A DISSERTATION
SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN COMPUTER SCIENCE AND ENGINEERING

Hand Motion Measurement and Action Recognition with Multimodal Sensor Fusion

by

Chenghong Lu

March 2024

The thesis titled

Hand Motion Measurement and Action Recognition with Multimodal Sensor Fusion

by

Chenghong Lu

is reviewed and approved by:

Chief referee		
Senior Associate Professor	Date	Feb. 20th, 2024
Lei Jing	Lei Jing	
Professor	Date	
Jungpil Shin	Jungpil Shin	Feb. 20. 2024
Senior Associate Professor	Date	
Yoichi Tomioka	Yoichi Tomioka	Feb. 21, 2024
Associate Professor	Date	Feb 20, 2024
Xiang Li	Xiang Li	

THE UNIVERSITY OF AIZU

March 2024

Contents

List of Symbols

Symbol	Description
$\mathbf{q}$	The symbol $\mathbf{q}$ represents quaternion. Quaternions can represent an orientation, a rotation, or a coordinate system change.
G	The global coordinate system,
I	The local coordinate system of the inertial sensor
S	local coordinate system of finger segment.
${}^{\mathrm{G}}\mathbf{q}$	The q containing only the upper left corner mark G represents the quaternion in the G coordinate system.
${}^{\mathrm{S}}_{\mathrm{G}}\mathbf{q}$	Represents coordinate system changes. Refers to the transformation from the coordinate system G of the lower left corner to the coordinate system S of the upper left corner.
$\mathbf{q}_{rl1\,\mathrm{rl2}}$	The lower right corner $rl1$ mark of $\mathbf{q}$ represents whether there is a cumulative error. The lower right corner $\mathrm{rl2}$ it represents a certain moment or rotation during a period of time.
$\mathbf{q}_k$	Quaternion at k moments
$\mathbf{q}_{(0\sim k)}$	Rotation quaternion changes from 0 to k moments
$\mathbf{q}_{\mathrm{r}}$	True-valued quaternion without cumulative error
$\mathbf{q}_{\mathrm{ce}}$	Quaternion containing cumulative error
$\Delta\mathbf{t}$	The sampling interval
$\mathrm{Rot}\,(\mathbf{q})$	the rotation matrix converted from $\mathbf{q}$
$\mathbf{W}$	The value measured by the gyro sensor
$\mathbf{A}$	The value measured by the accelerometer
$\mathbf{M}$	The value measured by the magnetometer
$chol\{\cdot\}$	The Cholesky decomposition of the matrix
$qr\{\cdot\}$	The QR decomposition of the matrix
$F_R^Q(\cdot)$	The conversion of the rotation matrix to the quaternion

Acknowledgment

First of all, I am very grateful to Prof. Lei Jing, for his careful guidance of my graduation thesis, which greatly improved my understanding of academic writing and taught me a lot of specific research skills. I gratefully acknowledge the excellent advice from Prof. Jungpil Shin, Prof. Yoichi Tomioka, and Prof. Xiang Li. Their professional opinions have significantly improved my dissertation. I would like to show my greatest appreciation to Mr. Zeyang Dai for serving valuable advice. Thanks to Wei Guo, Xiaoyang Liu, Haicui Li, Menglei Li, and Zitong Wang, Jiangkun Wang, who are both classmates and friends. We were inspired together and spent an unforgettable research process together. The experience of researching with laboratory members Shingo Amino, Misaki Kozakai, Jiangkun Wang, Hoshi Yuya, and Yuriya Nakamura allowed us to learn from each other and make progress. I would like to thank Mrs. Hoshi for taking care of me in various matters in the laboratory. Finally, thanks to my family for their support.

Abstract

Hand movement measurement and recognition play a vital role in various applications such as human-computer interaction (HCI), virtual reality (VR), augmented reality (AR), robot control, and sign language recognition. With the development of MEMS technology and deep learning, wearable glove solutions and vision-based solutions have been widely studied. Nonetheless, several challenges persist, including the accuracy of wearable sensors, the robustness of vision-based systems under varying environmental conditions, and the effective fusion of multi-modal sensor data.

To enhance wearable IMU data glove accuracy, we propose a multi-IMU calibration method based on hand kinematic constraints. As drift errors from sensors tend to accumulate over time, the limiting relationship in the movement of finger joints is used to obtain a partially corrected drift posture at a certain moment to improve accuracy. We built an IMU data glove, and experimental results show that the proposed system can provide better performance in joint angle accuracy.

Furthermore, we explored the amalgamation of IMU sensors and bend sensors within wearable data gloves for sign language recognition. A weighted Dynamic Time Warping (DTW) algorithm facilitates the fusion of time-series data, assigning differential weights to sensors based on their modality, culminating in enhanced sign language recognition performance.

Lastly, our research delves into the fusion of data from hand-worn data gloves and vision-based systems. While the data glove captures intricate finger curvature metrics, the visual system, employing MediaPipe, extracts commensurate features like finger keypoints and joint angles. The concatenated data undergoes feature fusion via a Convolutional Neural Network-Bidirectional Long Short-Term Memory (CNN-BiLSTM) architecture, leading to improved sign language recognition outcomes. The empirical results from our experiments attest to the potential and efficacy of our multi-modal data fusion approach.

Chapter 1

Introduction

1.1 Background

Hand motion capture and gesture recognition have been fields of active research and development for several decades. They play a crucial role in various applications, such as human-computer interaction (HCI), virtual reality (VR), augmented reality (AR), robotics control, and sign language recognition. Firstly, in the realm of Human-Computer Interaction (HCI), they pave the way for intuitive user interfaces, enabling more natural interactions without the constraints of traditional input devices, a transformation especially evident in virtual and augmented reality systems. Secondly, in the medical sphere, the accurate quantification of hand movements assists clinicians in monitoring and tailoring rehabilitation programs for patients recovering from injuries or surgeries, providing invaluable insights into their progress. The third application of paramount importance is sign language recognition, where automating hand movement detection can foster real-time translation, bridging communication gaps for the hearing-impaired community. Lastly, the precision of hand movement recognition is revolutionizing robotic control, facilitating human guidance in intricate tasks, especially in specialized areas like teleoperation and surgical interventions.

Many complex challenges exist in hand movement measurement and recognition research. Firstly, the inherent complexity of hand movements poses a significant hurdle. The human hand, with its 27 degrees of freedom, exhibits a wide spectrum of intricate gestures, both static and dynamic. This complexity is further compounded by sensor limitations. Devices such as bending sensors and IMUs, though promising, are not immune to issues like drift, noise, and degradation over time. Beyond the inherent attributes of hands and the equipment used, the surrounding environment introduces its own set of challenges. Environmental conditions, notably varying lighting and potential background interferences, can adversely affect recognition accuracy in optical-based systems. Additionally, the occurrence of occlusions, where parts of the hand can obscure other regions, especially in monocular camera setups, impedes seamless data capture. Lastly, as research delves into integrating multiple data sources, like cameras and varied sensors, the resultant complexity demands sophisticated data fusion techniques. These challenges underline the need for approaches that together aspects of hand biomechanics, sensor technology, and advanced computational methods.

1.2 Hand Biological Model

Our goal is to capture the movement of the hand. Therefore, we need a movement model of the hand to describe the relationship between hand segment orientations and positions. In order to reconstruct the hand movement posture, we introduce the calculation method of joint angle based on the proposed model.

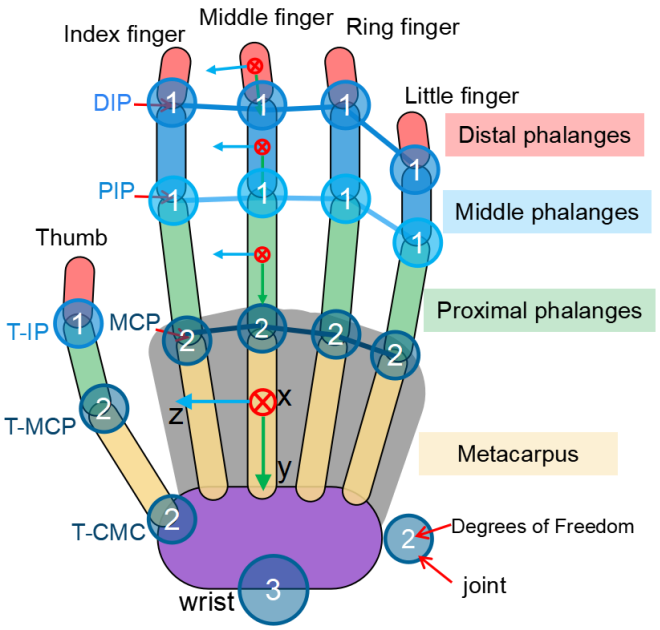

Figure 1.1: Modeled bones and joints of the human hand. The red arrow marks indicate the names of the knuckle and finger segments. Blue circles represent joints, white numbers represent DoF

As shown in Fig. 1.1, each finger consists of three phalanges (segments). Starting at the fingertip, the phalanges are called the distal, middle, and proximal phalanx. The phalanges are connected to 1 DoF by two hinge joints. These joints are called distal and proximal interphalangeal joints (DIP and PIP, respectively). Each finger is connected to the palm by metacarpal phalangeal joints (MCP) with 2 DoF. The palm is linked to the forearm via the wrist joint with 3 DoF.

1.2.1 Definition of Joint Angles

The joint angle is the angle rotated between the orientations of two segments. Specifically, the fingertip is defined as the most distal end and the arm as the most proximal end. A joint angle is the angle of rotation between the orientation of the distal segment and the orientation of the proximal segment. We are particularly interested in the main joint angles (flexion/extension) for a given joint. To describe the relationships between finger segments and joints, we first define a biomechanical hand model and coordinate system.

As Figure 1.1, fingers consist of three phalanges (segments). Starting from the fingertip, the phalanges are called the distal, middle, and proximal phalanx. The phalanges are connected by two hinge joints with one DoF. These joints are named distal and proximal interphalangeal joints (DIP and PIP) respectively. Fingers are connected

to the palm by metacarpal phalangeal joints (MCP) with two DoFs. The palm is linked to the forearm through the wrist joint with three DoFs.

We defined a static flat hand model. In this model, the relationships between the finger segments are as follows. The orientations of the distal phalanx, the middle phalanx, and the proximal phalanx are exactly the same. There is a fixed angle on the abduction/adduction axis between the proximal phalanx and metacarpal phalangeal, and the extension/flexion axis is the same. The direction of the metacarpal and forearm is also the same.

1.3 Hand Motion Measurement

1.3.1 Wearable Sensor

Inertial Measurement Units (IMUs) presenting a compact and efficient alternative to traditional motion capture systems. Comprising accelerometers, gyroscopes, and sometimes magnetometers, IMUs track the linear acceleration, angular velocity, and magnetic field orientation of the device to which they are attached. When positioned on a hand, these sensors capture precise movement data, translating it into orientation, velocity, and positional insights of the hand in three-dimensional space.

Bending sensors are flexible components that detect and quantify changes in their physical curvature, translating them into measurable electrical resistance. As the sensor flexes or bends in tandem with finger or hand movement, its resistance alters, which can then be captured and processed. This makes them particularly suited for tracking the flexion and extension of individual fingers and the hand as a whole.

Electromyography (EMG) is a technique used in medicine and physiology to measure and record the electrical activity generated by muscle tissue. Including muscle contraction, relaxation, and the timing and intensity of muscle activity.

1.3.2 Ambient Sensors

Monocular cameras, often referred to simply as single-lens cameras, capture visual information using a singular lens. Contrary to stereo or multi-lens systems, these cameras rely on one optical input to gather depth and spatial information. While they lack the direct depth perception capabilities inherent to multi-lens configurations, sophisticated algorithms and computer vision techniques have enabled monocular systems to estimate depth and 3D structures from their 2D images.

Optical Systems use markers placed on the hand and fingers, which are tracked by multiple cameras placed around the user. The 3D position of the markers is triangulated from the images captured by the different cameras.

1.4 Sensor Fusion Technology

Sensor Fusion Technology can be divided into Data-level, Feature-level, Decision-level, and Hybrid according to the level classification. Data-level fusion, also known as low-level fusion, involves combining raw sensor data directly at the sensor level. This fusion typically focuses on aligning and synchronizing the sensor data and can include pre-processing steps such as calibration, time-stamping, and coordinate alignment. It

aims to create a unified representation of the raw sensor measurements before further processing or fusion at higher levels. Feature-level fusion involves extracting relevant features from individual sensor data and then combining these extracted features from multiple sensors. The features can be domain-specific, such as edges, textures, or key points in computer vision, or derived from signal processing techniques in other domains. Feature-level fusion aims to capture and represent the salient information from each sensor, which can be used for subsequent processing and decision-making. Decision-level fusion, also known as high-level fusion, involves combining the decisions or results obtained from individual sensors. Instead of directly fusing the raw data or features, decision-level fusion focuses on combining the outcomes or decisions made by each sensor or subsystem. This level of fusion can include voting mechanisms, rule-based systems, or more advanced techniques such as Dempster-Shafer theory or fuzzy logic. The goal is to integrate multiple sensor inputs to make more informed and reliable decisions.

Hybrid fusion refers to the combination of multiple levels of fusion to achieve more comprehensive and robust fusion results. It involves integrating data, features, and decisions from multiple sensors at different levels of abstraction. Hybrid fusion algorithms can leverage the strengths of each level to address different aspects of the fusion problem. For example, combining data-level fusion for accurate alignment and synchronization, feature-level fusion for informative feature extraction, and decision-level fusion for final decision-making.

Sensor fusion technology exploits complementarity as the basis for good performance. Redundancy improves robustness, and decision conflicts are generally issues that need to be resolved.

Sensor fusion leverages complementarity to improve overall understanding. By combining complementary data, fusion algorithms can provide a more comprehensive and accurate representation of the phenomenon being observed. The fusion process takes advantage of the unique strengths of each sensor to enhance the final decision. Redundant sensors can be used in sensor fusion to increase reliability and fault tolerance. When redundant sensors produce consistent data, it boosts confidence in the decision. However, during decision conflicts or when sensor failures occur, fusion algorithms must be designed to handle redundancy properly. This might involve weighting sensors based on reliability or choosing the most trustworthy sensor for the current situation. When data from different sensors conflict or exhibit uncertainty, fusion algorithms play a critical role in resolving these conflicts. The fusion process may involve statistical methods, expert rules, or machine learning techniques to make a decision that best aligns with the available data. Handling decision conflicts effectively is essential for maintaining the accuracy and reliability of the sensor fusion system.

1.4.1 Kalman Filter

The Kalman algorithm, specifically the Kalman filter, is a recursive estimation algorithm widely used for state estimation in linear dynamic systems. It provides a solution for estimating the state of a system given noisy measurements by combining predictions from a mathematical model of the system with real-time sensor measurements.

1.4.2 Deep Neural Networks

Deep learning, on the other hand, is a subset of machine learning that focuses on training artificial neural networks with multiple layers to learn and extract features directly from data. It is particularly effective in handling complex, non-linear problems and has achieved significant success in various domains, including computer vision, natural language processing, and speech recognition.

Researchers have explored using deep learning models to learn the dynamic model or the measurement model in a Kalman filter to handle non-linear or complex systems. Deep learning can be employed to learn the system dynamics, update the state estimate, or refine the measurement noise parameters in a data-driven manner.

Deep Learning for End-to-End Estimation: Deep learning models can be employed as end-to-end estimators, bypassing the explicit use of a Kalman filter. By training deep neural networks to directly map sensor measurements to the desired estimation outputs, the need for a separate Kalman filtering step can be eliminated. This approach is particularly useful in scenarios where the underlying system dynamics are highly non-linear and difficult to model explicitly.

The relationship between the Kalman algorithm and deep learning can involve integrating their respective strengths or using deep learning as an alternative approach to estimation and prediction tasks, depending on the characteristics of the problem and the available data.

1.5 Thesis Organization

The relationship between the chapters of the dissertation is shown in the figure, and the organization is as follows.

First, the research background is introduced. In the first chapter, the basic concepts of hand movement flow measurement and recognition are introduced, such as biological models of hands, wearable data gloves, sensor fusion, etc. In addition, this chapter explains the organization and main contributions of the paper. Secondly, related work is described in Chapter 2. In order to improve the measurement accuracy of hand movements, Chapter 3 proposes a calibration method based on joint kinematic constraints for IMU-based data gloves. For multi-sensor fusion in IMU, Kalman algorithm is used for fusion. The system performance was tested in terms of joint angle accuracy and stability. The fusion of various wearable sensors, IMUs and curved sensors is presented in Chapter 4, with specific applications of sensor data fusion in sign language recognition. Weighted DTW is used for action recognition. The application of the fusion of wearable devices and visual systems in sign language recognition is in Chapter 5, Multimodal Data Fusion System in Hand Action Recognition. Improve sign language recognition rate and improve stability in complex environments. Finally, Chapter 6 is the conclusion of this paper.

1.6 Contributions

This dissertation is a study on hand movement measurement and hand posture recognition, including the establishment of a wearable data glove measurement system

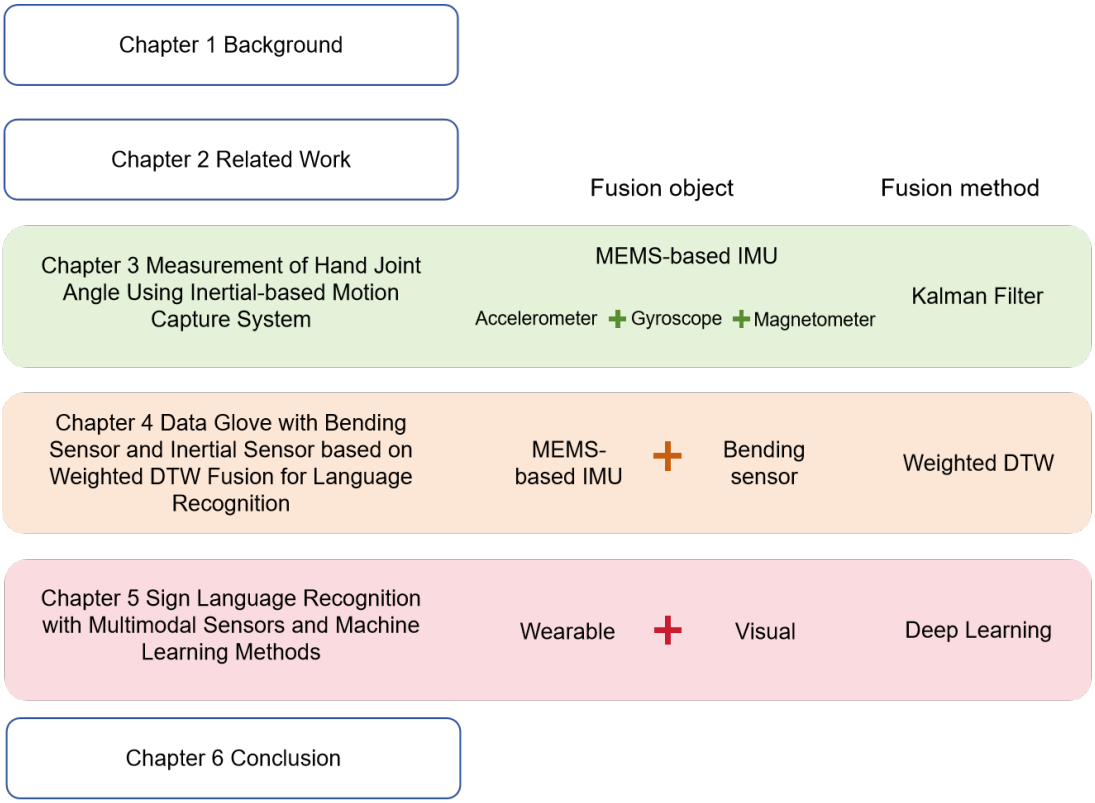

Figure 1.2: Thesis Structure

and the research on sign language recognition using the fusion of multi-modal sensors. The main contributions of each chapter are summarized as follows.

- An IMU-based wearable data glove has been developed for hand movement measurement. By using the kinematic constraints of hand joints for calibration, the inherent sensor drift issue is effectively mitigated, resulting in enhanced accuracy in joint angle measurements. Introduced in Chapter 3.

- For sign language recognition scenarios, a data glove equipped with inertial and bending sensors has been introduced to capture hand shapes and movements. With the challenge of multi-modal sensor time series data, a weighted Dynamic Time Warping (DTW) approach has been employed for effective data fusion. This distribution of weights enhances the importance of crucial sensors and improves recognition accuracy. Introduced in Chapter 4.

- We build a system combining vision with bending sensor data to recognize sign language. By integrating features such as finger keypoint coordinates, finger joint angles, and finger flexion angles, we create a rich feature set. This composite data is subsequently processed through a CNN-BiLSTM architecture, not only bolstering sign language recognition rates but also fortifying system robustness against challenges like visual occlusions. Introduced in Chapter 5.

1.7 Publications

The following papers have been published or submitted to major journals and conferences.

(1) Major journal paper

paper accepted

1. **Chenghong Lu**, Zeyang Dai and Lei Jing. "Measurement of Hand Joint Angle Using Inertial-Based Motion Capture System." IEEE Transactions on Instrumentation and Measurement 72 (2023): 1-11.

2. **Chenghong Lu**, Shingo Amino, and Lei Jing. 2023. "Data Glove with Bending Sensor and Inertial Sensor Based on Weighted DTW Fusion for Sign Language Recognition" Electronics, vol. 12. no. 3, 2023.

3. **Chenghong Lu**, Misaki Kozakai and Lei Jing. 2023."Sign Language Recognition with Multimodal Sensors and Deep Learning Methods" Electronics 12, no. 23: 4827.

(2) Major conference paper

1. **Chenghong Lu**, Jiangkun Wang and Lei Jing. "Hand motion capture system based on multiple inertial sensors: demo abstract." Proceedings of the 18th Conference on Embedded Networked Sensor Systems (2020)

2. Dai, Zeyang, **Chenghong Lu** and Lei Jing. "Time Drift Compensation Method on Multiple Wireless Motion Capture Nodes." 2020 13th International Conference on Human System Interaction (HSI) (2020): 266-271.

3. Hoshi, Yuya, **Chenghong Lu** and Lei Jing. "Haptic Finger Glove for the VR Keyboard Input." Interacción (2021).

4. Jiangkun Wang, Tsubasa Endo, **Chenghong Lu** and Lei Jing, "A Novel AR Whiteboard System and Usability Study," 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE19), Osaka, Japan, 2019, pp. 28-30 [IEEE]

5. Chenghong Lu, Yuriya Nakamura, Lei Jing, "Markerless Cat Life Logging System Using Skeleton Data and ST-GCN Method", The 12th International Conference on Awareness Science and Technology (iCAST23), Taichung, Taiwan, Nov., 2023 [IEEE]

Chapter 2

Related Work

There are many researches on motion measurement and recognition, such as systems for upper limbs (e.g., [1], [2]), lower limbs (e.g., [3], [4]), and hand (e.g., [5], [6], [7]). The joint angles of the hand and body can be calculated by the same processing. However, the fingers have three segment linkage structure, and the palm of the hand connects five fingers, presenting a more complex structure. Furthermore, the finger segment is a much smaller cylinder with more lack of a flat surface. To deal with the complex and compact structure, as well as the increased measurement error, We need a variety of sensors to collect hand information collaboratively, with appropriate calibration methods and sensor fusion methods.

