

Hand tracking and interaction by computer vision for patients rehabilitation

Master Thesis



Hand tracking and interaction by computer vision for patients rehabilitation

Master Thesis
January, 2022

By
Shuai Wang

Copyright: Reproduction of this publication in whole or in part must include the customary bibliographic citation, including author attribution, report title, etc.

Cover photo: Shuai Wang , 2022

Published by: DTU, Department of Applied Mathematics and Computer Science, Bygning 324, 2800 Kgs. Lyngby Denmark
www.compute.dtu.dk

ISSN: [0000-0000] (electronic version)

ISBN: [000-00-0000-000-0] (electronic version)

ISSN: [0000-0000] (printed version)

ISBN: [000-00-0000-000-0] (printed version)

Approval

This thesis has been prepared over five months during the fall semester at Department of Applied Mathematics and Computer Science, at the Technical University of Denmark, DTU, in partial fulfilment for the degree Master of Science in Mathematical Modelling and Computation.

Kongens Lyngby, January 2, 2022

A handwritten signature in black ink, appearing to read "Shuai Wang". The signature is fluid and cursive, with the characters slightly overlapping.

Shuai Wang - S200108

Abstract

With the development of artificial intelligence, highly accurate and real-time human tracking on mobile devices has become possible. Hand and body dexterity can be affected by stroke, but there is no affordable and convenient method to provide a long-term quantitative approach to track body dexterity.

In this thesis, we use a MediaPipe model trained by deep learning to achieve real-time, distance invariant tracking of hand and arm body landmarks. The purpose is to explore if mobile devices are feasible for at-home monitoring of patients during rehabilitation. First, by allowing the user's hand to do peak-valley movements (e.i., opening and closing the hand in front of the camera), the landmark of the hand is collected using landmark-based registration. Then, a long-term change in hand dexterity that a patient may experience during rehabilitation, is simulated by using of a number of rubber bands with different degrees of kinetic constraints. The changes in hand mobility are evaluated by comparing and analyzing the change in the hand area visible to the camera and the movement speed of the fingers. Second, in a mirror hand test, the motion data of the left hand and right hand are compared. Third, MeidaPipe landmarks of arm movements are analyzed to test if they could potentially allow monitoring changes in arm strength over time. The changes are simulated by using of different dumbbell weights.

The rubber band test found that it was possible to measure the difference in hand dexterity. With the use of more than two rubber bands, the hand peak area and speed will decrease with the increase of rubber bands, the test results are able to fulfill two-sample t-test, $p - value < 0.05$. The mirror hand test indicated that under the condition of equal external force limitation in the left and right hand comparison experiment, there is a significant difference between the dexterity of the two hands in hand peak area and speed. The dumbbell arm test suggested that the MediaPipe Pose model is able to detect differences between arm muscle strengths by detecting movement speed, but requires external force limitations to reach a specific level.

Finally, this thesis reflects on how mobile devices may provide input to a digital patient twin by offering a convenient, remote quantification of patients' physical dexterity for clinical professionals, and a user-friendly method for stroke patients to record the progress of their rehabilitation with their own phones or tablets.

All code and test data for this paper is open sourced on Github: https://github.com/lingwsh/hand_track_mediapipe.

Acknowledgements

Author: Shuai Wang, MSc of Mathematical Modelling and Computation, DTU Compute

Supervisor¹: John Paulin Hansen, Professor of Department of Technology, Management and Economics, DTU Management

Supervisor²: Milton Edgardo Mariani, PhD student of Management and Economics, Management and Economics, DTU Management

I would like to thank my supervisors Professor John Paulin Hansen and Milton Edgardo Mariani for their patient guidance during the writing of this thesis. In addition, great thanks to my parents for their encouragement and financial support during my graduate studies. Finally, the thesis was completed under the influence of Covid-19. I would also like to thank my close friends for their help during this difficult time.

