor-based glove to simultaneously monitor flexure of ten finger joints," *IEEE Sensors Journal*, vol. 21, no. 6, pp. 7620-7630, 2020.

[84] E. Ayodele, S. A. R. Zaidi, J. Scott, Z. Zhang, A. Hayajneh, S. Shittu, and D. McLernon, "A weft knit data glove," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-12, 2021.

[85] L. Behnke, L. Sanchez-Botero, W. R. Johnson, A. Agrawala, and R. Kramer-Bottiglio, "Dynamic hand proprioception via a wearable glove with fabric sensors," *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 149-154, 2023.

[86] Y. Lin, P. B. Shull, and J. B. Chossat, "Design of a wearable real-time hand motion tracking system using an array of soft polymer acoustic waveguides," *Soft Robotics*, vol. 11, no. 2, pp. 282-295, 2023.

[87] M. A. Zhou and P. Ben-Tzvi, "RML glove—An exoskeleton glove mechanism with haptics feedback," *IEEE/ASME Transactions on Mechatronics*, vol. 20, no. 2, pp. 641-652, 2014.

[88] I. Sarakoglou, A. Brygo, D. Mazzanti, N. G. Hernandez, D. Caldwell, and N. G. Tsagarakis, "Hexotrac: A highly under-actuated hand exoskeleton for finger tracking and force feedback," *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1033-1040, 2016.

[89] M. Palagi, G. Santamato, D. Chiaradia, M. Gabardi, S. Marcheschi, M. Solazzi, A. Frisoli, and D. Leonardis, "A mechanical hand-tracking system with tactile feedback designed for telemanipulation," *IEEE Transactions on Haptics*, vol. 16, no. 4, pp. 594-601, 2023.

[90] Y. Ma, Z. H. Mao, W. Jia, C. Li, J. Yang, and M. Sun, "Magnetic hand tracking for human-computer interface," *IEEE Transactions on Magnetics*, vol. 47, no. 5, pp. 970-973, 2011.

[91] K. Chen, S. N. Patel, and S. Keller, "Finexus: Tracking precise motions of multiple fingertips using magnetic sensing," *Proc. of CHI Conference on Human Factors in Computing Systems*, pp. 1504-1514, 2016.

[92] F. S. Parizi, E. Whitmire, and S. Patel, "Auraring: Precise electromagnetic finger tracking," *Proc. of ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 4, pp. 1-28, 2019.

[93] R. Casas, K. Martin, M. Sandison, and P. S. Lum, "A tracking device for a wearable high-DOF passive hand exoskeleton," *Proc. of The 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 6643-6646, 2021.

[94] F. Santoni, A. De Angelis, A. Moschitta, and P. Carbone, "MagIK: A hand-tracking magnetic positioning system based on a kinematic model of the hand," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, 9507313, 2021.

[95] Polhemus, https://polhemus.com/motion-tracking/hand-and-finger-trackers/, Accessed March 2024.

[96] J. K. Perng, B. Fisher, S. Hollar, and K. S. J. Pister, "Acceleration sensing glove (ASG)," *Proc. of the 3rd International Symposium on Wearable Computers*, vol. 178-180, 1999.

[97] J. Kim, N. D. Thang, and T. Kim, "3-D hand motion tracking and gesture recognition using a data glove," *Proc. of IEEE International Symposium on Industrial Electronics*, pp. 1013-1018, 2009.

[98] B. O'Flynn, J. T. Sanchez, J. Connolly, J. Condell, K. Curran, P. Gardiner, and B. Downes, "Integrated smart glove for hand motion monitoring," *Proc. of The 6th International Conference on Sensor Device Technologies and Applications*, 2015.

[99] T. L. Baldi, S. Scheggi, L. Meli, M. Mohammadi, and D. Prattichizzo, "GESTO: A glove for enhanced sensing and touching based on inertial and magnetic sensors for hand tracking and cutaneous feedback," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 6, pp. 1066-1076, 2017.

[100] Y. Lee, M. Kim, Y. Lee, J. Kwon, Y. L. Park, and D. Lee, "Wearable finger tracking and cutaneous haptic interface with soft sensors for multi-fingered virtual manipulation," *IEEE/ASME Transactions on Mechatronics*, vol. 24, no. 1, pp. 67-77, 2018.

[101] H. T. Chang and J. Y. Chang, "Sensor glove based on novel inertial sensor fusion control algorithm for 3-D real-time hand gestures measurements," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 1, pp. 658-666, 2020.

[102] Y. Liu, S. Zhang, and M. Gowda, "NeuroPose: 3D hand pose tracking using EMG wearables," *Proc. of the Web Conference 2021*, pp. 1471-1482, 2021.