2.1 Wearable Sensor

IMU was only used in aircraft navigation and large-scale equipment before, due to limitations in size, cost and power consumption [8]. With the development of micromachining technology. microsensors were developed for measuring physical(e.g., Angular velocity, Acceleration, Magnetism, strain, radiation and flow). Therefore, MEMS IMU sensors have very attractive features of low cost, compact, fast responsibility, and totally sourceless, making them attractive for determining the motion of small moving objects [9]. Wearable IMUs have been widely used in various domains, including virtual reality and real-time tracking of human action in the three-dimensional (3D) space [10].

Bending sensors [11], [12] are flexible components that detect and quantify changes in their physical curvature, translating them into measurable electrical resistance. As the sensor flexes or bends in tandem with finger or hand movement, its resistance alters, which can then be captured and processed. This makes them particularly suited for tracking the flexion and extension of individual fingers and the hand as a whole.

The EMG sensor performs hand movement recognition, and the sensor is generally placed on the forearm muscles. In order to recognize complex gestures and analyze multiple EMG sensor data, deep learning methods (e.g., [13], [14], [15], [16], [17], [18]) are extensively used.

The study[19] introduces a novel real-time hand gesture recognition method using sEMG to decode motor unit activities for various motor tasks. The method involves segmenting EMG signals into motion-related segments and applying a convolution kernel compensation algorithm for real-time global EMG decomposition. This technique was tested on high-density EMG data from eleven non-disabled participants performing twelve hand gestures. However, the reliance on high-density EMG data could also limit

its practicality in everyday applications due to the need for specialized equipment and setup.

2.2 Calibration

MIMU-based calibration has been extensively studied [20], [21]. Without limiting to MIMU only, some studies use additional sensors to provide augmented data [22], [23]. Optical systems with markers are commonly used to obtain the precision orientation of each segment. However, at the same time, additional sensors such as optical systems have the disadvantage of being complex to operate and limiting the scenarios.

StoS calibration with MIMU can be categorized as assumed alignment methods, static pose methods, and functional methods. The assumed alignment methods require visually alignment of the MIMU sensor axes with the underlying anatomical axes [24]. This is the most intuitive method to minimize errors by coordinating visual and manual manipulation. However, due to the three-dimensional starting point and the fuzzy initial direction of the finger segment, it is difficult for non-experts to ensure the accuracy and repeatability of the operation. In the static pose method, the hand motion is adjusted to align with the static poses [25], [26]. Finger segment orientation is determined based on the static action with known orientation. The functional method is calibrated by performing a calibration action to obtain the joint rotation axis [27]. First, a prescribed joint rotation motion is executed to obtain the rotation axis. Then the vertical axis is obtained by the gravity component and finally the remaining one axis is obtained by cross multiplication. The body segment orientations are estimated using a joint kinematic model [28], [29]. The joints are calibrated by unspecified movements. For a joint with 1 Degree of Freedom (DoF), we can perform the calibration in the process. However, for multi-DoF joints, it is difficult for the natural motion to rotate in only one dimension.

2.3 Sensor Fusion Technology

Sensor fusion, the amalgamation of data from diverse sensors, seeks to enhance decision-making and inferencing by leveraging the strengths and compensating for the weaknesses of individual sensing modalities. In this domain, two methodologies have particularly distinguished themselves due to their efficacy in handling complex real-world data: Kalman filtering and deep learning. While the Kalman filter and deep learning emerge from distinct paradigms the former grounded in control theory and the latter in artificial intelligence both have found profound applications in sensor fusion.

Looking back at the last decades, the vast majority of published finger segment orientation estimates for inertial sensors, the methods for sensor fusion can be divided into two categories: complementary filters and extended Kalman algorithms [27], [30].

Several research papers (e.g.,[31], [32]) have used the Madgwick algorithm [33], which belongs to the complementary filter. This algorithm uses a gradient descent algorithm to limit the magnetometer to eliminate drift in only the azimuthal part of the orientation. Other research (e.g.,[25], [26]) used the method of Thomas Seel et al.[34] This method uses an analytical solution to restrict the magnetometer to affect only the azimuthal part. Both of these complementary filters have constant gain. Their typical

drawback is that the gain is always given empirically only once and is poorly extendable to different scenarios.

The extended Kalman algorithm is based on the Kalman algorithm [35] for non-linear orientation estimation. Since the Kalman filter assumes that both the state and sensor measurements are Gaussian, the particle filter may be superior to the Kalman filter when part of the system is non-Gaussian. However, particle filters are computationally costly with a low potential for migrating to embedded systems. To improve the system performance, it is necessary to compare the EKF [36] with the sigma-point Kalman filter (SPKF)[37], [38]. Unscented Kalman filter (UKF) [39], and cubature Kalman filter (CKF) [40], [41] are both SPKF. The EKF is the first-order accuracy of the nonlinear functions. It is shown that a UKF performs much better than EKF, but its run time is longer [42]. The CKF employs a third-degree spherical-radical cubature rule and requires only fewer cubature points. Theoretically, CKF has better computational speed than UKF. CKF can lead to numerical instability during implementation, causing the filter divergence [40]. In combination with square root, SRCKF all resulting covariance matrices are guaranteed to remain positive semi-definite and can solve this problem [43].

In contrast, deep learning is a subset of machine learning, leveraging neural networks with many layers to learn intricate patterns from vast amounts of data. Over the past decade, deep learning has revolutionized fields like computer vision and sensor fusion. When traditional model-based approaches, such as the Kalman filter, might falter due to the non-linear and complex nature of the environment or the sheer dimensionality of the data, deep learning can shine. Deep neural architectures can automatically extract salient features from multi-sensor data and learn to integrate these features in ways that are often challenging to achieve with traditional algorithms. In the realm of sensor fusion, deep learning can be utilized to fuse data from diverse sensors, handling non-linearities, and capturing intricate relationships, often leading to improved performance in challenging scenarios.

Multimodal sensor data fusion methods are crucial in systems that combine curved sensors and vision. CNN [44] and BiLSTM [45] methods, which can obtain information from spatial and time series data. Fusion of CNN and BiLSTM [46], [47] has been used in the field of Natural language processing. Also, The skeleton of the hand using a method called MediaPipe [48] from videos. In addition, by using the sensor, we can expect to measure the angle of the finger more accurately even in the part that overlaps other objects. Therefore, combining sensor data with sign language recognition will make it possible to accurately predict hand movements.

2.4 Sign Language Recognition

In recent years, the evolution of wearable hand measurement devices has been evident, predominantly driven by miniaturization processes and advancements in algorithms. Notably, data gloves [49], [50], including IMU [51] and bending sensors [52] [53], have demonstrated significant advancements in wearability, accuracy, and stability metrics. Such advancements have consequently led to marked enhancements in the results of sign language recognition leveraging these measurement apparatuses.

There are many studies on sign language recognition solutions based on computer vision [54], [55], With the evolution of deep learning algorithms, the extraction and

analysis of features from visual data, including bone key point prediction [56], have substantially improved. While sign language recognition has experienced significant advancements, occlusions in images remain a notable challenge in computer vision. Himanshu and Sonia's review discusses the effects of occlusion on the visual system [57]. There are ways to avoid occlusion problems by using a depth camera, multiple cameras, or labeling invisible objects. There are also methods to detect occlusion, such as using shadows of objects and learning information before and after occlusion using time series data.

Therefore, the complementary information of the bending sensor system and the vision system is used to improve accuracy and stability.

Himanshu and Sonia present a review on occlusion [57]. There are ways to avoid occlusion problems by using a depth camera, multiple cameras, or labeling invisible objects. There are also methods to detect occlusion, such as using shadows of objects and learning information before and after occlusion using time series data.

Avola et al. [58] uses SHREC [8] for the dataset to perform sign language recognition. SHREC is a dataset that uses a depth camera to acquire gesture skeletons. DLSTM, a deep LSTM, is used for sign language recognition. SHREC is used and the angles formed by the fingers of the human hand are used as features. From the predicted skeleton, the finger angles are calculated and used as features. The training using SHREC and DLSTM enables highly accurate sign language recognition.

Liuhao [59] et al. explained the prediction of the skeleton of the hand from image recognition. It estimates the complete 3D hand shape and poses from a monocular RGB image, rather than a depth camera. It uses the original graph convolutional neural network for training. In some cases in this research, recognition accuracy is reduced due to blind spot problems.

Although motion capture using a special device such as Kinect [60] and Leap Motion Controller (LMC) [61] exist, sign language recognition using a monocular camera is superior in that can use a common camera. In addition, there are limitations in acquiring spatial information with images captured by a monocular camera. This is the case for blind spot problems or when spatial information does not appear in the image. By using MediaPipe, information from the camera can be acquired, and with the aid of sensors, accurate spatial information can be acquired for more accurate sign language recognition.

Chapter 3

Hand motion measurement

3.1 Introduction

With recent advances in Micro-Electro-Mechanical Systems techniques, there has been an increased focus on sensor-based motion capture with magnetic and inertial measurement units (MIMU) in recent years [62], [63]. In hand rehabilitation, quantifiable measurements help clinicians make more objective diagnoses [64],[65],[5]. In virtual reality (VR), MIMU-based hand motion capture enables users to interact naturally with digital objects [66].

However, the accuracy is still not comparable with optical systems, which is the de facto standard in motion capture. The interior accuracy of inertial-based methods has limited the widespread applications of hand motion capture. For this reason, this paper proposes a method for estimating finger joint angles. We not only implement an advanced sensor fusion algorithm to estimate the orientation, but also propose a novel sensor-to-segment (StoS) calibration method.

The measurement of the joint angle is particularly critical for hand motion capture. The hand consists of a chained multi-joint structure, so the posture and position of the hand can be reconstructed from joint angles and finger segment lengths using forward kinematics. Therefore, the accuracy of hand motion capture is largely dependent on the accuracy of the joint angle measurement.

In recent years, several studies have been conducted to improve the accuracy of joint angle measurement by clinical inertial motion capture systems. The most general method to estimate joint angle based on the inertial sensor is to attach the MIMU sensors on two linked finger segments, as shown in Fig. 3.1. The data is then processed in three steps. First, the sensor orientation is determined from the MIMU sensor data using a fusion algorithm. Second, the orientation of the finger segment is aligned according to the sensor orientation. Third, the joint angle is obtained by calculating the difference between the finger segment orientations. This is analogous to measuring the length of a desk with a ruler without a scale. The first step is to determine the scale of the ruler, the second step is to align the ruler to the two edges of the desk, and the third step is to calculate the difference between the two lengths. Measurement errors can occur during both sensor fusion and alignment, as if the scale of the ruler were to shorten or expand and the edges are not aligned.

The data fusion algorithm compensates for the drift of the gyroscope due to the accelerometer and magnetometer. The extended Kalman filter (EKF) is a common approach to solve the orientation estimation problem. However, it is an approximation

process for the linearization of the observation model in the EKF, which introduces relatively large residuals in this nonlinear problem in the presence of external disturbances. The Square Root Cubature Kalman Filter (SRCKF) is one of the sigma-point Kalman filters (SPKFs) [37] that have the ability to obtain higher-order accuracy in solving nonlinear problems. To the best of our knowledge, this is the first time that the SRCKF has been used in the orientation estimation of hand motion.

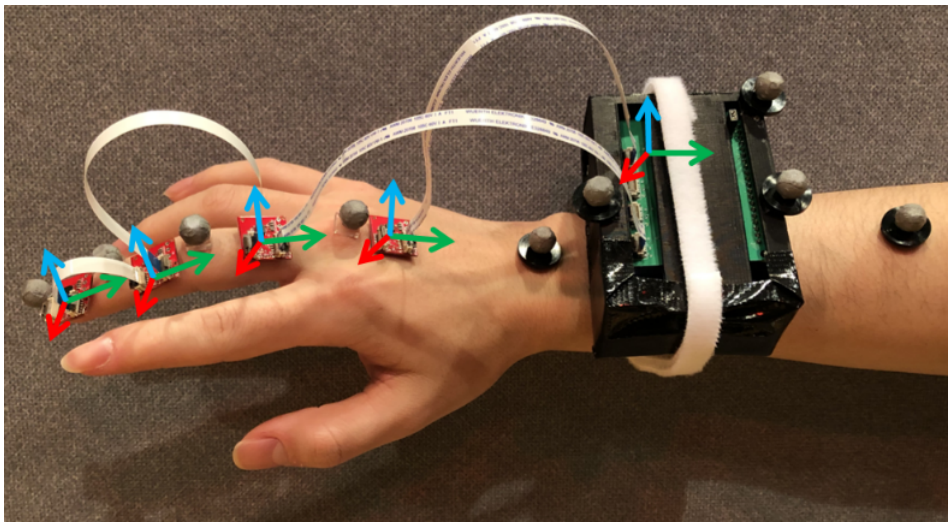

Figure 3.1: MIMU data glove. Every segment contains a gyroscope, accelerometer and magnetometer which are connected using flexible flat cables.

StoS calibration is used to resolve the misalignment between the sensor orientation and the actual segment orientation. The misalignment is divided into two parts: on the one hand, the initial misalignment due to the inability to directly observe the skeletal orientation when the sensor is mounted on the soft tissue skin. On the other hand, in-process misalignment increases due to cumulative sensor drift over time. The three common methods are assumed alignment, static pose, and functional methods. The assumed alignment method uses visual assessment and manual alignment. The static pose method aligns the sensor orientation to the known orientation of the finger segment in the static pose. This method is poorly interpreted for joint rotations. The lack of consideration of the rotation axis also limits our further use of joint kinematics to optimize the calibration step. The functional method uses calibration actions to align the rotation axis. In the case of the upper limb, which has similarity to the hand, there is no significant difference in accuracy between the three alignment methods [2]. An important aspect is that the three method lacks calibration for accumulated errors in the process.

The contributions of this Chapter:

1. A hand measurement system based on inertial sensors was established to measure joint angles.

2. A StoS calibration method based on hand joint kinematic constraints is developed. The method provides anatomical interpretation and accurate joint angles, by calibrating the joint rotation axis and performing in-process calibration.

3. SRCKF is used for the first time in MIMU-based hand motion capture. We establish the state equation and adjust the noise parameters to achieve high accuracy orientation.

The rest of this paper is organized as follows. Related work is briefly reviewed for inertial sensor hand motion capture. The SRCKF and StoS calibration methods are put forward in Section 3.2, and the implementation process is also described in more detail in this section. Section 3.3 shows MIMU-based data glove design. Section 3.4 shows the experimental design and results. In Section 3.5 our system is compared with other systems and the system results are discussed.

3.2 Methods

In this section, we describe how to obtain the hand joint angles as shown in Fig. 3.2, focusing on the orientation of the MIMU sensor obtained by implementing SRCKF fusion and the method of StoS calibration based on joint kinematic constraints.

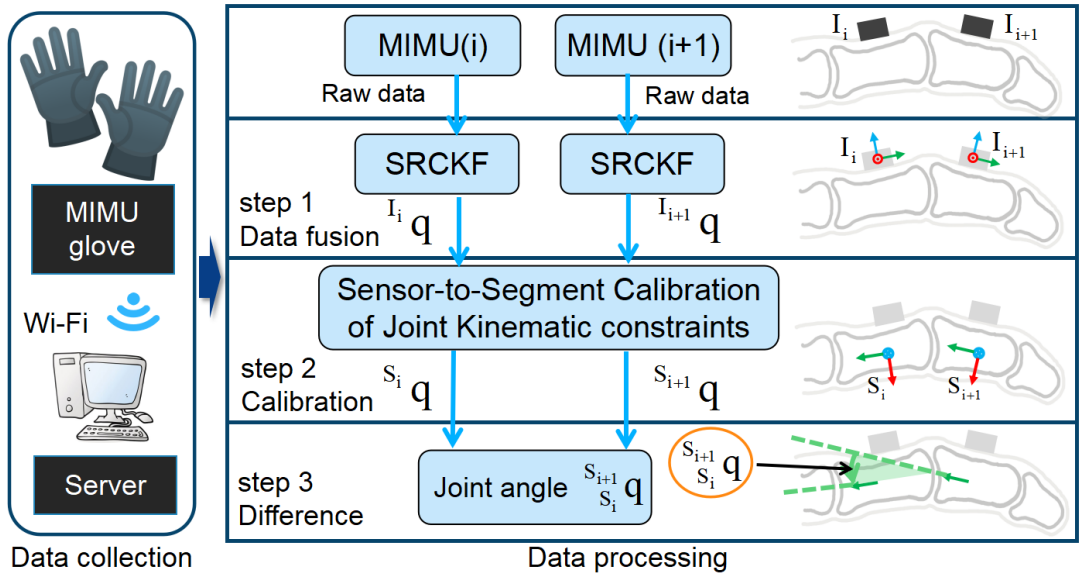

Figure 3.2: Hand motion capture system to measure joint angles (Raw data: gyroscope, accelerometer, magnetometer, ${}^{\mathrm{I}}\mathbf{q}$: inertial sensor orientation, ${}^{\mathrm{S}}q$: finger segment orientation)

3.2.1 Definition of Joint Angles

In the representation of the measurement of hand joint angles from the MIMU system, a global coordinate system and two types of local coordinate systems are defined (see Fig. 3.3). The Earth-based global coordinate system G in the North-East-Down (NED) frame is defined with gravity and magnetic north reference vectors. So it should be common for all MIMUs but is actually different and time-varying. [67]. Each MIMU i has a local coordinate frame I_i. Each hand segment i attached to the MIMU is a local coordinate system S_i. The coordinate system of hand segments is mostly based on the definition of the International Society of Biomechanics (ISB) [68]. The difference is that the functional axis is defined as the z-axis because it contributes to the interpretation of joint motion [69].

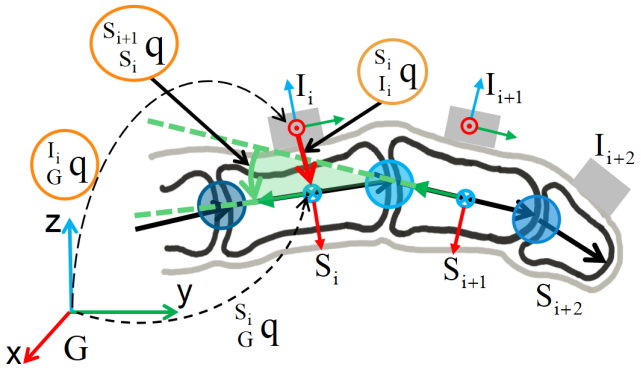

Figure 3.3: The definition of coordinate system (global coordinate system G, MIMU coordinate system I_i, segment coordinate system S_i) and joint angle (${}^{\mathrm{S}_{i+1}}_{\mathrm{S}_i}\mathbf{q}$).

$$ {}^{\mathrm{S}_i}\mathbf{q} = {}^{\mathrm{S}_i}_{\mathrm{I}_i}\mathbf{q} \otimes {}^{\mathrm{I}_i}\mathbf{q} \tag{3.1} $$

$$ {}^{\mathrm{S}_i+1}\mathbf{q} = {}^{\mathrm{S}_i+1}_{\mathrm{I}_i+1}\mathbf{q} \otimes {}^{\mathrm{I}_i+1}\mathbf{q} \tag{3.2} $$

$$ {}^{\mathrm{S}_i+1}_{\mathrm{S}_i}\mathbf{q} = {}^{\mathrm{S}_i}\mathbf{q} \otimes {}^{\mathrm{S}_i+1}\mathbf{q}^* \tag{3.3} $$

The rotation relationship between coordinate systems is defined by the unit quaternion. The sensor orientation ${}^{\mathrm{I}}_{\mathrm{G}}\mathbf{q}$ is first obtained by the sensor fusion algorithm (see section 3.2.2), after that the segment orientation ${}^{\mathrm{S}}_{\mathrm{G}}\mathbf{q}$ is obtained by the StoS alignment ${}^{\mathrm{S}}_{\mathrm{I}}\mathbf{q}$ (see section 3.2.3), and finally the joint angle ${}^{\mathrm{S}_{i+1}}_{\mathrm{S}_i}\mathbf{q}$ is obtained from the orientation of the adjacent segments. More details about quaternions for rotation can be found in [70].

3.2.2 Sensor Fusion Algorithm

We use SRCKF as the fusion algorithm of Attitude Estimation [37]. Fusion of magnetism, angular velocity and acceleration obtained from 9-Axis MIMU sensor (ICM20948) to obtain precision attitude estimation.

Process and measurement models

Quaternion is used as the attitude representation in this filter because of the computational simplicity and the gimbal lock avoided. A quaternion measurement algorithm based on an accelerometer and magnetometer is used as the observation model and the gyroscope quaternion equation of motion is used as the process model to build the Kalman filter.

$$ {}^{I}\mathbf{q}_k = \left\{ \mathbf{I}_{4\times4} + \frac{\Delta t}{2}[\mathbf{\Omega}\times] \right\} {}^{I}\mathbf{q}_{k-1} \tag{3.4} $$

where Δt denotes the sampling interval and $[\mathbf{\Omega}\times]$ is determined with the gyroscope output ${}^{I}\mathbf{W} = (\omega_x, \omega_y, \omega_z)^T$ in sensor frame s.

$$ [\mathbf{\Omega}\times] = \begin{pmatrix} 0 & -\omega_x & -\omega_y & -\omega_z \\ \omega_x & 0 & \omega_z & -\omega_y \\ \omega_y & -\omega_z & 0 & \omega_x \\ \omega_z & \omega_y & -\omega_x & 0 \end{pmatrix} \tag{3.5} $$

The observation model in quaternion form as [71].

The normalized acceleration and magnetic vectors ${}^{I}\mathbf{A} = (a_x, a_y, a_z)^T$ and ${}^{I}\mathbf{M} = (m_x, m_y, m_z)^T$ are read from the inertial sensor in the sensor frame I. The measurement model is formulated as

$$\mathbf{z_k} = \left[\begin{array}{cc} {}^{I}\mathbf{A} & {}^{I}\mathbf{M} \end{array} \right]^T = \mathrm{Rot}\left({}^{I}\mathbf{q_k}\right) \left[\begin{array}{cc} {}^{G}\mathbf{A} & {}^{G}\mathbf{M} \end{array} \right]^T \tag{3.6}$$

$\mathrm{Rot}({}^{I}\mathbf{q_k})$ is the rotation matrix converted from ${}^{I}\mathbf{q_k}$.