Contents

Preface	ii
Abstract	iii
Acknowledgements	iv
1 Introduction	1
1.1 Motivation	1
1.2 Research Directions	1
1.3 Main contributions	2
1.4 Structure of the thesis	2
2 Goals and research questions	3
3 Background and related work	5
3.1 Background	5
3.2 Related work	5
4 Theory and Methodology	8
4.1 Digital twin	8
4.1.1 Origins and evolution	8
4.1.2 Digital twin assistant system	9
4.1.3 DTA characteristics	10
4.2 MediaPipe	11
4.2.1 Hands	11
4.2.2 Pose	12
4.3 Landmark based registration	12
4.3.1 Introduction	12
4.3.2 Elements of image registration	15
4.3.3 Landmark and label	15
4.3.4 Image transformation and FRE	16
4.4 Hand area tracking procedures	18
4.4.1 Hand area calculation	18
4.4.2 Data acquisition	18
4.4.3 Description of data set	19
4.4.4 Data Re-display	20
4.4.5 Head area peak/valley noise remove	21
4.4.6 Hand peak/valley speed	23
4.4.7 One group landmark based registration	24
4.4.8 Multi-group landmark based registration	26
5 Experiment	27
5.1 Hand area tracking experiment	27
5.1.1 Single hand tracking with rubber band constraints	27
5.1.2 Mirror hand compare	29
5.2 Dumb-bell exercise arm angle tracking Experiment	31
5.3 MeidaPipe model performance and stability	33

6 Numerical Result	35
6.1 Hand area tracking experiment analysis	35
6.1.1 Different rubber band data analysis	35
6.1.1.1 Mix condition compare	36
6.1.1.2 Stable condition compare	37
6.1.1.3 Fluctuate condition compare	37
6.1.1.4 Registration based group compare	41
6.1.2 Different people tracking analysis	45
6.1.3 Mirror hand data analysis	47
6.2 Dumb-bell exercise analysis	49
7 Discussion	51
8 Conclusion	53
Bibliography	54
A Appendices	58
A.1 Hand area data	58

1 Introduction

1.1 Motivation

Stroke is a global health problem and is the second commonest cause of death and a leading cause of adult disability worldwide [1]. There will be more strokes in the future, because of demographic problems and inadequate control of the main risk factors for stroke. This disease requires more research attention to prevent it [2]. Paresis affecting the upper extremity is a common motor impairment in patients with stroke [3].

However, the hand is the commonly used body part in daily life. If we can track and evaluate the flexibility of the patient's hand, and then pass tests to evaluate the patient's ability to use tools and do daily life activities, such as eating, brushing teeth, washing the face and so on. However, these behaviors are abstract behaviors that are difficult to evaluate for stroke patients. These abstract life actions, such as eating, can be described by the position, gesture, and speed of the hand in the 3D space. Through 2D camera combined with AI computer vision technology, it can be used to record these data. The hand data in 3D space is used as the basis for designing a method for collecting stroke patients' data, which can be used to assess the health of stroke patients' hands and motivate them to do better rehabilitation training.

Digital Twins(DT) are defined as “digital replications of living as well as nonliving entities that data to be seamlessly transmitted between the physical and virtual worlds” [4]. For example, the Digital twin of hands and arms in 3D space through gestures can be used to collect user hand movement data, and vibration and sound feedback can also be used to make rehabilitation training more interesting, thereby promoting stroke patients to obtain better rehabilitation results.

The digital twin assistant(DTA) system can track arms and hands in real time. DTA can estimate arm and hand dexterity ability. Although real-time analysis, the output result can also give an instruction signal to remote assess for rehabilitation trainers and doctors, it could help the doctor or trainer to give real-time instruction to the patient in the rehabilitation progress.

DT data can be used to reconstruct 2.5D hand models of stroke patients' hands, for example, to show the comparison models of hand movement in different time periods. User gesture data can be collected in real time, allowing doctors or rehabilitation trainers to give real-time observation and guidance.

1.2 Research Directions

With the continuous development of artificial intelligence technology, new deep learning models have been able to run real-time human body joint tracking 3D motion data smoothly on common mobile devices, such as MediaPipe. The traditional tracking and localizing techniques still require multiple cameras for triangulation [5], depth cameras, radar data collection, and then long time computing by high performance computers to get the results. However, the new deep learning model MediaPipe can be used to track human motion data in real time with only a single RGB camera, and it can run on various platforms such as ordinary laptops, smartphones (IOS and Android) and web pages. Users do not need to spend extra money on special capture devices, and can use this method to capture data at any location.

Stroke patients need special rehabilitation to recover from upper limb paralysis. Stroke patients need physician guidance for rehabilitation, but the patient cannot feel the physician's movements directly on the limb. Our approach provides a DT-based digital model that captures the physician's movements and then transmits them remotely to the patient wearing the exoskeleton upper limb device in real time via the Internet, allowing them to feel the physician's hand movements: the raising and lowering of the arm, rotation of the wrist. This helps the patient to directly feel the rehabilitation movements directed by the physician.