[103] Y. Liu, C. Lin, and Z. Li, "WR-Hand: Wearable armband can track user's hand," *Proc. of ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 5, no. 3, pp. 1-27, 2021.

[104] gForce EMG Armband, https://www.oymotion.com/, Accessed March 2024.

[105] M. Bouzit, G. Burdea, G. Popescu, and R. Boian, "The Rutgers Master II-new design force-feedback glove," *IEEE/ASME Transactions on Mechatronics*, vol. 7, no. 2, pp. 256-263, 2002.

[106] C. N. Schabowsky, S. B. Godfrey, R. J. Holley, and P. S. Lum, "Development and pilot testing of HEXORR: Hand EXOskeleton rehabilitation robot," *Journal of Neuroengineering and Rehabilitation*, vol. 7, pp. 1-16, 2010.

[107] H. In, K. Cho, K. Kim, and B. Lee, "Jointless structure and under-actuation mechanism for compact hand exoskeleton," *Proc. of IEEE International Conference on Rehabilitation Robotics*, pp. 1-6, 2011.

[108] J. Iqbal, H. Khan, N. G. Tsagarakis, and D. G. Caldwell, "A novel exoskeleton robotic system for hand rehabilitation—conceptualization to prototyping," *Biocybernetics and Biomedical Engineering*, vol. 34, no. 2, pp. 79-89, 2014.

[109] H. In, B. B. Kang, M. Sin, and K. Cho, "Exo-glove: A wearable robot for the hand with a soft tendon routing system," *IEEE Robotics & Automation Magazine*, vol. 22, no. 1, pp. 97-105, 2015.

[110] F. H. Raab, E. B. Blood, T. O. Steiner, and H. R. Jones, "Magnetic position and orientation tracking system," *IEEE Transactions on Aerospace and Electronic Systems*, no. 5, pp. 709-718, 1979.

[111] J. B. Kuipers, "SPASYN-an electromagnetic relative position and orientation tracking system," *IEEE Transactions on Instrumentation and Measurement*, vol. 29, no. 4, pp. 462-466, 1980.

[112] C. Harrison and S. E. Hudson, "Abracadabra: wireless, high-precision, and unpowered finger input for very small mobile devices," *Proc. of The 22nd Annual ACM Symposium on User Interface Software and Technology*, pp. 121-124, 2009.

[113] D. Ashbrook, P. Baudisch, and S. White, "Nenya: Subtle and eyes-free mobile input with a magnetically-tracked finger ring," *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2043-2046, 2011.

[114] K. Chen, K. Lyons, S. White, and S. Patel, "uTrack: 3D input using two magnetic sensors," *Proc. of the 26th Annual ACM Symposium on User Interface Software and Technology*, pp. 237-244, 2013.

[115] L. Chan, R. Liang, M. Tsai, K. Cheng, C. Su, M. Y. Chen, W. Cheng, and B. Chen, "FingerPad: Private and subtle interaction using fingertips," *Proc. of the 26th Annual ACM Symposium on User Interface Software and Technology*, pp. 255-260, 2013.

[116] J. McIntosh, P. Strohmeier, J. Knibbe, S. Boring, and K. Hornbæk, "Magnetips: Combining fingertip tracking and haptic feedback for around-device interaction," *Proc. of CHI Conference on Human Factors in Computing Systems*, pp. 1-12, 2019.

[117] Y. Du, Y. Wong, W. Jin, W. Wei, Y. Hu, M. S. Kankanhalli, and W. Geng, "Semi-Supervised Learning for Surface EMG-based Gesture Recognition," *Proc. of International Joint Conference on Artificial Intelligence*, 2017.

[118] F. Quivira, T. Koike-Akino, Y. Wang, and D. Erdogmus, "Translating sEMG signals to continuous hand poses using recurrent neural networks," *Proc. of IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pp. 166-169, 2018.

[119] S. Raurale, J. McAllister, and J. M. del Rincon, "EMG acquisition and hand pose classification for bionic hands from randomly-placed sensors," *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1105-1109, 2018.

[120] I. Sosin, D. Kudenko, and A. Shpilman, "Continuous gesture recognition from sEMG sensor data with recurrent neural networks and adversarial domain adaptation," *Proc. of The 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pp. 1436-1441, 2018.

[121] A. D. Silva, M. V. Perera, K. Wickramasinghe, A. M. Naim, T. D. Lalitharatne, and S. L. Kappel, "Real-time hand gesture recognition Using temporal muscle activation maps of multi-channel sEMG signals," *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1299-1303, 2020.