The filter uses a state space model of the attitude estimation for dynamics q_k and measurements z_k at time k. $w_k \sim N(0, Q_k)$ and $v_k \sim N(0, R_k)$

$$\left\{ \begin{array}{l} {}^{I}\mathbf{q}_k = \left\{\mathbf{I}_{4\times4} + \frac{\Delta T}{2}[\mathbf{\Omega}\times]\right\} {}^{I}\mathbf{q}_{k-1} + w_k \\ \mathbf{z}_k = \mathbf{H}_k {}^{I}\mathbf{q}_k + v_k \end{array} \right. \tag{3.7}$$

Square root cubature Kalman filter

In this subsection, the square root cubature Kalman filter for nonlinear systems is derived. The cubature Kalman filter employs a third-degree spherical-radical cubature rule to compute Gaussian-weighted integrals to obtain excellent performance. The square-root filters improve the numerical stability because all the covariance matrices are guaranteed to stay positive semi-definite.

$$\left\{ \begin{array}{l} x_k = f(x_{k-1}) + w_{k-1} \\ z_k = h(x_k) + v_{k-1} \end{array} \right. \tag{3.8}$$

Initialization the filter, initiate state $\hat{x}_0$ and initiate square root of covariance matrix S_0.

$$\begin{array}{l} \hat{x}_0 = E[x_0] \\ S_0 = \mathrm{chol}\left\{ \left[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T \right] \right\} \end{array} \tag{3.9}$$

$chol\{\cdot\}$ denotes the Cholesky decomposition of the matrix. Time update :

Generate cubature point $\chi_{\mathbf{i}}$.

$$\chi_{\mathbf{i,k-1}} = \left\{ \begin{array}{ll} \hat{\mathbf{x}}_{\mathbf{k-1}} + \sqrt{L}\left(\mathbf{S_k}\right)_i, & i = 1, \ldots, L \\ \hat{\mathbf{x}}_{\mathbf{k-1}} - \sqrt{L}\left(\mathbf{S_k}\right)_i, & i = L+1, \ldots, 2L \end{array} \right. \tag{3.10}$$

$$\omega_i^{(m)} = \omega_i^{(c)} = \frac{1}{2L}, i = 1, \ldots, 2L$$

Propagated cubature points

$$\begin{array}{c} \chi_{\mathbf{k|k-1}} = \mathbf{f}\left(\chi_{\mathbf{k-1}}\right) \\ \hat{\mathbf{x}}_{\mathbf{k}}^{-} = \sum_{i=1}^{2L} \omega_i^{(m)} \chi_{\mathbf{i,k|k-1}}, \end{array} \tag{3.11}$$

The squared-root factor of the predicted error,

$$\mathbf{S}_{\mathbf{k}}^{-} = qr\left\{ \sqrt{\omega_i^{(c)}} \left(\chi_{L,\mathbf{k|k-1}} - \hat{\mathbf{x}}_{\mathbf{k}}^{-}\right), \mathbf{S}_{Q,k-1} \right\} \tag{3.12}$$

$qr\{\cdot\}$ denotes the QR decomposition of the matrix, $\mathbf{S}_Q$ denoted a square-root factor

of Q_k.

Measurement update:

Propagated cubature points

$$\begin{aligned} \mathcal{Z}_{k|k-1} &= \mathbf{H}\left(\chi_{\mathbf{k}|\mathbf{k-1}}\right) \\ \hat{\mathbf{z}}_{\mathbf{k}}^{-} &= \sum_{i=1}^{2L} \omega_i^{(m)} \mathcal{Z}_{i,k|k-1} \end{aligned} \tag{3.13}$$

Calculate the Square-root of the QR decomposition

$$\mathbf{S}_{\hat{\mathbf{z}}_{\mathbf{k}}} = qr\left\{\sqrt{\omega_i^{(c)}}\left(\mathcal{Z}_{L,k|k-1} - \hat{\mathbf{z}}_{\mathbf{k}}^{-}\right), \mathbf{S}_{R,k-1}\right\} \tag{3.14}$$

Calculate the cross-covariance matrix

$$\mathbf{I}_{\mathbf{x_k z_k}} = \sum_{i=0}^{2L} \omega_i^{(c)}\left[\chi_{\mathbf{i,k|k-1}} - \hat{\mathbf{x}}_{\mathbf{k}}^{-}\right]\left[\mathcal{Z}_{i,k|k-1} - \hat{\mathbf{z}}_{\mathbf{k}}^{-}\right]^T \tag{3.15}$$

Calculate the Kalman gain

$$\mathbf{K} = \left(\mathbf{P}_{\mathbf{x_k z_k}} / \mathbf{S}_{\hat{\mathbf{z}}_{\mathbf{k}}}^{\mathbf{T}}\right) / \mathbf{S}_{\hat{\mathbf{z}}_{\mathbf{k}}} \tag{3.16}$$

Calculate the updated state

$$\hat{\mathbf{x}}_{\mathbf{k}} = \hat{\mathbf{x}}_{\mathbf{k}}^{-} + \mathbf{K}\left(\mathbf{z}_{\mathbf{k}} - \mathbf{z}_{\mathbf{k}}^{-}\right) \tag{3.17}$$

Calculate the square-root factor of the error covariance

$$\begin{aligned} \xi &= \sqrt{\omega_i^{(c)}}\left\{\left(\chi_{L,\mathbf{k}|\mathbf{k}-1} - \hat{\mathbf{x}}_{\mathbf{k}}^{-}\right) - \mathbf{K}\left(\mathcal{Z}_{L,k|k-1} - \hat{\mathbf{z}}_{\mathbf{k}}^{-}\right)\right\} \\ \mathbf{S}_{\mathbf{k}} &= qr\left\{\xi, \mathbf{S}_{R,k-1}\right\} \end{aligned} \tag{3.18}$$

3.2.3 Sensor to Segment Calibration Method

StoS calibration is the method to obtain the rotational quaternion ${}^{\mathrm{S}}_{\mathrm{I}}\mathbf{q}$ between the MIMU sensor orientation ${}^{\mathrm{I}}_{\mathrm{G}}\mathbf{q}$ and the hand segment orientation ${}^{\mathrm{S}}_{\mathrm{G}}\mathbf{q}$. Without proper calibration, deviations from the initial value will continuously introduce errors in the measurement process. Calibration is able to convert the measured values into anatomically interpretable data, such as joint angles of flexion/extension, and abduction/adduction.

$${}^{\mathrm{S}}_{\mathrm{G}}\mathbf{q} = {}^{\mathrm{I}}_{\mathrm{G}}\mathbf{q} \otimes {}^{\mathrm{S}}_{\mathrm{I}}\mathbf{q} \tag{3.19}$$

Initial Calibration

The sensor is fixed to the finger segment, so the relationship between the sensor coordinate system and the segment coordinate system is approximately fixed.

The conversion of the quaternion to the axis angle $F_Q^A(\cdot)$.

$$\Delta\mathbf{q} = {}^{\mathrm{I}}_{\mathrm{G}}\mathbf{q}_{t+n} \otimes {}^{\mathrm{I}}_{\mathrm{G}}\mathbf{q}_t^* \tag{3.20}$$

$$(\mathbf{v_z}, \theta) = F_Q^A(\Delta\mathbf{q}) \tag{3.21}$$

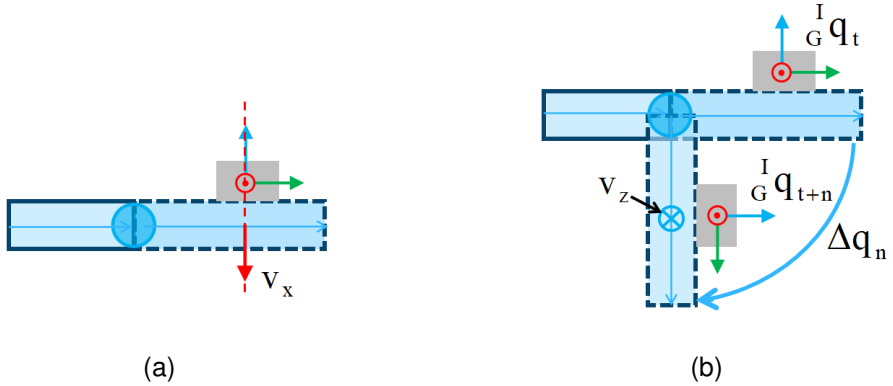

Figure 3.4: Joint kinematic constraints method for establishing the segment coordinate system S. (a) Hand is placed flat on the table and the gravity is measured to get $\mathbf{v_x}$. (b) Rotation axis obtained from the joint rotation action as $\mathbf{v_z}$.

In the joint kinematic constraints method, we add joint rotation axis constraints, which can be obtained by functional actions [27], or by using Gyroscope-Based Kinematic Constraint [72]. As shown in Fig. 3.4 to obtain $\mathbf{v_x}, \mathbf{v_z}$.

$$\mathbf{v_y} = \mathbf{v_x} \times \mathbf{v_z} \tag{3.22}$$

After that the procedure to calculate the rotational quaternion ${}^{\mathrm{S}}_{\mathrm{I}}\mathbf{q}$.

$$\mathbf{M} = \begin{bmatrix} \mathbf{v_x} \\ \mathbf{v_y} \\ \mathbf{v_z} \end{bmatrix} \tag{3.23}$$

The conversion of the rotation matrix to the quaternion $F_R^Q(\cdot)$.

$${}^{\mathrm{S}}_{\mathrm{G}}\mathbf{q} = F_R^Q(M) \tag{3.24}$$

$${}^{\mathrm{S}}_{\mathrm{I}}\mathbf{q} = {}^{\mathrm{I}}_{\mathrm{G}}\mathbf{q} \otimes {}^{\mathrm{S}}_{\mathrm{G}}\mathbf{q}^* \tag{3.25}$$

In-Process Calibration

In-process calibration refers to uninterrupted experimental collection to calibrate the orientation of finger segment orientations by joint motion. To complete the in-process calibration, we need to rely on three basic information:

First, the cumulated error of drift for a long time is large. The drift error in a short time is small. q represents the quaternion and the subscript represents the orientation at a certain moment or the rotation during a period of time. As shown in Fig. 3.5. (ce: cumulative error, k: k moments, $k+n$: $k+n$ moments, n is a very short period of time, $0 \sim k$: from 0 to k moments, r: real value.)

$$\begin{aligned} \mathbf{q}_k &= \mathbf{q}_{k\mathrm{r}} \otimes \mathbf{q}_{\mathrm{ce}(0\sim k)} \\ \mathbf{q}_{k+n} &= \mathbf{q}_{\mathrm{r}(k+n)} \otimes \mathbf{q}_{\mathrm{ce}(0\sim k)} \otimes \mathbf{q}_{\mathrm{ce}(k\sim k+n)} \end{aligned} \tag{3.26}$$

Eq.3.26 eliminates the error term $\mathbf{q}_{\mathrm{ce}(0\sim k)}$ and $\mathbf{q}_{\mathrm{ce}(k\sim k+n)}$.

$$\mathbf{q}_{k+n} \otimes \mathbf{q}_k{}^* = \mathbf{q}_{\mathrm{r}(k+n)} \otimes \mathbf{q}_{\mathrm{r}k}{}^* \tag{3.27}$$

Second, the representation of the joint rotation axis in the finger segment coordinate system is invariant. We can obtain a representation of the joint rotation axis in the sensor coordinate system. The cumulated error caused by the drift on the sensor also impacted the joint rotation axis in the sensor coordinate system.

$$\mathbf{q}_k = \mathbf{q}_{rk} \otimes \mathbf{q}_{ce(0\sim k)} \tag{3.28}$$

The cumulated error also affects the joint axis vector ${}_j$.

Third, the rotation of adjacent finger segment orientations can be divided into two parts, common rotation and joint rotation. Doing the difference of two adjacent finger segment orientations can get joint rotation.

$$\begin{aligned} {}_j\mathbf{q}_k &= {}^{I2}\mathbf{q}_k \otimes {}^{I1}\mathbf{q}_k{}^* \\ {}_j\mathbf{q}_k &\rightarrow (\mathbf{v}_j, \theta) \end{aligned}$$

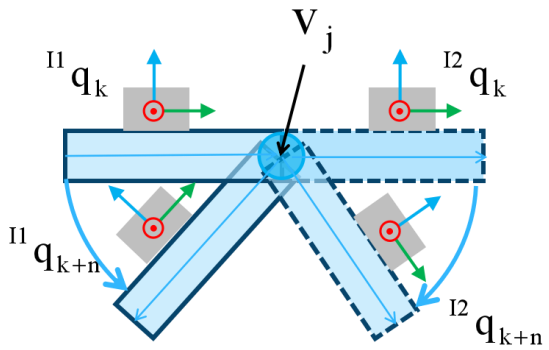

Figure 3.5: In-process calibration, joint axes are obtained using short time n frames of joint rotation. Eliminate the effect of accumulated sensor errors on the joint axis for long periods of $0 \sim k$.

$$\begin{aligned} {}_j\mathbf{q}_k &= {}^{I2}\mathbf{q}_k \otimes {}^{I1}\mathbf{q}_k{}^* \\ {}_j\mathbf{q}_{k+n} &= {}^{I2}\mathbf{q}_{k+n} \otimes {}^{I1}\mathbf{q}_{k+n} \end{aligned} \tag{3.29}$$

Expanding through Eq.3.26 and eliminating the error term, the result only contains the r term.

$$\begin{aligned} {}_j\mathbf{q}_{(k-k+n)} &= {}_j\mathbf{q}_{k+n} \otimes {}_j\mathbf{q}_k \\ &= \left({}^{I2}\mathbf{q}_{r(k+n)} \otimes {}^{I2}\mathbf{q}_{rk}\right) \otimes \left({}^{I1}\mathbf{q}_{r(k+n)} \otimes {}^{I1}\mathbf{q}_{rk}\right) \end{aligned} \tag{3.30}$$

${}_j\mathbf{q}_{(k-k+n)}$ no longer contains the cumulative error term from the long time. We convert quaternions to axis angles ${}_j\mathbf{q}_{(k-k+n)} \rightarrow (\mathbf{v}_{rj}, \theta)$.

The rotation quaternion between the two vectors is calculated for $\mathbf{v}_{rj}$ without the accumulated error and $\mathbf{v}_{ecj}$ containing the accumulated error, and the calibration is performed using this rotation quaternion ${}^{\mathbf{v}_{rj}}_{\mathbf{v}_{ecj}}\mathbf{q}$.

3.3 MIMU-based Data Glove System Design

In this section, we will introduce the implementation of the MIMU-based data glove system as shown in Fig. 3.2. The system consists of two parts: data collection and data processing and analysis.

3.3.1 Data collection

The data collection hardware is designed to be mounted on each finger segment and to collect data from multiple MIMU sensors.

The data glove is composed of three parts, MIMU sensor module, voltage adapter shield, and Raspberry Pi. Each segment on the hand is deployed with a MIMU module. Therefore we connected each MIMU module through Flexible Flat Cables (FFC). The MIMU sensor module is an expandable measurement module composed of MIMU sensors, sensor external circuits, and connectors. The 9-axis MIMU sensor (ICM20948) is a low-power digital sensor. It contains a 3-axis gyroscope, a 3-axis accelerometer, and a 3-axis compass. The characteristics of MIMU sensors are shown in the TABLE 3.1. The sensor is set to collect raw data at 229.8 Hz via SPI. The voltage adapter shield is responsible for converting the 3.3v signal and power transmitted from the Raspberry Pi into 1.8v signal and power for the MIMU module, and leads out the chip select port for each IMU sensor module from the system processor. Raspberry Pi is a tiny computer. It controls the data collection process for each sensor and transfers the data to a computer via WiFi. A battery shield on the Raspberry Pi powers the glove.

3.3.2 Data processing and analysis

Data is uploaded to the server for processing and analysis. The inertial sensor raw data is used to calculate the joint angles on an Ubuntu 18.04 server with a CPU of Intel Core i9-10900X @ 3.70GHz. And the algorithm is programmed in python 3.6.9.

MIMU calibration is necessary before performing the sensor fusion algorithm. the MIMU is placed stationary on a table to calibrate the accelerometer offset and the gyroscope offset. The magnetometer data are collected in each direction of complete rotation. Then, the ellipse fitting algorithm is applied to calibrate the magnetometer [73].

After that, we first perform a StoS calibration after data fusion and finally calculate the joint angle. as described in Section V-B and Section V-C.

Table 3.1: Characteristics of the ICM20948

Characteristics	Range	ADC	Noise level
Gyroscope	± 2000 deg/s	16 Bit	0.015 deg/s
Accelerometer	± 2 g	16 Bit	0.00023 g
Magnetometer	± 4900 µT	16 Bit	0.15 µT

3.4 Experiment and Evaluation

In this section, we evaluate the performance of the finger joint measurement system. First, we describe the experimental setup and the motion collection protocol. Then, we evaluate the accuracy of our system in various settings. Finally, our system is compared with the state-of-the-art.

3.4.1 Experiment Setup and Data Collection Protocol

In the experiments, we used the MIMU-based data glove from Section 3.3 to collect data. We recruited 10 male volunteers (ages 27.0 ± 3.4) for the experiment. The subject sits and performs predefined movements as in Fig. 3.6(a).

Figure 3.6: Experiment System Setup. (a) Wearing markers and data glove to perform flat hand pose. (b) Four cameras were set up near the hand to provide low views. (c) Display the relative positions of the 12 cameras in the optical motion capture interface.

Optical Motion Capture System

The optical motion capture system VICON is adopted as the reference system [74], which measurement accuracy is more than ten times higher than the MIMU-based systems. VICON system consists of 12 infrared cameras, including 8 cameras equally distributed along the outer edge of the ceiling in a measurement room and 4 cameras set up around the subject. Since the space on the hand is relatively small, we use 8mm markers which are trackable and small enough and shown in Fig. 3.6(a). The real scenario is shown in Fig. 3.6(b) and the camera in the VICON motion capture system is shown in Fig. 3.6(c) We use the joint angles obtained by mounting a marker at each joint [74], [75].

Predefined Motions

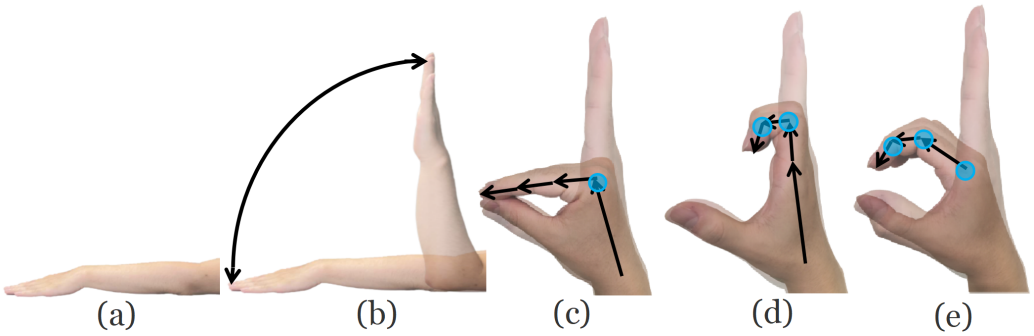

Figure 3.7: Predefined motions in the experiment. (a) flat hand (a) swing hand (b) MCP flexion (c) DIP_PIP flexion (d) DIP_PIP_MCP flexion

Predefined motions are shown in Fig. 3.7 and TABLE 3.2 including joint motions and non-joint motions. We set up three finger joint motions: MCP flexion, DIP_PIP flexion and DIP_PIP_MCP flexion. In addition, we set up the flat hand and swing. Flat

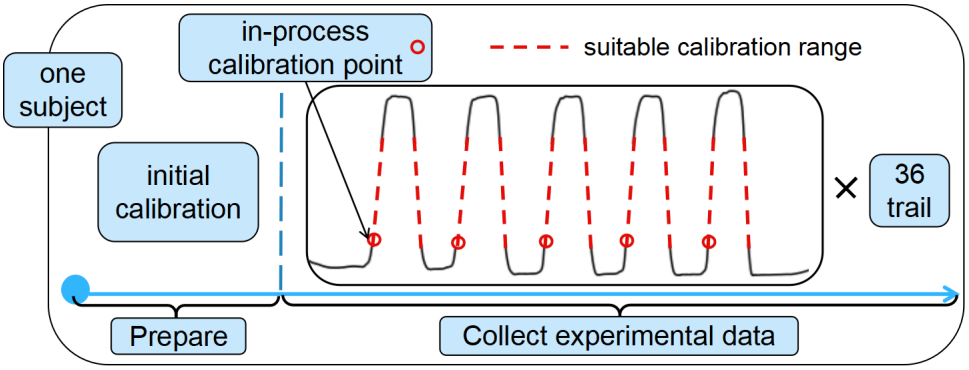

Figure 3.8: The data collection protocol is divided into a preparation stage and a collection experimental data stage. Calibration points for in-process calibration are red circles.

hand is the flat hand pose. These two motions have no joint angle changes but contain different external disturbances to the inspection.

Table 3.2: Hand movement tasks

Device mounting alignment	
(1)flat hand	
Slow motion(0.5 Hz)	Fast motion (1 Hz)
(2)Swing up and down	(6)Swing up and down
(3)MCP flexion	(7)MCP flexion
(4)DIP PIP flexion	(8)DIP PIP flexion
(5)DIP PIP MCP flexion	(9)DIP PIP MCP flexion
Device mounting misalignment	
Same as motion (1) to (9)	

MCP: metacarpophalangeal joint;
PIP: proximal interphalangeal joint;
DIP: distal interphalangeal joint

Data Collection Protocol

At the beginning of the experiment, subjects were asked to wear the markers on their hand as shown in Fig. 3.1. The protocol of data collection for each subject is shown in Fig. 3.8. In the preparation phase, a static pose (flat hand), and joint flexion and extension movements were performed for initial calibration. In the data collection phase, the hand movement task was performed as shown in TABLE 3.2. This process was repeated on different days. Each subject performed each movement twice. resulting in a total of 360 trials ($10 \times 18 \times 2$).