Through rehabilitation, stroke patients can also compare changes during rehabilitation with historical data from the DT model. The results of the comparison can help the patient to better understand the results of the rehab and motivate them to do better rehab. At the same time, doctors can also get detailed data about the changes in the patient's body after rehabilitation, which can help to arrange the subsequent rehabilitation treatment plan.

1.3 Main contributions

The main contributions of this thesis are the following:

- A deep learning-based camera tracking method MediaPipe Hand for tracking and analyzing the dexterity of stroke patients' hands for comparing the recovery of patients' landmark hands' data after rehabilitation therapy for long-term recording and comparing. It is also used to motivate patients to better participate in rehabilitation therapy.
- A MediaPipe Pose based single RGB camera DTA system for real-time feedback of the human arm. The purpose is for rehabilitation doctors to better guide the rehabilitation training of stroke patients.

1.4 Structure of the thesis

The thesis is structured as follows:

- chapter 1 briefly introduces the thesis topic in three perspectives: Background and Motivation (section 1.1), Research directions (section 1.2) and Main contributions of the thesis (section 1.3).
- chapter 2 shows the main scientific research questions and answers that need to be studied in this paper.
- chapter 3 in section 3.1 analyze the main research background of the thesis, and in section 3.2 analyze the development and application of different hand tracking techniques in varied fields.
- chapter 5 contains the background knowledge and methodology needed for the technology in the thesis.
- chapter 6 discusses the results of different experiments in detail.
- In chapter 7, the conclusions of different experiments and the drawbacks of the method of this thesis are summarized and discussed.
- chapter 8: Overview of the main conclusions and strengths and weaknesses of the article.
- chapter A: Other relevant comparative data results.

2 Goals and research questions

The purpose of this thesis is a study on the physical condition assessment of stroke patients using deep learning based and computer vision. In the research process, it is important to consider how to make it easier for patients to use and not to increase the additional cost of using the device for the user. In the long-term, data collection at the same time to pay attention to the user's privacy protection.

The main goal is to record the user's body landmarks in real time through MeidaPipe and analyze the health status of the body through different experiments on different parts of the body landmark.

Therefore, the first research question of this master thesis is as follows.

How to build a Digital Twin to track the body for elderly stroke patients?

The original parameters of Digital Twins are mainly the extraction of human biometric features using different sensors, such as: blood pressure, heartbeat, and blood oxygen saturation. However, the physical condition of stroke patients cannot be directly reflected by these parameters, while using computer vision and deep learning is able to obtain results that cannot be directly obtained by traditional methods of sensors, such as: determining the health condition of the body based on its movement patterns. And it has the following advantages

1. the test methods mentioned in this article, such as hand Peak-Valley tracking, are simpler to use than professional medical devices, and the test results can be obtained using a common computer, smartphone or tablet. It is very convenient for stroke patients to use and record their own Digital Twin data independently or with the help of others.
2. Since both stroke patients and normal users will own at least one of the devices, such as a laptop, tablet, or smartphone, it is not necessary to purchase additional devices to record their health data in order to participate in the test.
3. Although we use the mobile device's camera to analyze video images and record the user's body landmark in real time, the data is only saved anonymously with the user's body landmark data, and no information about the user's image is stored. This approach allows users to manage and share their data with doctors and researchers, but without revealing their personal image privacy data.

How to use machine learning and computer vision to get human motion landmark parameters?

- Traditional methods to get human landmarks cannot provide high precision human joint landmarks and can only find the position of the hand. Methods based on deep camera and simultaneous multiview images require the use of high performance computers and the equipment is expensive. Usually these methods need to be used with professional computing software, which is very difficult for ordinary users in daily use.
- Deep learning based MediaPipe methods can run fluently on multiple platforms, such as: ordinary laptops, tablets and Android phones or IOS phones.

- Due to the rapid development of machine learning and machine vision, MediaPipe's model can now calculate and display the results in real time. The model also supports the fusion of different models and supports the simultaneous calculation and display of hand model, body model, and face model, which provides great convenience for subsequent data analysis.

How can the hand models be tracked for long-term analysis and comparison, so that stroke patients can better understand the results of their rehabilitation training?