[122] D. Oh and Y. Jo, "Classification of hand gestures based on multi-channel EMG by scale Average wavelet transform and convolutional neural network," *International Journal of Control, Automation, and Systems*, vol. 19, no. 3, pp. 1443-1450, 2021.

[123] Leap Motion, https://www.ultraleap.com/product/leap-motion-controller/, Accessed March 2024.

[124] Cyber Glove Systems, https://www.cyberglovesystems.com/, Accessed March 2024.

[125] Dexmo, https://www.dextarobotics.com/en-us/, Accessed March 2024.

[126] Sense Glove, https://www.senseglove.com/, Accessed March 2024.

[127] Manus VR, https://manus-vr.com/, Accessed March 2024.

[128] HaptX Gloves, https://haptx.com/, Accessed March 2024.

[129] Quester Motion Glove, https://quester.kr/, Accessed March 2024.

[130] M. Caeiro-Rodríguez, I. Otero-González, F. A. Mikic-Fonte, and M. Llamas-Nistal, "A systematic review of commercial smart gloves: Current status and applications," *Sensors*, vol. 21, no. 8, 2667, 2021.

[131] Meta Quest, https://www.meta.com/, Accessed March 2024.

[132] Apple Vision Pro, https://www.apple.com/apple-vision-pro/, Accessed March 2024.

[133] StretchSense MoCap Gloves, https://stretchsense.com/, Accessed March 2024.

Jinuk Heo received his B.S. degree in mechanical engineering from Seoul National University, Seoul, Korea in 2019, where he is currently working toward a Ph.D. degree in mechanical engineering. His research interests include hand tracking, human-robot interaction, and interactive simulation.

Hyelim Choi received her B.S. degree in mechanical engineering from Seoul National University, Seoul, Korea in 2020, where she is currently working toward a Ph.D. degree in mechanical engineering. Her research interests include hand tracking, visual perception, and sensor fusion.

Yongseok Lee received his B.S. degree in mechanical and aerospace engineering and a Ph.D. degree in mechanical engineering from Seoul National University, Seoul, Korea in 2021. He is currently a Staff Engineer with the Samsung Research, Seoul, Korea. His research interests include the hand tracking, human-machine interaction, and generative artificial intelligence.

Hyunsu Kim received his B.S. degree in mechanical engineering and a B.A. degree in psychology from Sungkyunkwan University, Seoul, Korea in 2021. He is currently working toward a Ph.D. degree in mechanical engineering at Seoul National University, Seoul, Korea. His research interests include design of haptic device and haptic simulation.

Harim Ji received his B.S. degree in mechanical engineering from Seoul National University, Seoul, Korea in 2023, where he is currently working toward a Ph.D. degree in mechanical engineering. His research interests include hand tracking, computer graphics, and interactive simulation.

Hyunreal Park received his B.S. degree in mechanical engineering from Seoul National University, Seoul, Korea in 2023, where he is currently working toward a Ph.D. degree in mechanical engineering. His research interests include hand tracking and haptics.

Youngseon Lee received her B.S. degree in energy resources engineering from Seoul National University, Seoul, Korea in 2021, where she is currently working toward a Ph.D. degree in mechanical engineering. Her research interests include dexterous manipulation, control of robotic hands, and haptics.

Cheongkee Jung received his B.S. degree in civil engineering from Korea Military Academy, Seoul, Korea, in 2017. He is currently working toward an M.S. degree in mechanical engineering at Seoul National University, Seoul, Korea. His research interests include human-robot interaction, teleoperation, control of swarm robots, and interactive simulation.

Hai-Nguyen Nguyen obtained his B.Eng. degree in mechatronics and an M.Sc. degree in engineering mechanics from the Hanoi University of Science and Technology, Hanoi, Vietnam, in 2008 and 2011, respectively. He received a Ph.D. degree in mechanical and aerospace engineering from Seoul National University, Seoul, Korea, in 2018. He is currently Senior Researcher with Department of Mechanical Engineering, Seoul National University, Seoul, Korea. His research interests include on the dynamics, control, and planning of mechatronic and robotic systems, with a special emphasis on aerial robotics.

Dongjun Lee received his B.S. degree in mechanical engineering and an M.S. degree in automation and design from the Korea Advanced Institute of Science and Technology, Daejeon, Korea, and a Ph.D. degree in mechanical engineering from the University of Minnesota at Twin Cities, Minneapolis, MN, USA, 2004. He is currently a Professor with the Department of Mechanical Engineering, Seoul National University, Seoul, Korea. His research interests include the dynamics and control of robotic and mechatronic systems with emphasis on aerial/mobile robots, teleoperation/haptics, physics simulation, multirobot systems, and industrial control applications.