Data processing

The MIMU system and optical system are independent. The alignment of coordinate systems and clocks between systems is important. To synchronize the two measurement systems [71], [76], three optical markers are mounted on the MIMU of wrist,

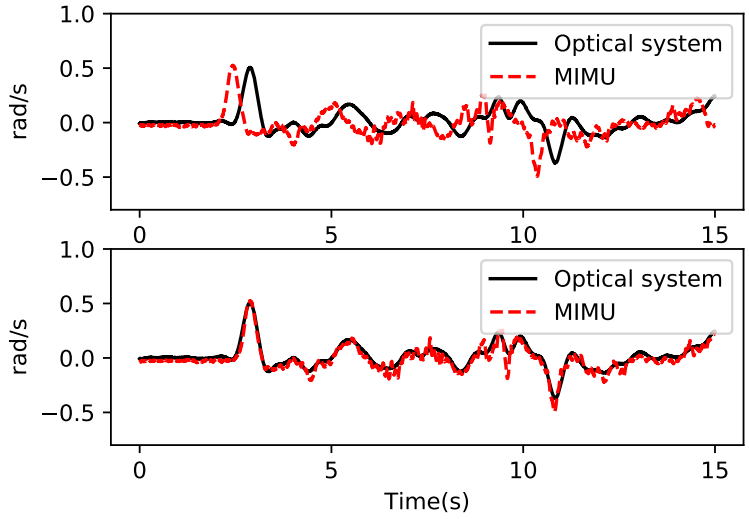

Figure 3.9: The angular rates of the optical and MIMU systems in unsynchronized and synchronized

as shown in Fig. 3.1. The position of markers is measured, and then the reference orientation parameters are extracted by post-processing. The angular velocity is calculated by reference orientation. In the experiments, firstly, the MIMU is kept stationary and the orientation of the MIMU is obtained using the SRCKF, and then we calculate the offset between the two coordinate systems. Second, for each experimental data, we perform a cross-correlation between the two angular velocities and determine the time drift by the extreme value of the correlation coefficient obtained at the beginning of the experiment from static to motion as shown in Fig. 3.9.

The parameter settings for in-process calibration are shown in Fig. 3.8. The suitable calibration range is 0.2s movement over 10 degrees, marked by the red dashed line. We set the calibration point as the initial point of the suitable calibration range shown by the red circle, and do not repeat the calibration within 1s.

3.4.2 Results of Joint Angle Accuracy

Examples are shown in Fig. 3.10 with joint angles after data processing from the proposed system and the optical motion capture system in the three joint flexion tasks in one trial. For optical systems, it is not possible to install markers on all finger segments. This is because when two marker points are too close together, the system often misidentifies them as one marker point. Therefore, we collect test data with one finger, the middle finger, as a representative.

The accuracy of the MIMU-based hand motion analysis obtained was assessed using the Root Mean Square Error (RMSE) as in equation (3.31).

$$RMSE = \sqrt{(\frac{1}{n})\sum_{i=1}^{n}(y_i - x_i)^2} \tag{3.31}$$

The average RMSE of the flexion/extension angles of the middle finger is $5.5°(4.0°)$

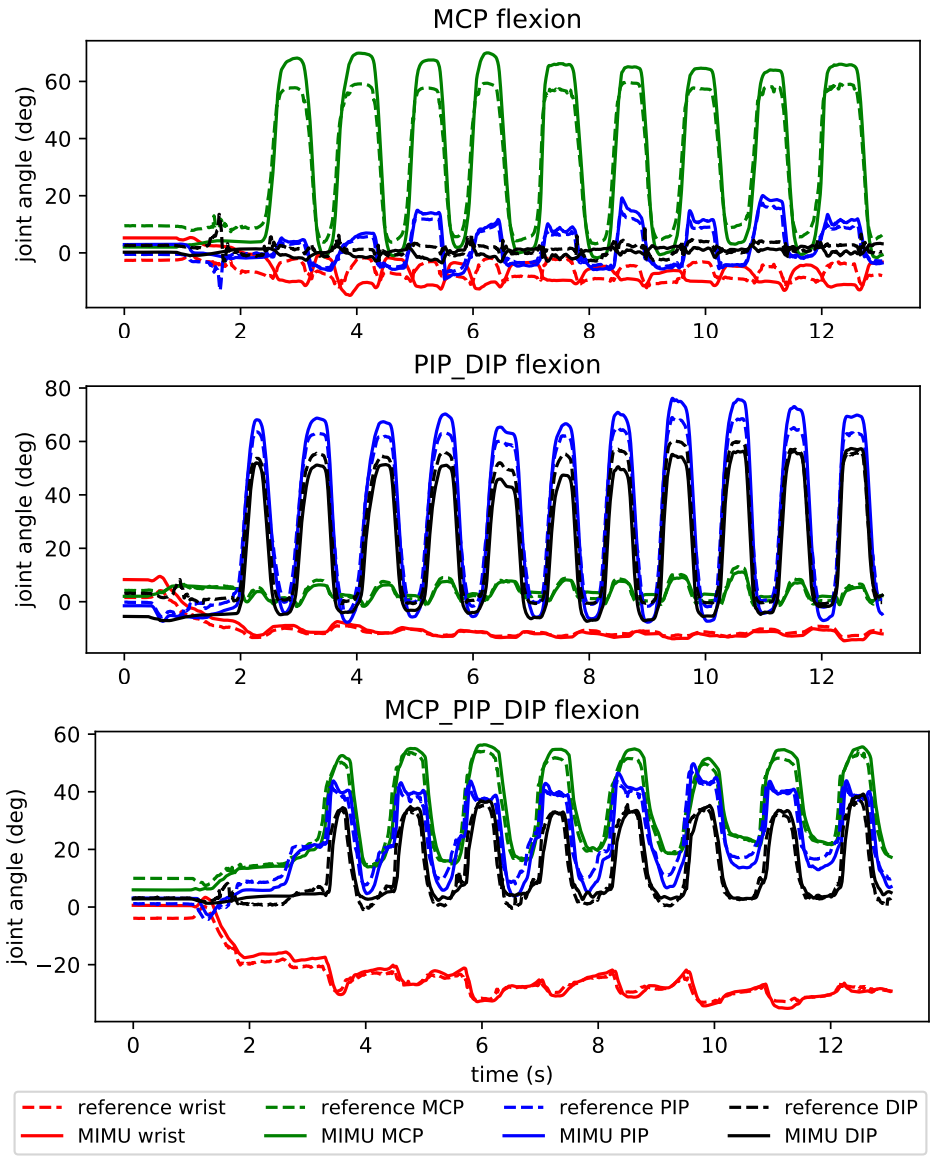

Figure 3.10: Examples of middle finger joint angle measurements in three knuckle tasks versus an optical reference system. (A) MCP flexion; (B) PIP_DIP flexion; (C) MCP_PIP_DIP flexion;

over all dynamic movements, as shown in TABLE 3.4. The dynamic motion average RMSE is much larger than the static ones.

Different type of movements

To investigate the effects of different types of movements, we compared the system performance in 1 pose and 4 movements settings, including flat hand pose, swings up and down, MCP flexion, DIP_PIP flexion, and DIP_PIP_MCP flexion. Flat hand pose means that the segment of the hand remains immobile during the task. we collected joint angle data from 10 participants. As shown in Fig. 3.11, the average RMSE for 1 pose and 4 movements are 0.82°, 3.2°, 5.8°, 6.2°, and 7.0°, respectively. The joint angle of the flat hand pose is accurate due to the small range of motion (ROM). As the ROM of the motion becomes larger, the external disturbances also become larger, causing the average RMSE to gradually increase.

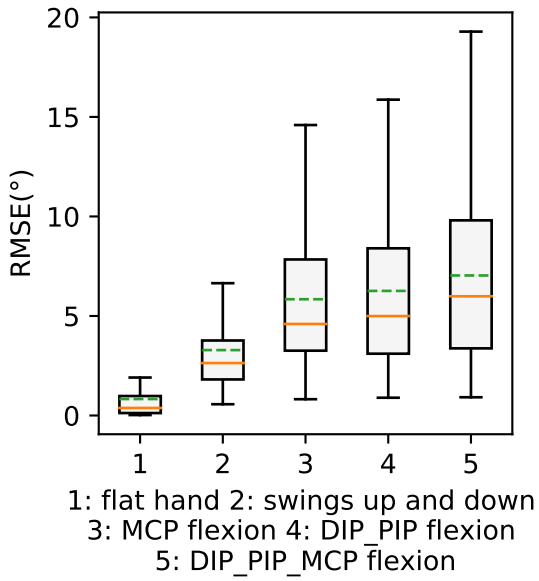

Figure 3.11: Compare 1 pose and 4 different movements.

Movement speed

To investigate the effect of different motion speeds, we compared the system in a stationary pose, in slow motion (motion speed 0.5 Hz), in and fast motion (motion speed 1 Hz). Fig. 3.12 shows the average RMSE of joint angles at different motion speeds. As can be seen in the figure, the average RMSE of the flat hand pose is the smallest, the inertial sensor receives less external interference during slow motion, and the joint angle is more accurate, with an average RMSE of 5.2°. In contrast, the external disturbance of fast motion is large, and the average RMSE is 5.8°.

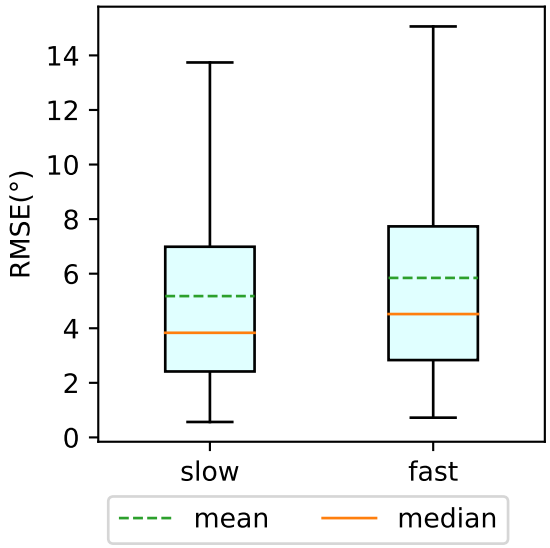

Figure 3.12: RMSE boxplot in Static, Slow Motion, and Fast Motion (movement speed).

Sampling rates

We use downsampling to illustrate the effect of the sampling rate on the joint angular accuracy of the finger. We used the same data set of MCP flexion slow, DIP_PIP flexion slow, DIP_PIP_MCP flexion slow, MCP flexion fast, DIP_PIP flexion fast, and DIP_PIP_MCP flexion fast by 2 participants at a sampling rate of 229.8 Hz. Then, the data are downsampled to 200 Hz, 100 Hz, 50 Hz, 25 Hz, and 10 Hz to determine the average RMSE of the finger joint angles at each sampling rate. In the experiment, slow means that the movement period is about 0.5 Hz and fast means that the movement period is about 1 Hz. The results are shown in Fig. 3.13. The average RMSE of the finger flexion motion decreases as the sampling frequency increases. In practice, we set the sample rate of the instrument to 200 Hz to obtain as much information as possible about finger flexion.

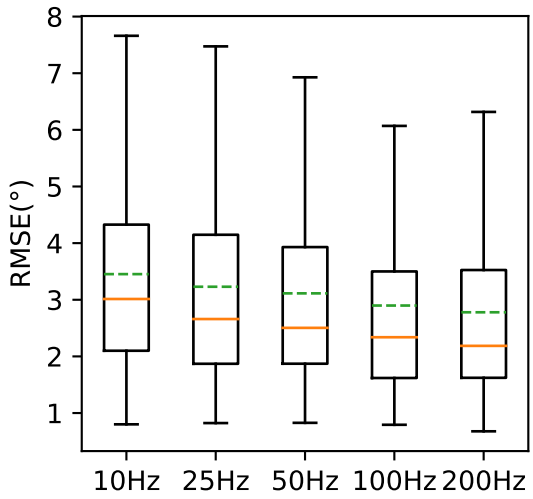

Figure 3.13: The finger flexion data at 229.8 Hz were downsampled to obtain the data at 10, 25, 50, 100, and 200Hz. The RMSE of the finger flexion motion decreases as the sampling frequency increases (sampling rates).

Mounting orientation

We set up two ways for the sensor to finger segment with different mounting orientations in order to evaluate the performance of the system. These orientations of alignment (0 degrees offset) and misalignment (about 30 degrees offset) as shown in Fig. 3.15. Data were collected in all poses and motions including flat hand pose, Swing up and down, MCP flexion, DIP_PIP flexion, and DIP_PIP_MCP flexion. Data were collected from 10 participants under the initial orientation of both installed devices. Fig. 3.14 shows the results for different mounting orientations. Our system is not affected by the mounting orientations, and we observe no significant difference between the two mounting orientations. Consequently, our system is insensitive to mount orientation.

Comparison with Representative Method

In this section, we compare our system with existing work. The hand motion capture system [26] is based on the static pose method on StoS calibration. To compare

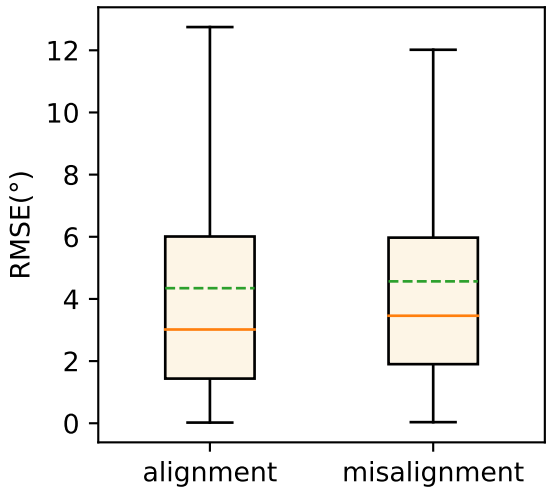

Figure 3.14: RMSE boxplot in alignment and misalignment (mounting orientation).

Table 3.3: Our system is compared with Kortier's systems (EKF: extended Kalman filter)

Researches(year)	Hardware	Number of subjects	Sample frequency	Data fusion algorithm	flexion tasks (100Hz) (MCP/PIP/DIP)
Ours	**ICM20948**	10	200Hz	SRCKF	4.3°/3.8°/3.9°
Kortier et al. (2016)[27]	**STLSM330DLC**	3	100Hz	EKF	5.0°/7.3°/5.6°

the two calibration methods, both the proposed method and the static pose method use the orientation estimate of the SRCKF with high accuracy as input. The computational complexity of the calibration part is $\mathcal{O}(n)$ in each calibration part. We used the same data set for comparison, which contains 1 pose and 4 dynamic movements, including flat hand pose, swings up and down, MCP flexion, DIP_PIP flexion, and DIP_PIP_MCP flexion. As shown in Fig. 3.16. Our method in the flat hand pose has a similar average RMSE as the static pose method. In addition, our method provides more accurate joint angle estimates in 4 dynamic motions. However, in practice, many parameter adjustments are highly correlated with the equipment. Therefore, this result is only illustrative in these experimental and parameter settings.

Figure 3.15: Two types of mounting orientations include alignment (0 degrees offset) and misalignment (about 30 degrees offset).

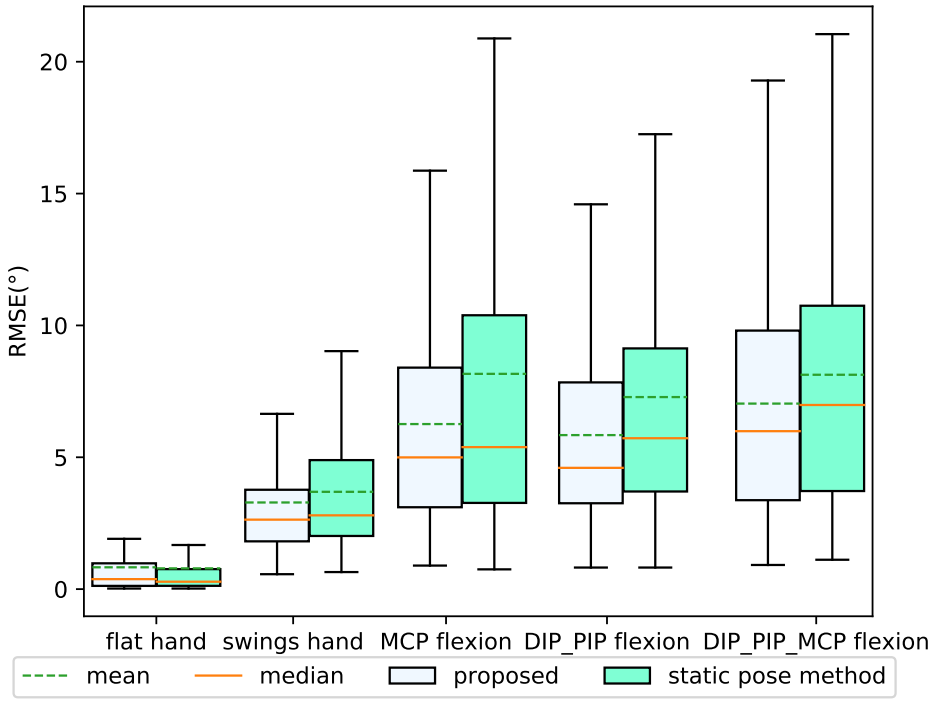

Figure 3.16: Box plot of the proposed method and the static pose method for RMSE in different types of movements.

3.5 Discussion

We proposed and evaluated a MIMU-based hand motion capture system for measuring finger joint angles. The average RMSE between our system and the optical reference system with markers was $5.5°(4.0°)$. The system was also tested in various experimental settings, verifying that it is accurate and stable.

From the intuitive results of the experiments, the average RMSE of the joint angle is related to the ROM of joint movement, and DoF of joints. However, ROM and joint DoF should not be restricted. To improve the accuracy of the system, considering the characteristics of the inertial sensor and the biological structure of the hand, the following dilemma exists. Inertial sensor measurements are subject to inherent errors, especially in activities such as rotation and translation or vibration. These external disturbances are not completely captured by the model built by the filter. The soft tissue artifact (STA) effect of the hand has a large individual variation, a lack of measurement tools and methods, and no established mathematical model. In addition, STA affects the crosstalk of multi-DoF joints and becomes larger.

Our system is compared with another hand joint measurement system as shown in TABLE 3.3. Currently, there is no unified convention as a reference system for measuring hand joint angles of inertial-based motion capture systems. The optical system is highly accurate but costly, so there are relatively few similar studies using optical systems as reference systems. However, it is necessary for the reference system to be one order of magnitude more accurate than the inertial system. Therefore, we only compare with the study that uses optical system as reference.

Kortier et al. [27] researched and built a hand motion capture system. And van den Noort et al. [77] evaluated Kortier's system for the accurate measurement of various finger motor tasks. The optical system was used as a reference system similar to ours. We

selected the same motion patterns (MCP flexion, DIP_PIP flexion, and DIP_PIP_MCP flexion) and same motion frequency (0.5 Hz) as experimental data and downsampled them to 100 Hz for comparison with Kortier's system. We improved the accuracy in each joint.

3.6 Summary

This paper proposed a hand motion capture system for measuring hand joint angles. Data fusion is performed with the SRCKF, and StoS calibration is performed with a joint kinematic constraints method. A highly customizable hardware platform is constructed. The modular unit is easily expandable and directly taped to the back of the hand to maintain the intra-palmar tactile sensation of the hand. The errors are measured quantitatively by the optical system. The RMSE of the joint angle is $5.5^{\circ}(4.0^{\circ})$ for dynamic motion. The stability of the system was verified by setting the type of movements, movement speed, sampling rates, and mounting orientation in four experiments.

The proposed method is a valuable approach for estimating finger joint angles by improving key steps to enhance accuracy. However, there are still many problems that need to be resolved. All non-invasive techniques, such as MIMU-based systems or marker-based optical systems, are subject to soft tissue artifact (STA) effects. A more ideal reference system would be direct magnetic resonance imaging (MRI) of the bone. MRI would be a powerful tool to further research on eliminating STA effects. Future work is needed to understand the soft tissue properties and the effect of skin on MIMU during movement. A promising method for future research is to combine the joint kinematic constraints method to compensate for drift.

Table 3.4: Results of the comparison with the optical system, for each joint angle of the middle finger and wrist joint angle during different movements. (RMSE: Root Mean Square Error, ROM: Range Of Motion with an optical system, SD: Standard Deviation)

action / joint	Swing up and down		MCP flexion		DIP_PIP flexion		DIP_PIP_MCP flexion		All movements	
	RMSE(°) (SD)	**ROM(°) (SD)**	**RMSE(°) (SD)**	**ROM(°) (SD)**	**RMSE(°) (SD)**	**ROM(°) (SD)**	**RMSE(°) (SD)**	**ROM(°) (SD)**	**RMSE(°) (SD)**	**ROM(°) (SD)**
Wrist	4.1 (2.8)	10.1 (4.2)	4.8 (2.4)	18.3 (7.0)	3.6 (1.6)	18.5 (9.0)	5.2 (3.7)	21.9 (9.1)	4.2 (2.8)	15.9 (19.3)
MCP	2.7 (1.9)	8.8 (3.7)	9.3 (4.3)	50.5 (14.3)	4.2 (2.1)	27.2 (12.1)	8.4 (4.5)	37.9 (12.7)	5.3 (4.3)	25.8 (20.1)
PIP	3.0 (2.3)	9.9 (3.7)	5.0 (3.0)	31.9 (13.4)	8.4 (5.0)	79.5 (12.7)	7.8 (5.3)	46.5 (16.5)	5.4 (5.1)	36.3 (33.7)
DIP	3.4 (2.3)	10.9 (3.3)	4.3 (2.0)	21.5 (10.3)	8.9 (4.1)	56.1 (14.9)	6.8 (3.0)	31.7 (17.1)	5.4 (4.2)	28.9 (34.6)
Mean	3.3 (2.4)	9.9 (3.8)	5.8 (3.7)	30.6 (17.1)	6.3 (4.2)	45.3 (27.1)	7.0 (4.4)	34.5 (16.8)	5.5 (4.0)	26.7 (28.8)

Chapter 4

Bending Sensor and Inertial Sensor based on Weighted DTW Fusion

4.1 Introduction

In Japan, there are about 341000 hearing impaired people [78]. The general way to communicate between a healthy person and hearing impairment is communication by writing or sign language. However, communication by writing takes a lot of time. And, sign language that hearing impairment people use is not familiar to healthy people or acquired hearing impairment people. Each of the two approaches has problems that hinder smooth communication in society.

Sign language recognition has always been a research problem that has received a lot of attention. There have been a large number of studies on sign language recognition in recent years [79–82].

Sign language recognition systems can be divided into non-wearable and wearable approaches. The non-wearable generally include vision-based [83, 84] and WiFi signal-based [85, 86] methods in non-wearable ones. Another approach is to recognize sign language with wearable sensor-based data gloves [87, 88].

Due to the development of deep learning methods in visual sign language recognition, the recognition rate has been improved. However, deep learning is driven by data, and the quality of data collection greatly affects the results. Insufficient video frames and occlusions will also reduce the recognition accuracy. Gerges et al. [56] established a dynamic hand recognition based on MediaPipe's Landmarks and compared the recognition accuracy of three deep learning methods: Gated recurrent unit (GRU), Long Short Term Memory (LSTM), and Bi-directional LSTM (BILSTM). Data set collection requires complete characters, no occlusion, and a fixed duration. It is difficult to achieve these requirements in actual use. Chang et al. [89] studied the research that recognizes sign language by detecting the place of nails and wrist by pictures of the hand. It recognizes language by Skelton of hand and distribution of skin color from taken picture of the hand. However, the systems that hearing-impaired people need to use in their daily lives can detect not only the hand shape part of sign language but also the dynamic part of hand movement in sign language. In other vision-based methods, there is the way that uses color gloves and Kinect stored from Microsoft. Shibata et al.[11] uses color gloves for recognizing sign language. The color glove has every color at every finger and wrist. And, it recognizes by moving distance and area of glove colors. However, in the detection step, the background or the user's clothing is the same color as the part

of the glove, which cannot be recognized in this way. Kinect can detect hand motions and hand-places. Muaaz et al.[90] developed a system that can recognize American Sign Language with Kinect. This system has a high recognition rate of an average of 80%. And, this system can make easy sentences by recognizing Sign-Language words. However, this system is also limited by the camera, and we can only use this system in limited positions without occlusion. In daily life, it can be a large barrier for hearing impairments to use the system.

Vision-based sign language recognition is limited by the nature of camera view observation and is not good at capturing complex two-handed interaction movements because of occlusion. It is also susceptible to the influence of the environment between the camera and the object. The way of wearable sensors and Data-Glove forces users to some burdens. But data gloves can collect data steadily in complex environments, without the problem of line of sight obstruction, noisy backgrounds, and inadequate light. It can even be used outdoors, in low visibility. The camera method is subject to a variety of environmental constraints. Therefore, we plan to use wearable devices to capture the complex motion of the fingers.

In recent years, wearable sensor-based data gloves with the continuous improvement of processing information technology and the miniaturization and high functionality of equipment. Wearable sensor-based data gloves have been able to operate a large amount of information and more complex processing.

Common wearable sensor data gloves for sign language recognition include flexible sensor flex sensor [91], Inertial measurement unit (IMU) [92] surface electromyography (sEMG) [93, 94], and touch sensor [95]. As shown in Table 4.1, we compared the studies of various sensors. Portability in the table refers to whether good results can be obtained without any data from new users. EMG data has a large individual variation. When using the Bilinear Model for classification, a new subject needs to perform at least one motion. The recognition rate will drop significantly without using the Bilinear Model.

The information directly related to the hand in sign language includes 21 degrees of freedom of the joints on the hand, and the spatial displacement and orientation of the hand. Complicated information makes it difficult to obtain appropriate characteristics through a single type of sensor. Korzeniewska et al. [91] chose Velostat to make bending sensors to collect data to identify Polish Sign Language and obtained a letter recognition rate of 86.5%. However, sign language generally uses words as the unit of recognition. Youngmin Na et al. [92] installed an accelerometer on the index finger to recognize static letter gestures in the Korean sign language alphabet, but sign language contains a lot of dynamic gestures, and only static gesture recognition is not enough. Jakub et al. [96] collect IMU sensor data installed on the palm and fingertips, and use parallel Hidden Markov Model (HMM) approaches for sign language recognition. The finger shape data can be obtained by combining the IMU data on the fingertips and the IMU data on the palm. For collecting hand shape features, multiple inertial sensors are more expensive than multiple bending sensors.

Data gloves from a single type of sensor either collect much missing hand information or cost a high price to implement. So multi-sensor fusion is a better solution. The use of wearable sensors and data gloves is moving toward practical applications as MEMS technology advances sensors are being miniaturized. It also breaks down the spatial limitations of the hand, making multi-sensor data collection possible. Among the multiple combinations, inertial sensors to collect hand motion and bending sensors

Table 4.1: Comparison of related research (KNN: K-nearest neighbor)

Research	**Sensor**	**Subject**	**Kinds**	**Portability**	**Algorithm**	**Dynamic Motion**
Muaaz et al. [16]	Kinect	5	10	○	DTW	○
Tateno et al. [19]	EMG	20	20	×	LSTM	○
Lee et al. [21]	Touch	-	36	○	Tree	×
Faisal et al. [23]	Inertial and Flex	35	3	○	KNN	○
Chu et al. [26]	Inertial and Flex Force	3	7	○	DTW	○
Ours	Inertial and Flex	8	20	○	weighted DTW	○

to collect hand shape are the common approaches[23-26]. Faisal et al.[97] used the KNearest Neighbors (KNN) classifier 14 static and 3 dynamic gestures Sign Language Recognition. Faisal et al. [52] collected data from 25 subjects for 24 static and 16 dynamic American sign language gestures for the validating system. Boon Giin Lee et al. [98] use the support vector machine (SVM) to classify the American sign language.

The combination of inertial sensors and bending sensors helps us to obtain hand shape and motion information at a low cost. However, how to rationalize multiple sensor data for sign language recognition is still a difficult problem. The execution length of the actions of sign language varies greatly due to people's habits or usage scenarios. The Dynamic Time Warping (DTW) algorithm is a solution to compare the similarity between time series data of different lengths. However, the current research on the application of DTW algorithm to sign language recognition is insufficient. Chu et al. [99] studied DTW for sign language recognition on 7 Japanese Sign Language datasets, Validation performed using leave one out (LOO) approach recognition rates is 82.5%. First of all 7 recognition actions are insufficient. On the other hand, the variation between different sensors is significant in providing useful information for sign language recognition. So it is necessary to propose weighted DTW.

In this study, inertial sensors and bending sensors can be deployed simultaneously in the hand space to collect hand shape and motion features. It becomes a practical and promising solution to combine these two parts of features to recognize sign language. Thus this research Sign-Glove system is implemented, as shown in Fig.1. The development of such systems will give us a future where we wear sensors like accessories that make it easier to communicate between a healthy person and a hearing impaired person. When we develop the system to recognize Sign-Language on portable devices with recent technology, For the recognition algorithm of sign language, we extend DTW to use on time series of multiple sensors. DTW is a general method for measuring the similarity between two temporal sequences. However, for data from multiple sensors, different sensors provide different recognition contributions. So we propose weighted DTW, an algorithm that improves the recognition rate by setting weighted values to raise the effect of key sensors.

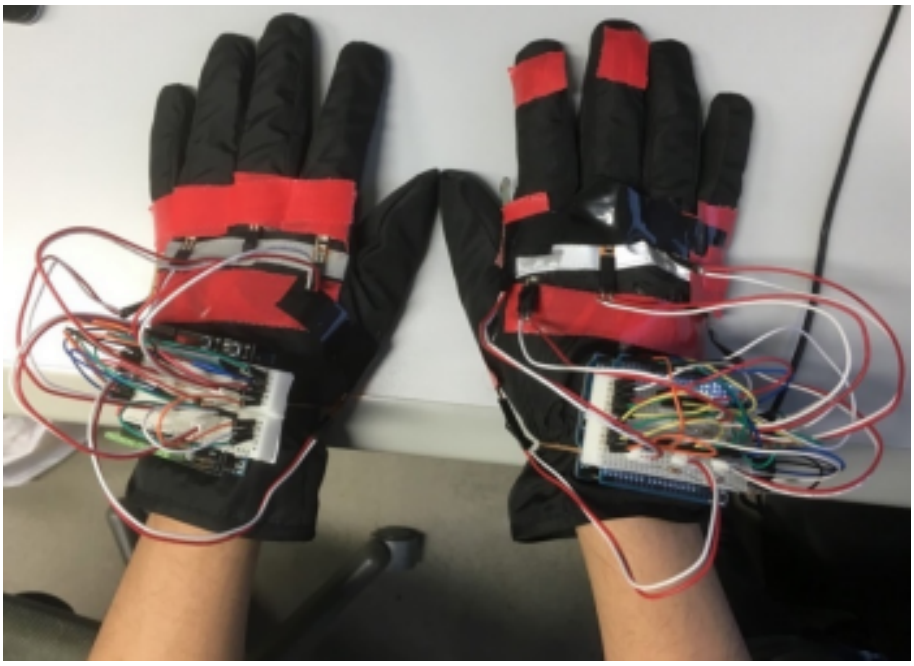

Figure 4.1: The Sign-Glove on hands

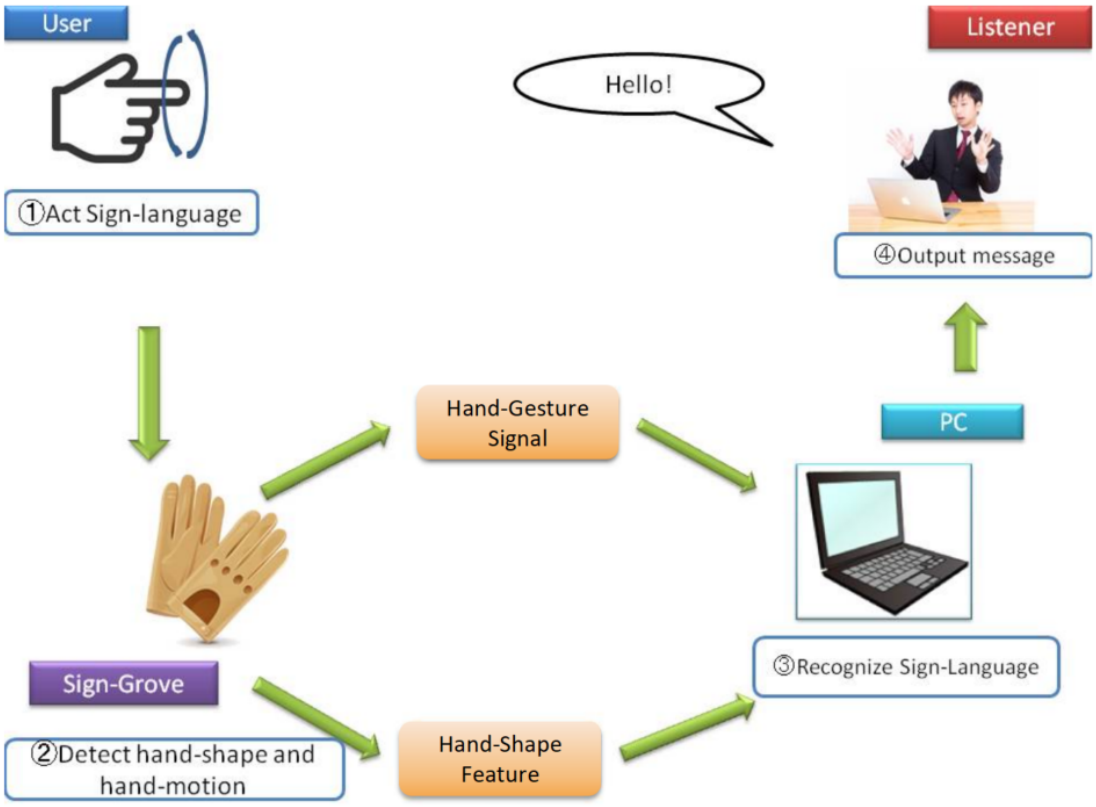

Figure 4.2: Usage of Sign-Glove for sign language recognition.

4.2 Application Model and Sign Languages Dateset

4.2.1 Application Model

A system presented by this research shows the meaning of a sign language word on PC for supporting communication between a healthy person and a hearing-impaired person. This system supposed that a user uses a pair of sign gloves and shows the meaning of a Sign-Language word on the PC.

A user wears sign gloves on his/her hands. And, The user moves a motion of a Sign-Language word. Then, PC shows the mean of sign language. Sign-Glove is a device-shaped glove with WonderSense and a bending-sensor. WonderSense is the device developed in this laboratory. This model supposed that the user wants to communicate a Sign-Language word motion to the other. We explain the process of this system according to Fig.2. A user moves the motion of the Sign-Language word that he wants to communicate the word to the other. Sign-Glove measures the acceleration of hand motion and hand shape at this time. WonderSense of Sign-Glove transmits measured hand acceleration to WondeBox with Bluetooth Low Energy. WondeBox is a receiver device of WonderSense. WondeBox sends measured hand acceleration data to a PC with a serial connection. At the same time, the bending sensors of Sign-Glove measure the hand shape. And, Arduino sends measured data with a serial connection. Arduino is one of the AVR micon boards. Arduino is used for taking data from bending sensors and sending the data to the PC. After the finished Sign-Language gesture, the data values sent by sensors are computed to recognize a Sign-Language word motion. This Sign-Language word motion is converted into a message that is associated with the Sign-Language word motion in the PC. Finally, the PC displays the message requested by the user. At this time, if the message is a serious one, the PC makes a sound.

4.2.2 Sign Languages Dateset

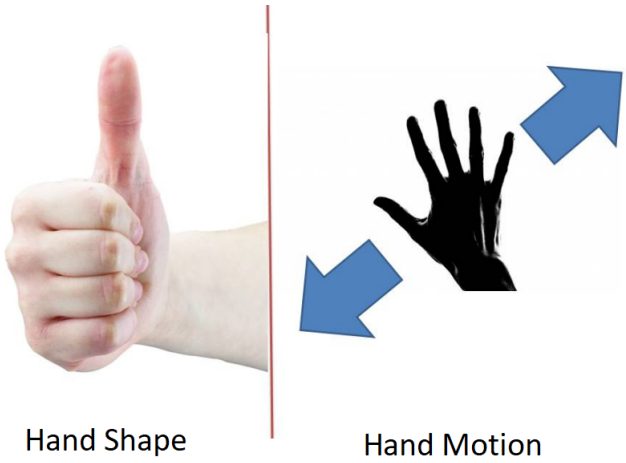

Figure 4.3: Hand shape and hand motion factors for sign languages.

Sign language consists of two main components in the hand part, namely, the shape of the hand and the overall movement of the hand. Static sign language is defined as a special case of dynamic sign language, which specifically means that the shape of the

hand and the hand motion remain unchanged for a period of time. Figure 3 shows the hand shape parts and the hand motion parts of sign language.

4.2.3 Sign Language Dataset Definition

Figure 4.4: Selection of sign language vocabularies with both hand shape and hand motion factors.

The key point to recognizing sign language is to recognize the hand shape and the hand motion at the same time. Missing one of them will significantly reduce the recognition rate, such as “please” and “good”, “sick” and “obstacle”, “down” and “I see”, as shown in Figure 4, because the hand motion of these sign languages is the same but the shape of the hands is different. If we detect only the hand motion of these sign language words, the result is that this sign language is completely the same. In contrast, Sign-Glove used in this research can detect hand shape. Thus, we can increase the recognition rate of sign language words. Furthermore, for the same reason, we can also achieve the correct result of recognizing sign language words, which is the same hand shape and different hand motions.

4.3 Methods

The system architecture is shown in Fig.5. The data glove collects the physical features and the communication structure is shown in Fig.6. We explain the design of the system in section 4.1. We explain the recognition algorithm in section 4.2.

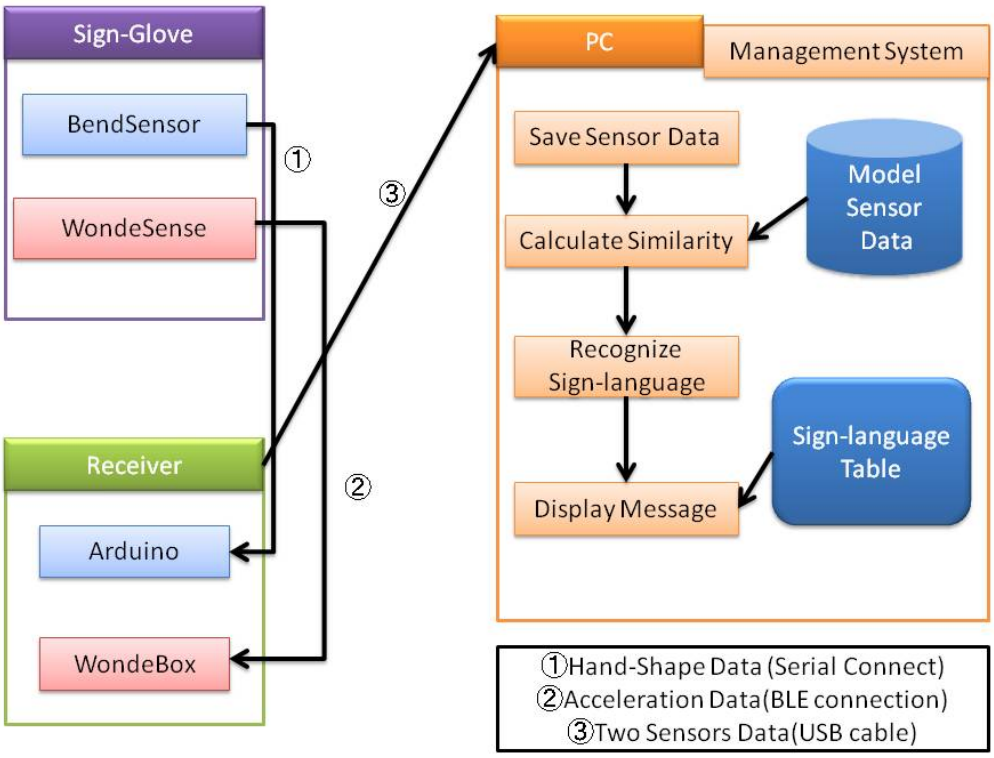

Figure 4.5: Architecture of the system

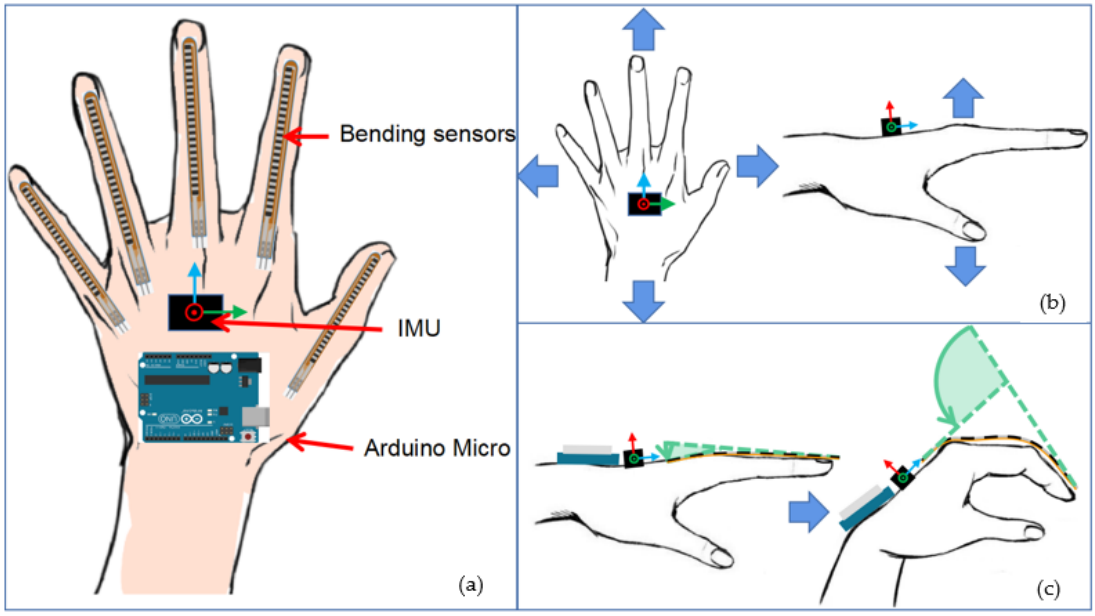

Figure 4.6: (a) System structure diagram; (b) IMU collecting the hand motion data; (c) bending sensor collecting the hand shape data.

4.3.1 System Design

In this section, we explain that a Sign-Language word recognized by a system constructed in this research and the way to detect. First of all, we explain operation follow of this system. This research system recognizes a Sign-Language word based on bending condition data of fingers detected by bending-sensor and acceleration data of hands detected by WonderSense.

First, Sign-Glove wore user's hands takes bending fingers condition data and acceleration data of hands from the bending-sensor and 3-axis acceleration sensor. This bending fingers condition data is sent from Arduino to a PC with a serial connection on the USB cable. At the same time, the acceleration of hands detected by WonderSense sends to WondeBox which is a data receiver with a BLE connection. And, WondeBox sends the data to a PC with a serial connection. Until the end of a Sign-Language word motion, Sign-Glove continues to take the values. Acceleration and bending finger state data are stored in the computer. After that, we read the model data.

The model data is defined with the 3-axis acceleration data and bending data of finger for all sign language words. For the recognition process, we use the Weighted DTW algorithm to calculate the similarity between the sign language data to be recognized and the model data. The highest similarity of the model data is selected. Finally, the meaning of the sign language word is extracted from the sign language table and displayed on the screen.

As shown in Figure 6, the data glove collects the physical features of the hand. The IMU collects the motion features of the hand as shown in Figure 6(b). The bending sensor collects the shape features of the hand as shown in Figure 6(c). Both sensors are stitched to the cloth glove at the corresponding locations for fixation. The bending sensor is fixed in a special way. When the finger is bending, the skin is stretched. But the length of the bending sensor is fixed, so we fix the top of the bending sensor to the fingertip position of the glove, and the middle of the sensor is restricted by the wire without shifting from left to right. The back end of the sensor is not fixed and can be free to stretch, only let the sensor and the back of the hand as far as possible to fit.

4.3.2 Implementation

Hardwares

We explain the construction of the hardware in this research. Sign-Glove is the device that takes the acceleration of hand-gesture and hand-shape. Sign-Glove has two kinds of sensors. One of the two sensors is the bending-sensor. Fig.6 (a) shows the bending-sensor. Bending-sensor changes its resistance by bending condition. The bending sensor has a polymer ink on one side the sensor is about $30k\Omega$ resistance when straight and the resistance increases when it bent. Arduino detects the change as a voltage change value and sends it to the PC with a USB cable. Arduino is one of the AVR Micon boards. We use it as a banding sensor receiver. The other sensor is WonderSense. Fig.6 (b) shows WonderSense. WonderSense collects acceleration data using a 9-axis inertial sensor module MPU9250. WondeBox is a data receiver of WonderSense. The core chip of the WondeBox is the PCA10040 for Bluetooth data reception. WondeBox sends data to a PC with a USB cable. A sign-Glove device is a pair of gloves. Sign-Glove is constructed by ten bending sensors, two Arduino, two WonderSense, and one WondeBox. To facilitate synchronization, we sampled both the inertial and bending

sensor data sampling rate to 50 Hz.

Softwares

We explain data taken from WonderSense. As above, the data taken from WonderSense is acceleration data. The acceleration data taken from WonderSense is sent to WonderTerminal on PC through WonderBox. WonderTeminal has a function that builds a server. The server of WonderTerminal sends the acceleration data to our Sign system. The acceleration data format is String and the frequency is 50Hz. We use the acceleration data for recognition. Our research system can save the acceleration data from WonderSense into DataBase. We explain data taken from Bending-sensors. As above, the data taken from Bending-sensor is a resistance value of finger bending. The resistance data taken from every bending sensor is sent to our research system in PC through Arduino. In this system, we can save the data from Bending-Sensors into Database.

4.3.3 Recognition Method

Dynamic Time Warping

The Dynamic Time Warping (DTW) algorithm is for measuring waveform similarity. The DTW algorithm calculates the similarity of time-series data using Euclidean distance. The feature of the DTW algorithm is that the length of sample data does not become a problem for the calculation. The duration of sign language varies while expressing the same word according to habit, proficiency, and other factors. Even in this situation, the DTW algorithm can calculate similarity. Next, we explain how to calculate the DTW algorithm for a single sensor.

Weighted DTW

DTW can calculate model data and sensor data similarity for a single sensor. And by assigning weights, the Weighted DTW can effectively fuse data from multiple sensors. The model data is the ideal data generated by analyzing the average value of the standard action and the waveform trend of each sensor for multiple executions.

The contribution of each sensor to sign language recognition is different. In this research, we used both bending sensors and inertial sensors. And two inertial sensors measure the movement of two hands, and 10 bending sensors measure the bending of ten fingers. On the one hand, the types of sensors are different, so the effectiveness of information is different. On the other hand, even with the same sensor, for sign language recognition, the thumb, index finger, and middle finger of the right hand provide more critical information in many cases. While the other fingers most of the time make little contribution to distinguishing sign language. Due to a large number of static states, the waveform has less effective information and is more affected by noise. So setting the same weight is unreasonable. We set different weights between 10 bending sensors, different weights between 2 inertial sensors, and different weights between 2 types of sensors. The weights calculation process is as follows:

Table 4.2: Setting of the weight parameters in the experiment.

Weight	Value	Weight	Value
α	0.05	β_6	0.0002
β_1	0.2448	β_7	0.0002
β_2	0.3772	β_8	0.0002
β_3	0.3776	β_9	0.0002
β_4	0.0002	β_10	0.0002
β_5	0.0002	γ	0.5

4.4 Experiment and Evaluation

In the experiment, we evaluate the performance of the sign language data glove. We first describe the experimental setup. Next, the experiments compare the recognition performance of hand shape, hand motion, and combined data of both. After that, we verify the recognition performance of our weighted DTW.

4.4.1 Experiment Setting

Figure 4.7: The basic pose and hand position during the usage of the system: (a) side view

We recruited 8 volunteers and collected data on 20 sign language words. The average age of subjects is 22. Each person repeated each sign language three times, and we collected a total of 8*20*3 data. model data is the average value of multiple executions of the standard action. Table 4.2 is the weight parameters when we experiment. Next, we introduce the usage of Sign-Glove.

Figure 4.8: The basic pose and hand position during the usage of the system: front view.

In this section, we explain how to wear Sign-Glove and the starting position of recognizing a Sign-Language word when we recognize a Sign-Language word in Fig.4 with this system. First, we wear Sign-Glove on our hands. We should insert our fingers into Sign-Glove because Sign-Glove has a bending-sensor for each finger part. Fig. 6 (a)is a correct image of wearing Sign-Gloves.

Fig. 4.7 and Fig.4.8 show a pose and hand position when we start to recognize a Sign-Language word. A basic position is sitting in a chair, and putting your hands on your knees. We must start to recognize a Sign-Language word from the basic position. In this research, we start the system during recognition. And, when you finish a Sign-language word motion, we stop this system. Then, we don't return our hands to the basic position.

4.4.2 Experiment Results

Comparison between the hand shape, hand motion, and combination methods

As shown in Fig.4.9 the recognition rate of twenty kinds of sign Language in this experiment. Experiments were performed to calculate the recognition rate of sign language for three kinds of feature data: combined hand motion data and hand shape part data, only based on hand motion data, and only based on hand shape part data. The motion data of the hand originates from the inertial sensor, which is shown as a red rectangle on the graph. The shape of the hand originates from the bending sensor, which is shown as a blue rectangle on the graph.

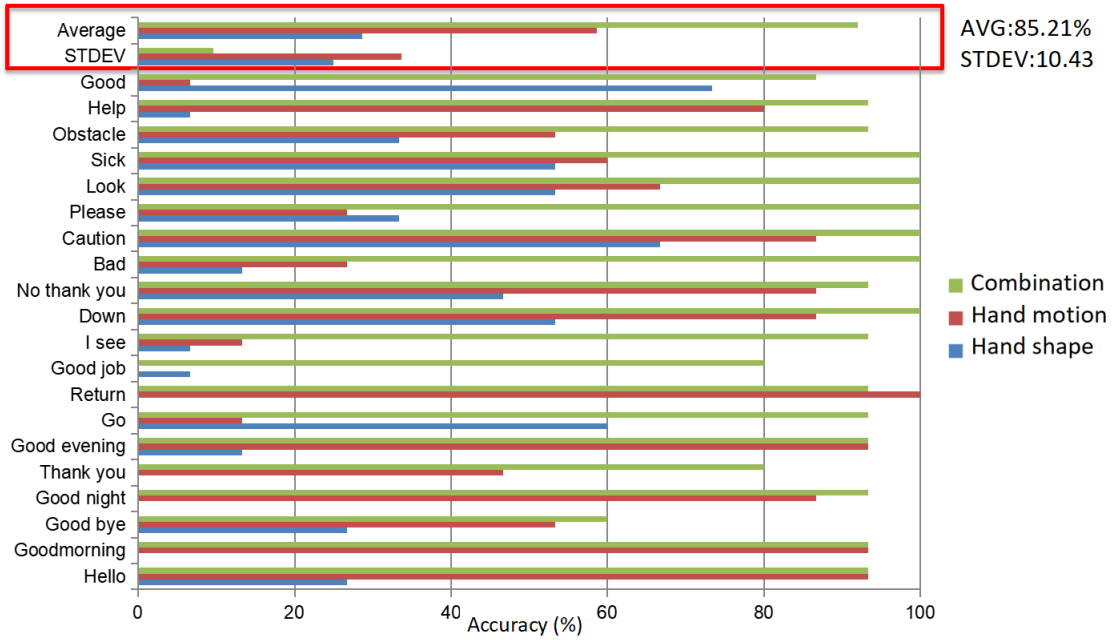

Figure 4.9: Comparison between the hand shape, hand motion, and combination methods. (The red boxes on the right give the values of AVG and STDEV for combination method.)

4.4.3 Comparison between using our proposed weighted DTW or unweighted DTW

We show a result which cases when we use different weights or the same weight as data fusion. As shown in Fig.4.10.

When we use our weight by data fusion algorithm, we can take the average recognition rate as 85.21% and the standard deviation is about 10.43. When we don't use our weight by data fusion algorithm, the average recognition rate is about 57.92% and the standard deviation is 27.50%. So, we can understand this data fusion algorithm is increasing the recognition rate and decreasing the standard deviation. Thus, we can say that this algorithm is useful for recognizing sign language.

4.4.4 Discussion

We build a data glove based on bending sensors and inertial sensors to capture hand shape and motion features, and then use weighted DTW fusion features to recognize sign language. We experimentally verify that both hand shape and hand motion contribute to sign language recognition. Moreover, the two features are complementary, and a higher recognition rate can be obtained by fusing the two features to recognize sign language. Adjusting the weight values to fuse the features, we find that the quality of information provided by sensors with different placements is different. By adjusting the weights to focus on the sensors with large value changes during the execution of the sign language, the recognition accuracy can be improved. We collected data for 20 dynamic sign language words from 8 volunteers, and the recognition accuracy was 85.21%. The feasibility of the system was verified.

In comparison with similar systems, although there have been a large number of studies on sign language recognition. But the defined sign language countries are dif-

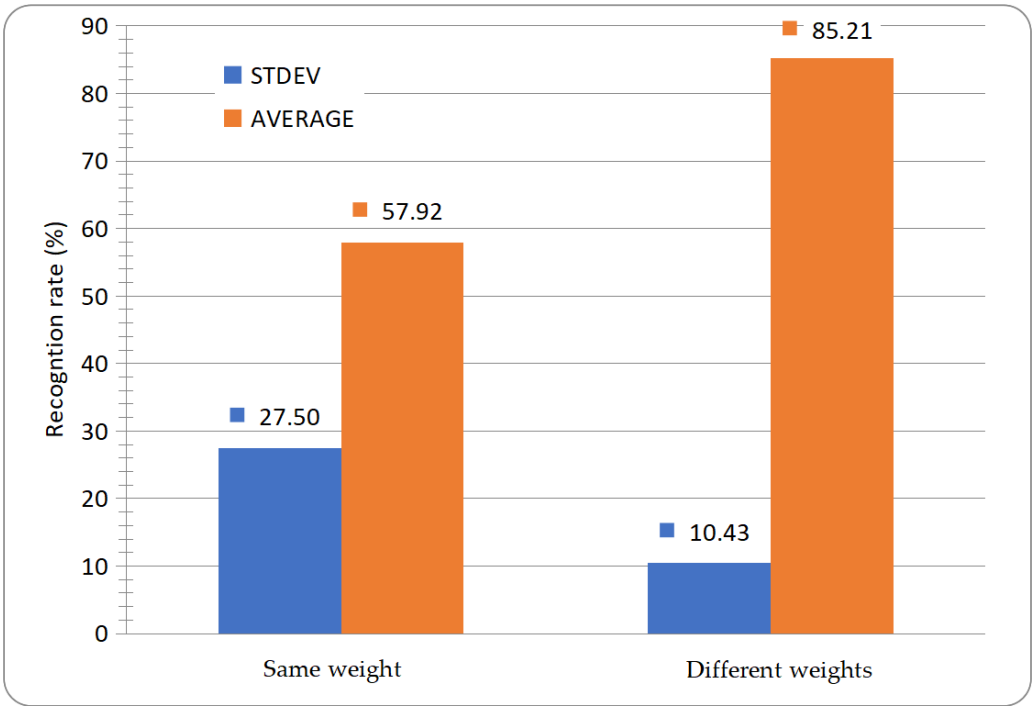

Figure 4.10: Comparison between using our proposed weighted DTW or original DTW.

Table 4.3: Comparison between using our proposed weighted DTW or original DTW.

Researches (years)	**Subject**	**Number of words**	**Algorithm**	**Sensor**	**Cross-recognition**
Our	**8**	**20**	**Weighted DTW**	**Bend and IMU**	**85.21%**
Chu et al. (2021)	**3**	**7**	**DTW**	**Bend and IMU**	**82.5%**

ferent and the number of participants in the experiment is different. The amount of sign language word data contained in the dataset is different. We choose Chu's systems, which are similar in structure to our system and both use bending sensors and IMUs, and also use Japanese sign language, for comparison. The results are shown in Table 2, which shows that the weighted DTW has a better recognition rate when the number of participants and the number of recognized sign language words are both greater.

There are still many limitations of our system. The data glove prototype system uses a breadboard, so the system is rather bulky. For some palm-related sign language words, some times caused inaccurate movements. However, the semantic impact on the sign language expression is minimal. It is still able to recognize sign language words in sign language communication. For the impact of data collection, there will be data loss or disconnection problem during long time data collection.

In addition to hand shape features and hand motion features, collecting other features in sign language has the potential to further improve recognition rates in the future. For example, the relationship between head and hand position, body posture, facial expressions, etc. In addition, the data features of some locations on the hand do not contribute much to recognition, offering the possibility of simplifying the device in the future.

4.5 Summary

In this research, we build a Sign-Glove system to recognize sign language. By analyzing the process of sign language, we noticed that sign language is composed of both hand motion and hand shape in time. Therefore, we decided to use IMU to detect the hand motion part and the bending sensor to detect the hand shape part. Then, we combine this information and use the weighted DTW algorithm to fuse the features and recognize the sign language words. In the experiments, we verified the performance of the Sign-Glove system and obtained high recognition rates of sign language. Such a wearable glove system has the potential to greatly reduce the cost of communication for people with hearing impairment. In the future, with further improvements, we exchange the cables for wireless connections like BLE and Xbee. In addition, word-by-word sign language recognition was achieved, but sign language is often used to construct meaning through continuous use. We will replace the breadboard connection to Printed Circuit Board and Flexible Flat Cables connections to achieve more stable data collection over a long period of time in daily use. We hope to build a system capable of continuous sign language recognition in the future. A more concise system provides more convenient and complete sign language expressions.

Chapter 5

Wearables and Vision Fusion Methods

5.1 Introduction

Recognition of hand motion capture is an interesting topic. Hand motion can represent many gestures. In particular, sign language plays an important role in the daily lives of hearing-impaired people. About 2.5 billion people are expected to have some degree of hearing loss by 2050, according to the WHO, and more than 1 billion young people are at risk of permanent hearing loss [100]. In addition, due to the impact of infectious diseases in recent years, online communication has become important. Facilitating communication between sign language users and non-users via video calls remains a pertinent research focus. However, the intricate nature of sign language gestures presents challenges to achieving optimal recognition solely through wearable data gloves or camera-based systems.

Both wearable data gloves and camera-based systems have been extensively explored for sign language recognition. Bending sensors glove only focus on finger bending degree. Consequently, several sign language words exhibiting similar curvature patterns become indistinguishable. This limitation curtails the utility of such devices. Given the significance of hand and arm gestures in sign language, it is imperative for vision-based approaches to prioritize the extraction of key points data from the hands, thereby reducing interference from extraneous background elements. Occlusion presents a significant challenge to vision-based methodologies. During the acquisition of hand key points, monocular cameras may fail to capture certain spatial information due to inter-finger occlusions. Such occlusions often act as impediments, constraining the potential for enhancement in recognition accuracy. In gesture recognition, it is easy for fingers to block each other, objects to block hands, or even parts to be nearly blocked due to overexposure or too dark, resulting in unrecognizability. As shown in Figure 5.1, the occlusion problem is less effective in obtaining key points. Integration with bending sensors offers a solution, enabling precise measurement of finger angles, even in regions overlapped by external entities.

In this research, we integrate a wearable sensor-based system with a camera-based approach to enhance the precision of hand sign language capture. One inherent challenge in extracting skeletal information for sign language is addressing occlusions among fingers and accessing spatial data unattainable by standalone camera systems.

To address this, our proposed system leverages hand skeletons as delineated by MediaPipe for sign language prediction. We adopt a hybrid methodology, intertwining Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory

(BiLSTM) models, to bolster our sign language recognition capabilities. CNN is good at extracting relationships between features, and BiLSTMs are adept at temporal data feature comprehension, rendering them ideal for action-oriented tasks such as sign language interpretation. Through this CNN + BiLSTM amalgamation, we have achieved superior recognition accuracy compared to single-sensor solutions.

This Chapter contribution is shown below itemization.

Our devised system integrates visual and bending sensor inputs. Visual data is utilized to extract essential key points and joint angles while eliminating redundancy. This approach mitigates the influence of background and lighting variations, enhancing the system's generalizability and data efficiency. The flex sensor captures finger flexion patterns, enabling adaptability across diverse environments.

We amalgamated key point coordinates, finger joint angles, and curvature features, strategically combining multifaceted information at the feature level. This integration forms the foundation for our CNN-BiLSTM model, facilitating information synergy and effectively enhancing recognition rates.

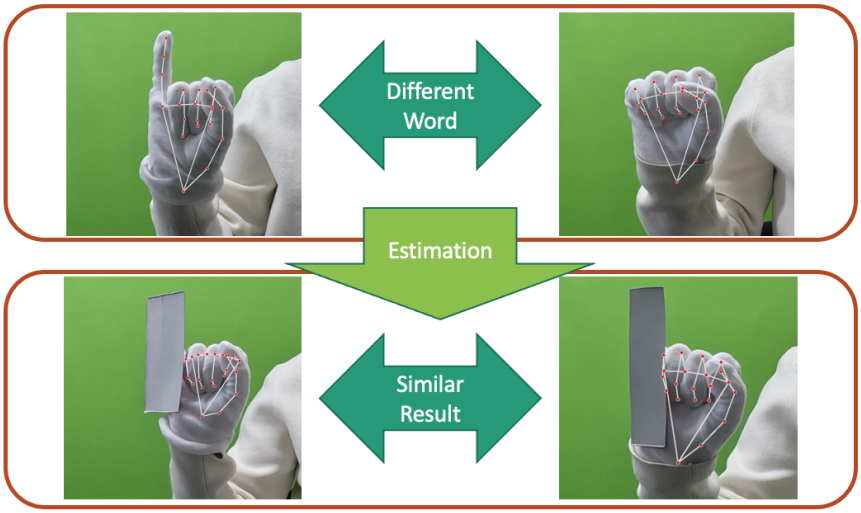

Figure 5.1: Occlusion Problem in Hand Sign Language

5.2 Method

The system simultaneously acquires data from bending sensors and vision and uses deep learning methods to fuse the data for sign language recognition.

The system simultaneously acquires data from bending sensors and vision and uses deep learning methods to fuse the data for sign language recognition.

The structure of the system is shown in Figure 5.2. The system contains two inputs, video collected by the camera and sensor data collected by the bending data glove. The camera data is used to obtain the key points of the hand through MediaPipe, and the joint angles of the fingers are obtained through the key points. Afterward, the joint angle data of the key point data and the finger bending angle of the sensor are spliced together, and the semantics of the sign language are obtained through CNN+BiLSTM recognition.

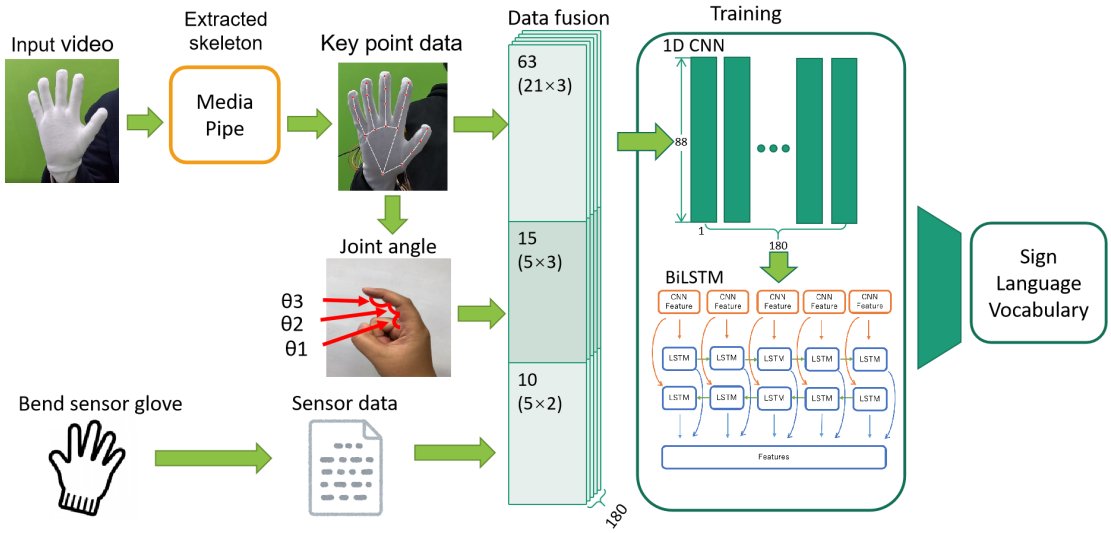

Figure 5.2: Method structure: Data collection and Training

5.2.1 2-axis bending sensor

The sensor used is a 2-axis bending sensor as shown in Figure 5.3 developed by Bend Labs. Compared to conventional sensors, this sensor measures angular displacement with higher accuracy in terms of power loss. The sensor output is the angular displacement as computed from the vectors defined by the ends of the sensor (v1 and v2). [101]

Figure 5.3: 2-axis bending sensor from Bend Labs

5.2.2 MediaPipe

We use MediaPipe to predict skeletons from images. MediaPipe can predict face, posture, and hand skeletons with high accuracy. This method is intended for use with GPUs for real-time inference. However, there are also lighter and heavier versions of the model to deal with CPU inference on mobile devices which is less accurate than running on desktops [102]. Fig:3.1 is the output of MediaPipe hand skeleton data. (a) are the predicted 21 keypoint positions. In (b), the points in (a) correspond to the

numbers. (c) is an example of using MediaPipe. In this research, 21 keypoints indicated by red dots are used as skeleton data and used as a dataset.

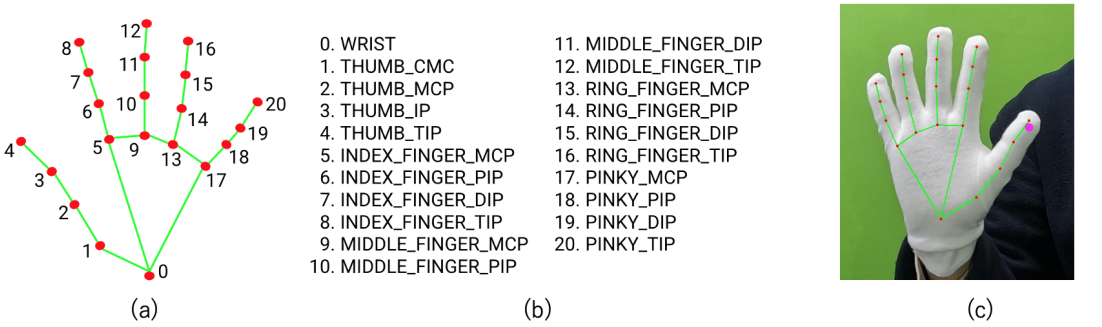

Figure 5.4: Skeleton and Bending Sensor Data Fusion

5.2.3 CNN+BiLSTM

Since video data is used for sign language recognition, a method that processes both spatial information and time series data is effective. Spatial information is learned using CNN, and time series information is learned using BiLSTM. First, a sign language dataset is input to MediaPipe. MediaPipe outputs the keypoint data of the sign language, which is used as skeleton data. The skeleton data is then input to the CNN to extract spatial information, and then temporal information is extracted by BiLSTM. The spatial and temporal information is learned and used as a model. By combining CNN and BiLSTM, we have achieved higher recognition accuracy by learning spatial and temporal features than only with one kind of them.

5.3 Implementation

5.3.1 Outline

The model of this sign language recognition system is shown in Figure 5.2. First, we build a data collection system, including data gloves and cameras that collect bending data. Then create a dataset. This data set contains video data and finger-bending data during sign language. Next, the hand skeleton is predicted from the sign language video. The hand skeleton is estimated using a MediaPipe. Finally, the sensor data and the skeletal data are fused and trained with CNN + BiLSTM. The model for gesture estimation is formed.

5.3.2 Bending Sensor Glove Structure

This part describes the design of the original glove, the sensors, the sensor controllers, and the sensor data structures. Figure 5.6 is the actual 2-axis bending sensor glove. Secure the fingertips and loosely secure the rest so the sensor does not come loose. Therefore, fixing parts was created with a 3D printer. The fingertip part is designed so that the sensor can be inserted and fixed. Also, if every part fixes the sensor,

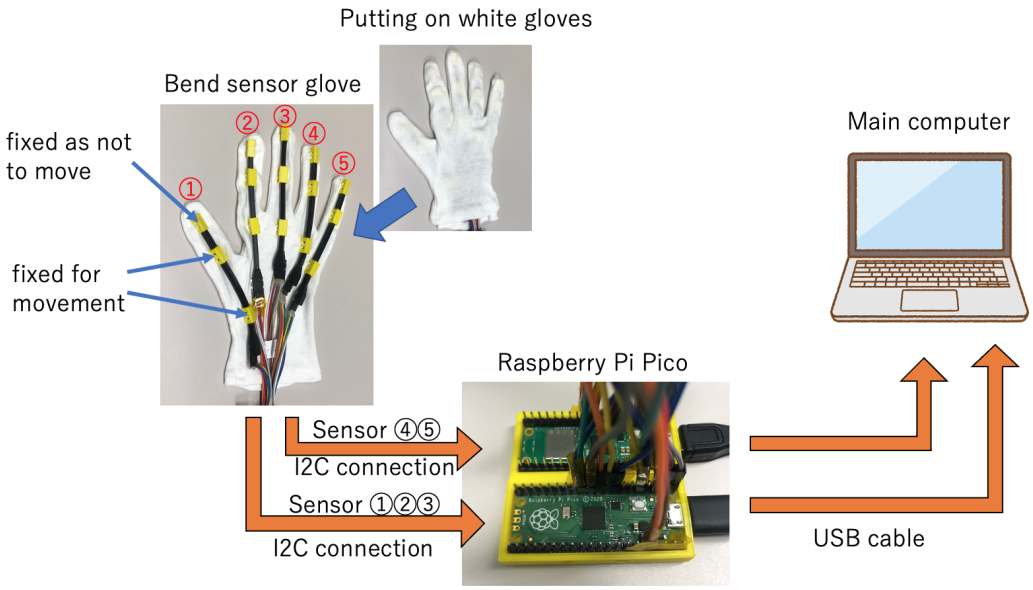

Figure 5.5: Sensor Glove Design

the movement of the finger will be restricted, making it impossible to express sign language. Therefore, the part other than the tip is not fixed. Also, when actually using it, wear white gloves to hide the sensor. This will prevent from MediaPipe not recognizing the sensor glove as a hand. Then Raspberry pi pico is used as a controller to control the sensor. Note that sensor gloves have different values depending on the person using the same hand pose.

5.3.3 Sign language Dataset

First, create a dataset of sign language videos to create skeleton data. The dataset is original data from laboratory members. Sign language words are used in 32 Japanese sign language vocabulary(SLV). The Japanese language is represented by 46 letters. They are represented by vowels (a, i, u, e, o), and consonants (k, s, t, n, h, m, y, r, w). The letter list used in this research is shown below 5.2. Japanese has letters with vowels only, vowels and consonants, and special characters represented by ”nn”. The table shows consonants in columns and vowels in rows. The first column from the right is for vowels only(”/” mean no consonants), and ”nn” appears at the end of the column for the consonant n.

5.3.4 Image Data Collection

The dataset has videos of 4 people for each word shot at 60fps with a green screen background. The sensor glove is put on the righ hand. sign language vocabulary is basically fixed, such as clenching a fist or raising the only index finger, and the hand is not moved. However, some sign language vocabulary is expressed by moving the hand. ”ri”, ”no”, ”nn”, and ”mo” in the wordlist table5.2. “mo” is a finger movement only, but “ri”, “no” and “nn” are expressed by moving the wrist.

Figure 5.6: Sensor Glove

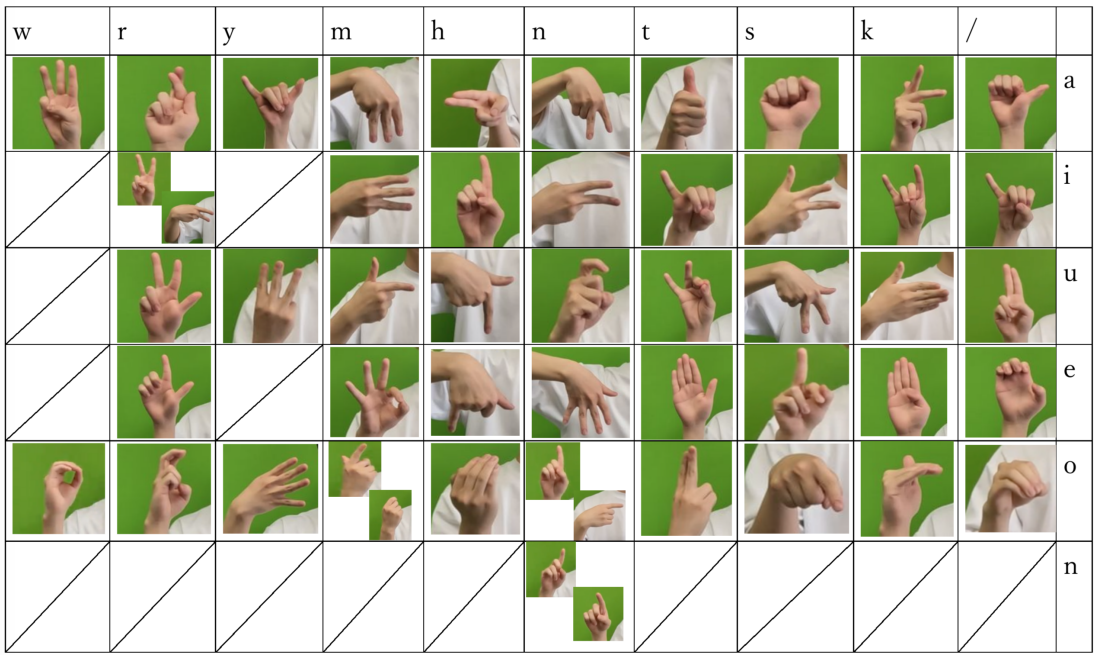

Figure 5.7: Japanese Sign Language Letter List

Key Point Estimation

We predict skeletal data from videos of sign language wearing sensor gloves. MediaPipe estimate 21 key points and make them skeleton data. Keypoint coordinates are 3D(x, y, z) and 60 frames are acquired per second.

Calculating joint angle

Calculate finger angles from skeleton data obtained with MediaPipe. This is useful for data argumentation of the dataset. There is one finger angle for each joint, and angles are calculated by the inner product. For example, to calculate the angle of the pinky finger, the keypoint k is predicted by the media pipe and calculated using

5.3.5 Collecting Sensor Data

This section describes the original Bending sensor glove and finger angle data collection. We made an original Bending sensor glove to collect finger angles. The glove is worn on the right hand. The data collected while wearing the glove is saved as a text file on the main computer along with time stamps and angles of five fingers 2 axis angles. In addition, a video of the sign language is also filmed at the same time as the bending sensor data is collected. The angle of the finger acquired at the same time as the bending sensor data and the image acquired at the same time support image recognition.

5.3.6 Data Fusion

Skeleton Data is acquired by MediaPipe, finger joint angles are calculated from Skeleton data, and sensor data is fused. Skeleton data is 63 (21 key points * 3 dimensions), finger joint angle is 15 (5 fingers * 3 joint angles), and sensor data is 10 (5 fingers * 2-axis).

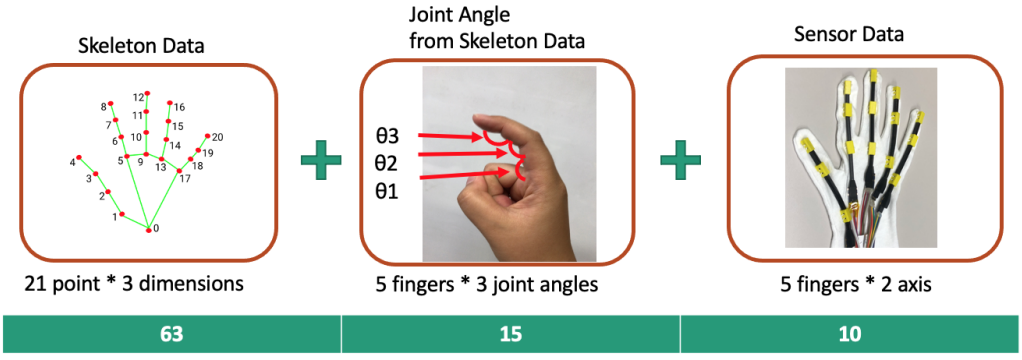

Figure 5.8: Data Fusion

5.4 Experiment and Evaluation

5.4.1 Experiment Purpose

In this experiment, we designed the experiment with the purpose of evaluating the sign language recognition performance of the fusion system of curved sensor gloves and computer vision. Experimental evaluation and discussion will be conducted by comparing the results of sign language recognition using only skeleton data and using all fused data.

5.4.2 Experiment Setting

We prepared a Bending sensor glove and a camera to collect data. The camera uses GoPro Hero10. High resolution, fast and small. Also, use a green screen for the background and unify the background colors. Wear sensor gloves and collect sensor data and video data. There are 32 sign language words. This data set contains 32 gestures from 4 people, with each gesture repeated 10 times.

To demonstrate the effectiveness of the sensor, occlusion is generated in the sign language vocabulary video and recognized by MediaPipe. First, we generate an occlusion in the image. Occlusion is expressed by randomly selecting the coordinates of the keypoint obtained with MediaPipe and displaying a black square on it. Occlusion is (80, 80) for image size (1080, 1920).

5.4.3 Experiment Process

The subject first collected stationary movements while maintaining a flat hand, providing calibration data for the glove. We turn on the gloves and the camera at the same time to obtain synchronized data. Subjects were guided through each gesture.

In the collected video data, we added black squares to people to simulate occlusion. Occlusion is generated at random positions for each video. Occlusion was generated by inserting black squares at random positions in the image. However, MediaPipe may not be able to get the Keypoint if an occlusion occurs. MediaPipe acquires skeleton data for each frame, but if the keypoint cannot be acquired in the first frame, the output result of MediaPipe is (x, y, z) = (-1, -1, -1), If the frame is in the middle, the output result is the value of the previous frame. Finally, the generated occlusion data is put into the dataset.

5.4.4 Experiment Results

The model was trained with k-Fold cross validation. For training with a small data set, the training accuracy during training could be higher. If this is the case, the accuracy in training may be high, but the accuracy in testing may be lowered, resulting in overfitting. To prevent this situation, there is a technique called k-Fold cross-validation. In k-Fold cross-validation, data is divided into k pieces, some of which are used for validation data and others for training data. Since all the divided data are used once for validation data, training is performed k times. The average of the k training accuracies is calculated as the result. The cross entropy method is calculated for the loss function. If the probability distributions of p and q are approximate, the cross-entropy loss is

smaller. In other words, the closer the learning accuracy approaches 1, the closer the result approaches 0. Results for the skeleton data only are shown below. There were 2282 number of samples extracted from the skeleton data. There are 261 test data, and the remaining data is training and evaluation data. Also, training data and evaluation data are split at a ratio of 4:1, there are 640 training data and 275 evaluation data. Training data is used for training, evaluation is used for evaluation during training, and test data is used for model evaluation.

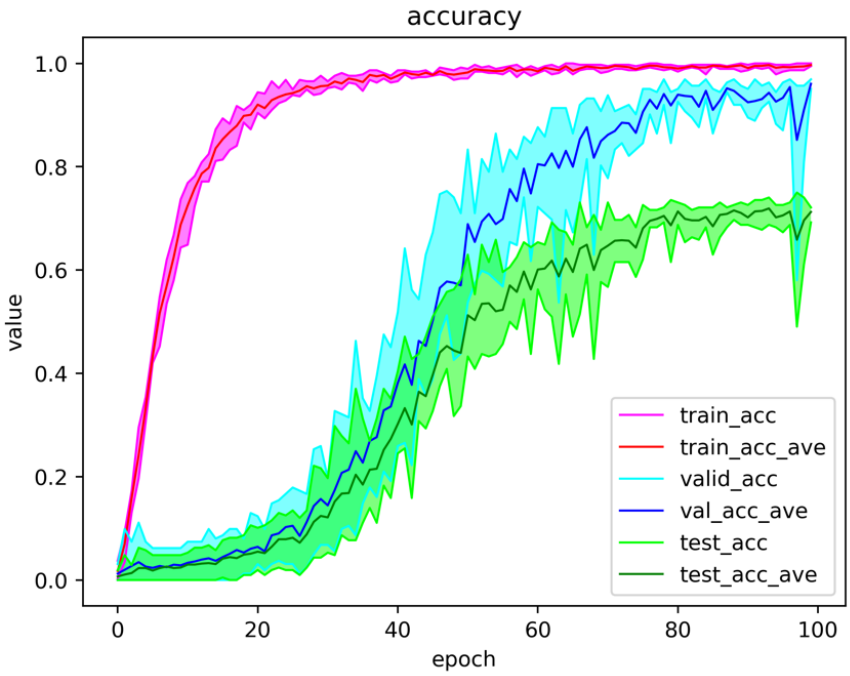

Figure 5.9: Accuracy curve of Only Skeleton Data

The training curve is shown below. The blue line shows the accuracy of Training, the orange line shows the accuracy of Validation, and the green line shows the accuracy of validation with Test data. For the Skeleton-only validation, cross-validation was performed 5 times, with an average training accuracy of 85.9% when training and 73.5% when using test data 6.3. For the Fusion data validation, cross-validation was performed 5 times, with an average training accuracy of 99.2% when training and 96.5% when using test data 6.5.

5.4.5 Discussion

The overall recognition rate of the fused system is improved compared to using only skeleton data. At the same time, the fused system uses fewer epochs to obtain a stable recognition rate and has lower overfitting. But there are some situations worth improving. First, some sign language movements are indistinguishable using only bending sensors. The values are exactly the same, which will cause a conflict in recognition judgments. In addition, when recognizing partially similar sign languages, the added sensor data values are similar, and the recognition results of some actions are lower. When the recognition effect using only skeleton data is poor, the sensor has a complementary effect.

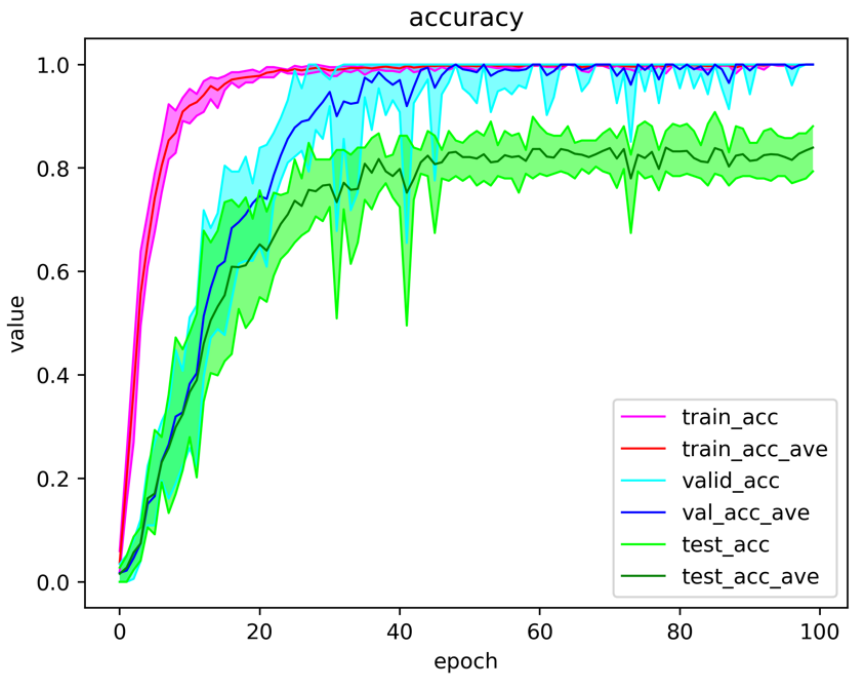

Figure 5.10: Accuracy curve of Fusion Data

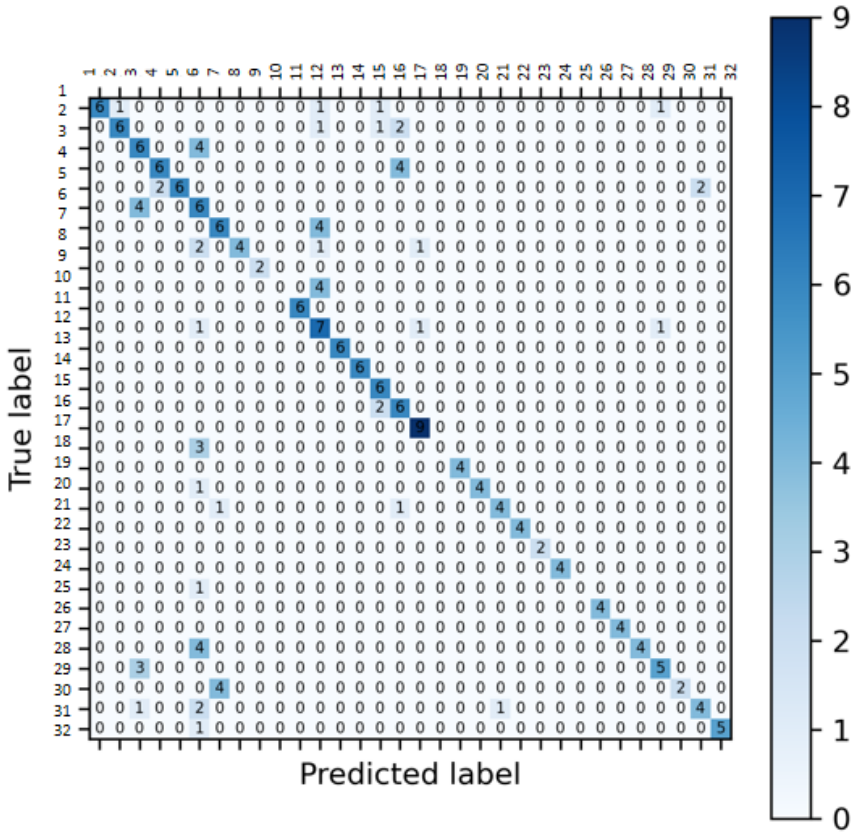

Figure 5.11: Confusion Matrix: Only Skeleton Data

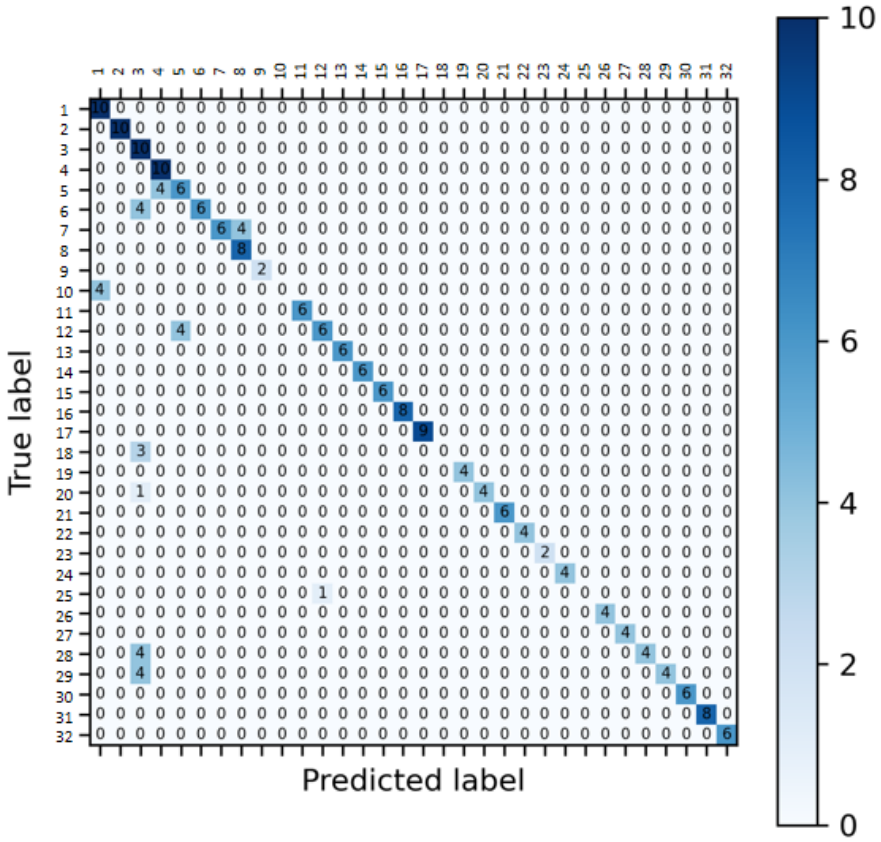

Figure 5.12: Confusion Matrix: Fusion Data

5.5 Summary

In this research, we aimed to improve sign language recognition with occlusion accuracy by combining CNN+BiLSTM and also combining bending sensor data with skeleton data. The combination of the CNN+BiLSTM method allowed us to perform finger character recognition better than using it alone. However, there were limitations in acquiring spatial information, such as blind spot problems. Therefore, we used a 2-axis bending sensor to assist with spatial information. The performance evaluation of the original 2-axis bending glove further strengthened the spatial information of sign language. By using sensor data, we were able to improve sign language recognition accuracy in the presence of occlusion compared to skeleton data alone. Our future task is to provide the system with complementary hand movement measurement data of more different modalities.

Chapter 6

Conclusion

In the study of hand movement measurement and recognition, it is imperative to acquire precise and consistent data. The inherent drift of the sensor can be mitigated through calibration techniques. By employing suitable sensor data fusion algorithms and capitalizing on data complementarity, noise can be minimized, thereby optimizing measurement attributes. It is further noted that the use of appropriate decision-making algorithms enhances reliability by harnessing redundancy and ensures decisions are congruent with the extant data in cases of decision conflicts.

In Chapter 3, we proposed a hand motion capture system for measuring hand joint angles. Data fusion is performed with the SRCKF, and StoS calibration is performed with a joint kinematic constraints method. A highly customizable hardware platform is constructed. The modular unit is easily expandable and directly taped to the back of the hand to maintain the intra-palmar tactile sensation of the hand. The errors are measured quantitatively by the optical system. The proposed method is a valuable approach for estimating finger joint angles by improving key steps to enhance accuracy. However, there are still many problems that need to be resolved. All non-invasive techniques, such as MIMU-based systems or marker-based optical systems, are subject to soft tissue artifact (STA) effects. A more ideal reference system would be direct magnetic resonance imaging (MRI) of the bone. MRI would be a powerful tool for further research on eliminating STA effects. Future work is needed to understand the soft tissue properties and the effect of skin on MIMU during movement.

In Chapter 4, we build a Sign-Glove system to recognize sign language. By analyzing the process of sign language, we noticed that sign language is composed of both hand motion and hand shape in time. Therefore, we decided to use IMU to detect the hand motion part and the bending sensor to detect the hand shape part. Then, we combine this information and use the weighted DTW algorithm to fuse the features and recognize the sign language words. In the experiments, we verified the performance of the Sign-Glove system and obtained high recognition rates of sign language. Such a wearable glove system has the potential to greatly reduce the cost of communication for people with hearing impairment. In the future, with further improvements, we exchange the cables for wireless connections like BLE and Xbee. In addition, word-by-word sign language recognition was achieved, but sign language is often used to construct meaning through continuous use. We will replace the breadboard connection with Printed Circuit Board and Flexible Flat Cables connections to achieve more stable data collection over a long period of time in daily use. We hope to build a system capable of continuous sign language recognition in the future. A more concise system provides

more convenient and complete sign language expressions.

In Chapter 5, we aimed to improve sign language recognition with occlusion accuracy by combining CNN+BiLSTM and also combining bending sensor data with skeleton data. The combination of the CNN+BiLSTM method allowed us to perform finger character recognition better than using it alone. However, there were limitations in acquiring spatial information, such as blind spot problems. Therefore, we used a 2-axis bending sensor to assist with spatial information. The performance evaluation of the original 2-axis bending glove further strengthened the spatial information of sign language. By using sensor data, we were able to improve sign language recognition accuracy in the presence of occlusion compared to skeleton data alone. Our future task is to provide the system with complementary hand movement measurement data of more different modalities.

References

[1] M. Yahya, J. A. Shah, K. Kadir, Z. M. Yusof, S. Khan, and A. Warsi, "Motion capture sensing techniques used in human upper limb motion: a review," *Sensor Review*, vol. 39, pp. 504–511, 2019.

[2] B. Bouvier, S. Duprey, L. Claudon, R. Dumas, and A. Savescu, "Upper limb kinematics using inertial and magnetic sensors: Comparison of sensor-to-segment calibrations," *Sensors (Basel, Switzerland)*, vol. 15, pp. 18 813 – 18 833, 2015.

[3] D. Fong and Y. Chan, "The use of wearable inertial motion sensors in human lower limb biomechanics studies: A systematic review," *Sensors (Basel, Switzerland)*, vol. 10, pp. 11 556 – 11 565, 2010.

[4] P. Picerno, "25 years of lower limb joint kinematics by using inertial and magnetic sensors: A review of methodological approaches." *Gait & posture*, vol. 51, pp. 239–246, 2017.

[5] A. Rashid and O. Hasan, "Wearable technologies for hand joints monitoring for rehabilitation: A survey," *Microelectron. J.*, vol. 88, pp. 173–183, 2019.

[6] 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 (Basel, Switzerland)*, vol. 20, 2020.

[7] Z. Yang, S. Yan, B. V. van Beijnum, B. Li, and P. Veltink, "Estimate hand–finger position with one magnetometer and known relative orientation," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[8] N. Ahmad, R. A. B. R. Ghazilla, N. M. Khairi, and V. Kasi, "Reviews on various inertial measurement unit (imu) sensor applications," in *SiPS 2013*, 2013.

[9] M. Tanenhaus, D. Carhoun, T. Geis, E. Wan, and A. Holland, "Miniature imu/ins with optimally fused low drift mems gyro and accelerometers for applications in gps-denied environments," *Proceedings of the 2012 IEEE/ION Position, Location and Navigation Symposium*, pp. 259–264, 2012.

[10] F. Wittmann, O. Lambercy, R. Gonzenbach, M. V. van Raai, R. Hover, J. Held, M. Starkey, A. Curt, A. Luft, and R. Gassert, "Assessment-driven arm therapy at home using an imu-based virtual reality system," *2015 IEEE International Conference on Rehabilitation Robotics (ICORR)*, pp. 707–712, 2015.

[11] C. Lu, S. Amino, and L. Jing, "Data glove with bending sensor and inertial sensor based on weighted dtw fusion for sign language recognition," *Electronics*, vol. 12, no. 3, 2023. [Online]. Available: https://www.mdpi.com/2079-9292/12/3/613

[12] K. Wang and G. Zhao, "Gesture recognition based on flexible data glove using deep learning algorithms," in *2023 4th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)*, 2023, pp. 113–117.

[13] X. Tang, Y. Liu, C. Lv, and D. Sun, "Hand motion classification using a multi-channel surface electromyography sensor," *Sensors*, vol. 12, no. 2, pp. 1130–1147, 2012.

[14] Y. Liu, X. Li, L. Yang, G. Bian, and H. Yu, "A cnn-transformer hybrid recognition approach for semg-based dynamic gesture prediction," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–16, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:258577572

[15] L. Chen, J. Fu, Y. Wu, H. Li, and B. Zheng, "Hand gesture recognition using compact cnn via surface electromyography signals," *Sensors (Basel, Switzerland)*, vol. 20, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:210945584

[16] Y. Wang, P. Zhao, and Z. Zhang, "A deep learning approach using attention mechanism and transfer learning for electromyographic hand gesture estimation," *Expert Syst. Appl.*, vol. 234, p. 121055, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:260323932

[17] C. Chen, Y. Yu, X. Sheng, J. Meng, and X. Zhu, "Real-time hand gesture recognition by decoding motor unit discharges across multiple motor tasks from surface electromyography," *IEEE Transactions on Biomedical Engineering*, vol. 70, no. 7, pp. 2058–2068, 2023.

[18] D. Xiong, D. Zhang, X. Zhao, and Y. Zhao, "Deep learning for emg-based human-machine interaction: A review," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 3, pp. 512–533, 2021.

[19] C. Fang, B. He, Y. Wang, J. Cao, and S. Gao, "Emg-centered multisensory based technologies for pattern recognition in rehabilitation: State of the art and challenges," *Biosensors*, vol. 10, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:220853742

[20] R. V. Vitali and N. Perkins, "Determining anatomical frames via inertial motion capture: A survey of methods." *Journal of biomechanics*, vol. 106, p. 109832, 2020.

[21] L. Pacher, C. Chatellier, R. Vauzelle, and L. Fradet, "Sensor-to-segment calibration methodologies for lower-body kinematic analysis with inertial sensors: A systematic review," *Sensors (Basel, Switzerland)*, vol. 20, 2020.

[22] W. Teufl, M. Miezal, B. Taetz, M. Fröhlich, and G. Bleser, "Validity, test-retest reliability and long-term stability of magnetometer free inertial sensor based 3d joint kinematics," *Sensors (Basel, Switzerland)*, vol. 18, 2018.

[23] A. Alizadegan and S. Behzadipour, "Shoulder and elbow joint angle estimation for upper limb rehabilitation tasks using low-cost inertial and optical sensors," *Journal of Mechanics in Medicine and Biology*, vol. 17, p. 1750031, 2017.

[24] D. Laidig, D. Lehmann, M.-A. Bégin, and T. Seel, "Magnetometer-free realtime inertial motion tracking by exploitation of kinematic constraints in 2-dof joints," *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1233–1238, 2019.

[25] M. Valtin, C. Salchow, T. Seel, D. Laidig, and T. Schauer, "Modular finger and hand motion capturing system based on inertial and magnetic sensors," *Current Directions in Biomedical Engineering*, vol. 3, pp. 19 – 23, 2017.

[26] C. Salchow-Hömmen, L. Callies, D. Laidig, M. Valtin, T. Schauer, and T. Seel, "A tangible solution for hand motion tracking in clinical applications," *Sensors (Basel, Switzerland)*, vol. 19, 2019.

[27] H. G. Kortier, V. I. Sluiter, D. Roetenberg, and P. H. Veltink, "Assessment of hand kinematics using inertial and magnetic sensors," *Journal of NeuroEngineering and Rehabilitation*, vol. 11, pp. 70 – 70, 2013.

[28] P. Müller, M. Begin, T. Schauer, and T. Seel, "Alignment-free, self-calibrating elbow angles measurement using inertial sensors," *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pp. 583–586, 2016.

[29] D. Laidig, P. Müller, and T. Seel, "Automatic anatomical calibration for imu-based elbow angle measurement in disturbed magnetic fields," *Current Directions in Biomedical Engineering*, vol. 3, pp. 167 – 170, 2017.

[30] K. Kitano, A. Ito, and N. Tsujiuchi, "Hand motion measurement using inertial sensor system and accurate improvement by extended kalman filter," *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6405–6408, 2019.

[31] B.-S. Lin, I.-J. Lee, S.-Y. Yang, Y.-C. Lo, J. Lee, and J.-L. Chen, "Design of an inertial-sensor-based data glove for hand function evaluation," *Sensors (Basel, Switzerland)*, vol. 18, 2018.

[32] B.-S. Lin, I.-J. Lee, P.-Y. Chiang, S.-Y. Huang, and C. wei Peng, "A modular data glove system for finger and hand motion capture based on inertial sensors," *Journal of Medical and Biological Engineering*, vol. 39, pp. 532–540, 2019.

[33] S. Madgwick, A. J. L. Harrison, and R. Vaidyanathan, "Estimation of imu and marg orientation using a gradient descent algorithm," *2011 IEEE International Conference on Rehabilitation Robotics*, pp. 1–7, 2011.

[34] T. Seel and S. Ruppin, "Eliminating the effect of magnetic disturbances on the inclination estimates of inertial sensors," *IFAC-PapersOnLine*, vol. 50, pp. 8798–8803, 2017.

[35] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

[36] T. L. Ainscough, R. Zanetti, J. Christian, and P. Spanos, "Q-method extended kalman filter," *Journal of Guidance Control and Dynamics*, vol. 38, pp. 752–760, 2015.

[37] Z. Dai and L. Jing, "Real-time attitude estimation of sigma-point kalman filter via matrix operation accelerator," *2019 IEEE 13th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC)*, pp. 342–346, 2019.

[38] Z. Yang, B. F. Beijnum, B. Li, S. Yan, and P. Veltink, "Estimation of relative hand-finger orientation using a small imu configuration," *Sensors (Basel, Switzerland)*, vol. 20, 2020.

[39] S. Julier and J. Uhlmann, "Unscented filtering and nonlinear estimation," *Proceedings of the IEEE*, vol. 92, pp. 401–422, 2004.

[40] I. Arasaratnam and S. Haykin, "Cubature kalman filters," *IEEE Transactions on Automatic Control*, vol. 54, pp. 1254–1269, 2009.

[41] I. Arasaratnam, S. Haykin, and T. Hurd, "Cubature kalman filtering for continuous-discrete systems: Theory and simulations," *IEEE Transactions on Signal Processing*, vol. 58, pp. 4977–4993, 2010.

[42] R. V. D. Merwe and E. A. Wan, "Sigma-point kalman filters for integrated navigation," 2004.

[43] X. Tang, J. Wei, and K. Chen, "Square-root adaptive cubature kalman filter with application to spacecraft attitude estimation," *2012 15th International Conference on Information Fusion*, pp. 1406–1412, 2012.

[44] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *ArXiv*, vol. abs/1511.08458, 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:9398408

[45] S. Zhang, D. Zheng, X. Hu, and M. Yang, "Bidirectional long short-term memory networks for relation classification," in *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation*, 2015, pp. 73–78.

[46] J. P. C. Chiu and E. Nichols, "Named entity recognition with bidirectional lstm-cnns," *Transactions of the Association for Computational Linguistics*, vol. 4, pp. 357–370, 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:6300165

[47] P. Kavianpour, M. Kavianpour, E. Jahani, and A. Ramezani, "A cnn-bilstm model with attention mechanism for earthquake prediction," *ArXiv*, vol. abs/2112.13444, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:245501929

[48] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, and M. Grundmann, "Mediapipe: A framework for building perception pipelines," *ArXiv*, vol. abs/1906.08172, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:195069430

[49] M. S. Amin, S. T. H. Rizvi, and M. M. Hossain, "A comparative review on applications of different sensors for sign language recognition," *Journal of Imaging*, vol. 8, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:247944115

[50] M. Al-Qurishi, T. Khalid, and R. Souissi, "Deep learning for sign language recognition: Current techniques, benchmarks, and open issues," *IEEE Access*, vol. 9, pp. 126 917–126 951, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:237549703

[51] C. Lu, Z. Dai, and L. Jing, "Measurement of hand joint angle using inertial-based motion capture system," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–11, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:256187189

[52] M. A. A. Faisal, F. F. Abir, M. U. Ahmed, and M. A. R. Ahad, "Exploiting domain transformation and deep learning for hand gesture recognition using a low-cost dataglove," *Scientific Reports*, vol. 12, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:254628522

[53] C. Lu, S. Amino, and L. Jing, "Data glove with bending sensor and inertial sensor based on weighted dtw fusion for sign language recognition," *Electronics*, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:256349052

[54] M. Zakariah, Y. A. Alotaibi, D. Koundal, Y. Guo, and M. M. Elahi, "Sign language recognition for arabic alphabets using transfer learning technique," *Computational Intelligence and Neuroscience*, vol. 2022, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:248327632

[55] N. Mukai, S. Yagi, and Y. Chang, "Japanese sign language recognition based on a video accompanied by the finger images," *2021 Nicograph International (NicoInt)*, pp. 23–26, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:237327603

[56] G. H. Samaan, A. R. Wadie, A. K. Attia, A. M. Asaad, A. E. Kamel, S. O. Slim, M. S. Abdallah, and Y.-I. Cho, "Mediapipe's landmarks with rnn for dynamic sign language recognition," *Electronics*, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:252823458

[57] P. Purkait, C. Zach, and I. D. Reid, "Seeing behind things: Extending semantic segmentation to occluded regions," *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1998–2005, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:174801463

[58] D. Avola, M. Bernardi, L. Cinque, G. L. Foresti, and C. Massaroni, "Exploiting recurrent neural networks and leap motion controller for the recognition of sign language and semaphoric hand gestures," *IEEE Transactions on Multimedia*, vol. 21, pp. 234–245, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:4379732

[59] 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," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10 825–10 834, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:67855274

[60] Z. Zhang, "Microsoft kinect sensor and its effect," *IEEE Multim.*, vol. 19, pp. 4–10, 2012. [Online]. Available: https://api.semanticscholar.org/CorpusID:8629444

[61] J. Guna, G. Jakus, M. Pogacnik, S. Tomažič, and J. Sodnik, "An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking," *Sensors (Basel, Switzerland)*, vol. 14, pp. 3702 – 3720, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:11452648

[62] L. Dipietro, A. Sabatini, and P. Dario, "A survey of glove-based systems and their applications," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 38, pp. 461–482, 2008.

[63] 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 (Basel, Switzerland)*, vol. 21, 2021.

[64] J. Henderson, J. Condell, J. P. Connolly, D. Kelly, and K. Curran, "Review of wearable sensor-based health monitoring glove devices for rheumatoid arthritis," *Sensors (Basel, Switzerland)*, vol. 21, 2021.

[65] M. Borghetti, E. Sardini, and M. Serpelloni, "Sensorized glove for measuring hand finger flexion for rehabilitation purposes," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, pp. 3308–3314, 2013.

[66] L. Almeida, E. Lopes, B. Yalçinkaya, R. Martins, A. Lopes, P. Menezes, and G. Pires, "Towards natural interaction in immersive reality with a cyber-glove," *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 2653–2658, 2019.

[67] C. Yi, F. Jiang, C. Yang, Z. Chen, Z. Ding, and J. Liu, "Reference frame unification of imu-based joint angle estimation: The experimental investigation and a novel method," *Sensors (Basel, Switzerland)*, vol. 21, 2021.

[68] G. Wu, F. V. D. van der Helm, H. Veeger, M. Makhsous, P. V. Roy, C. Anglin, J. Nagels, A. Karduna, K. McQuade, X. Wang, F. Werner, and B. Buchholz, "Isb recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motion–part ii: shoulder, elbow, wrist and hand." *Journal of biomechanics*, vol. 38 5, pp. 981–992, 2005.

[69] B. G. de Monsabert, J. Visser, L. Vigouroux, F. V. D. van der Helm, and H. Veeger, "Comparison of three local frame definitions for the kinematic analysis of the fingers and the wrist." *Journal of biomechanics*, vol. 47 11, pp. 2590–7, 2014.

[70] J. B. Kuipers, *Quaternions and rotation sequences : a primer with applications to orbits, aerospace, and virtual reality*. Princeton, NJ: Princeton Univ. Press, 1999. [Online]. Available: http://www.worldcat.org/title/quaternions-and-rotation-sequences-a-primer-with-applications-to-orbits-aerospace-and-virtua oclc/246446345

[71] Z. Dai and L. Jing, "Lightweight extended kalman filter for marg sensors attitude estimation," *IEEE Sensors Journal*, pp. 1–1, 2021.

[72] T. Seel, J. Raisch, and T. Schauer, "Imu-based joint angle measurement for gait analysis," *Sensors (Basel, Switzerland)*, vol. 14, pp. 6891 – 6909, 2014.

[73] J. Fang, H. Sun, J. Cao, X. Zhang, and Y. Tao, "A novel calibration method of magnetic compass based on ellipsoid fitting," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 6, pp. 2053–2061, 2011.

[74] X. Zhang, S. Lee, and P. Braido, "Determining finger segmental centers of rotation in flexion-extension based on surface marker measurement." *Journal of biomechanics*, vol. 36 8, pp. 1097–102, 2003.

[75] H. Moon, "Capturing human hand kinematics for object grasping and manipulation," 2013.

[76] Z. Dai, C. Lu, and L. Jing, "Time drift compensation method on multiple wireless motion capture nodes," *2020 13th International Conference on Human System Interaction (HSI)*, pp. 266–271, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:220669018

[77] J. C. van den Noort, H. Kortier, N. van Beek, D. Veeger, and P. Veltink, "Measuring 3d hand and finger kinematics—a comparison between inertial sensing and an opto-electronic marker system," *PLoS ONE*, vol. 11, 2016.

[78] M. of Health. Labour and W. H. Page, "2016 survey on difficulty in life (nationwide fact finding survey on children with disabilities at home) results," 2023. [Online]. Available: https://www.mhlw.go.jp/toukei/list/dl/seikatsu_chousa_c_h28.pdf

[79] R. Rastgoo, K. Kiani, and S. Escalera, "Sign language recognition: A deep survey," *Expert Syst. Appl.*, vol. 164, p. 113794, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:224880147

[80] S. Jiang, P. Kang, X. Song, B. P. L. Lo, and P. B. Shull, "Emerging wearable interfaces and algorithms for hand gesture recognition: A survey," *IEEE Reviews in Biomedical Engineering*, vol. 15, pp. 85–102, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:233998626

[81] A. Seçkin, "Multi-sensor glove design and bio-signal data collection," *Natural and Applied Sciences Journal*, vol. 3, no. Special Issue: Full Papers of 2nd International Congress of Updates in Biomedical Engineering, pp. 87 – 93, 2021.

[82] M. Seçkin, A. Çağdaş Seçkin, and Çetin Gençer, "Biomedical sensors and applications of wearable technologies on arm and hand," *Biomedical Materials & Devices*, pp. 1–13, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:251361217

[83] N. Aloysius and Geetha, "Understanding vision-based continuous sign language recognition," *Multimedia Tools and Applications*, vol. 79, pp. 22 177 – 22 209, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:218657614

[84] S. Sharma and S. Singh, "Vision-based sign language recognition system: A comprehensive review," *2020 International Conference on Inventive Computation Technologies (ICICT)*, pp. 140–144, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:219592778

[85] Y. Ma, G. Zhou, S. Wang, H. Zhao, and W. Jung, "Signfi: Sign language recognition using wifi," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, pp. 23:1–23:21, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:4362611

[86] W. He, K. Wu, Y. Zou, and Z. Ming, "Wig: Wifi-based gesture recognition system," *2015 24th International Conference on Computer Communication and Networks (ICCCN)*, pp. 1–7, 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:4341610

[87] K. Kudrinko, E. Flavin, X. Zhu, and Q. Li, "Wearable sensor-based sign language recognition: A comprehensive review," *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 82–97, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:221343742

[88] P. M. Lokhande, R. Prajapati, and S. Pansare, "Data gloves for sign language recognition system," 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:19620821

[89] M. Kohei, C. Youngha, and M. Nobuhiko, "Recognition of fingerspelling in japanese sign language based on nail detection and wrist position," 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:146113366

[90] M. Salagar, P. Kulkarni, and S. Gondane, "Implementation of dynamic time warping for gesture recognition in sign language using high performance computing," *2013 International Conference on Human Computer Interactions (ICHCI)*, pp. 1–6, 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:17010990

[91] E. Korzeniewska, M. Kania, and R. Zawiślak, "Textronic glove translating polish sign language," *Sensors (Basel, Switzerland)*, vol. 22, 2022. [Online]. Available: https://api.semanticscholar.org/CorpusID:252256076

[92] Y. Na, H. Yang, and J. Woo, "Classification of the korean sign language alphabet using an accelerometer with a support vector machine," *J. Sensors*, vol. 2021, pp. 9 304 925:1–9 304 925:10, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:237258029

[93] S. Tateno, H. Liu, and J. Ou, "Development of sign language motion recognition system for hearing-impaired people using electromyography signal," *Sensors (Basel, Switzerland)*, vol. 20, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:223546974

[94] S. A. Khomami and S. Shamekhi, "Persian sign language recognition using imu and surface emg sensors," *Measurement*, vol. 168, p. 108471, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:225024512

[95] K. S. Abhishek, L. C. F. Qubeley, and D. Ho, "Glove-based hand gesture recognition sign language translator using capacitive touch sensor," *2016 IEEE International Conference on Electron Devices and Solid-State Circuits (EDSSC)*, pp. 334–337, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:35855408

[96] J. Gałka, M. Masior, M. Zaborski, and K. Barczewska, "Inertial motion sensing glove for sign language gesture acquisition and recognition," *IEEE Sensors Journal*, vol. 16, pp. 6310–6316, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:27003849

[97] M. A. A. Faisal, F. F. Abir, and M. U. Ahmed, "Sensor dataglove for real-time static and dynamic hand gesture recognition," *2021 Joint 10th International Conference on Informatics, Electronics & Vision (ICIEV) and 2021 5th International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, pp. 1–7, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:239038814

[98] B. G. Lee and S. M. Lee, "Smart wearable hand device for sign language interpretation system with sensors fusion," *IEEE Sensors Journal*, vol. 18, pp. 1224–1232, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:45573954

[99] X. Chu, J. Liu, and S. Shimamoto, "A sensor-based hand gesture recognition system for japanese sign language," *2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech)*, pp. 311–312, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:233227138

[100] WorldHealthOrganization, "World report on hearing," 2021, available online: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss.

[101] "Soft angular displacement sensor theory manual," 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:204894965

[102] 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*, vol. abs/2006.10214, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:219792872