- The MediaPipe hand model is used to track and record the Peak-Valley motion of the hand in real time, and to evaluate the change of hand dexterity by comparing the speed of hand motion and the area of the peak and valley. In the process of comparison, to avoid the effect of the change of the camera's far and near view on the hand area, this article uses landmark based registration to eliminate this effect.
- The above method not only compares the dexterity of the same hand of the user, but also compares the dexterity of the left and right hands by mirror transformation.
- The MediaPipe body model can track the landmark of the major joints of the whole body, and by recording the changes in speed and number of dumbbell exercises done by stroke patients, the recorded data can be used to compare the changes in arm muscle strength in a long-term situation.

3 Background and related work

3.1 Background

In the early stage of hand tracking, the main method of hand detection and gesture interface is based on feature extraction.[\[6\]](#) proposed a real-time 2D tracking method for hand monocular view maps. Kanade-Lucas-Tomasi (KLT) feature tracker is a method to compare two photos by finding different vector differences between them to see if the features in them are the same. This is a traditional method of image registration, but it is very costly for computer performance. With the help of the learned color distribution, a group of KLT features are retained and the hand position is determined by classifying the features. Although this method can be run indoors and outdoors, working with different people. But the output is only 2D coordinates of the hand position, and this phase belongs to the early onset of hand detection. [\[7\]](#) The method based on frequency analysis is used for instantaneous estimation of class separability without any training to classify hand gestures. Although it has classification capability for simple hand gestures, it is still significantly far from the complex gesture classification used in daily life.

Hand tracking application: Hand tracking is a popular topic in the machine vision and graphics communities, including 2D/3D gesture recognition using RGB/depth cameras, and hand model reconstruction. These techniques have been utilized in these domains [\[8\]–\[13\]](#). One real time hand and body digital twin system are designed for reducing construction worker risk in [\[14\]](#). This article offers construction workers the possibility to operate robots remotely through body and hand tracking technology. The LSTM ResNet model designed for prediction of hand grasp, which is based on real time RGBD video, is a new system of prosthesis control [\[15\]](#). But this method can not adapt well in new environment and new object, our method designs a robust and real time hand tracking method for arm rotation direction detection based on MediaPipe [\[16\]](#). These similar technologies can also be used by rehabilitation trainers and physicians for the remote rehabilitation of stroke patients. Combining the necessary exoskeleton equipment with the low-cost MediaPipe hand and body tracking model could provide more assistance to patients in the future and reduce the cost of traveling to doctors' appointments and rehab training.

However, for now, these applications, which are mainly designed to solve specific usage needs, are not designed to be optimized for stroke patients. Although in the [\[12\]](#), there is mention of Finger Dexterity, its main purpose is to explore Mid-air text entry. this cannot be directly used for long-term assessment of hand dexterity comparison in stroke patients. NU-Wrist is a preliminary mechanical design robot for wrist rehabilitation, which include wrist range safety design and Human-Robot axes Self-Alignment method [\[17\]](#). The proposed robot design provides rotation within the anatomical range of human wrist and forearm motions. A novel compliant robot handle link ensures dynamic passive self-alignment of human-robot axes during therapy exercising. This approach helps the stroke patient to better determine the range of motion of the hand, but then this approach requires specific equipment to be defined and only wrist motion can be measured.

3.2 Related work

It remains challenging due to inherent depth and scale ambiguities, diverse appearance variations, and complex joint and recovery space configurations. While a large number of existing works have considered markerless gesture recognition, most of them requiring

depth cameras [18]–[30] or simultaneous multiview images [20], [31]–[35] to solve the above challenges. Articles in Table 3.1 based on different techniques of hand tracking will be explored in this paper for the characteristics of the different methods. This paper will also explain why MediaPipe is used as a hand tracking method to discuss different rehabilitation applications.

The 3D hand pose could be got by generative adversarial network (GAN)¹ convolutional neural network(CNN)² with a single RGB 2D image[36]. Mobile GPUs are possible to calculate 2.5D hand landmarks by a special ML pipeline[16]. Hand tracking for 3D model by monochrome camera can achieve egocentric tracking[37].

However, most of these methods are not directly applicable in the real world where only monocular RGB images are available. Moreover, users cannot directly use depth cameras and synchronized multiview devices directly in their daily devices, which is a huge limitation for application scenarios. Monocular RGB camera solutions [16], [36], [38]–[43] have great potential to apply in daily life with a acceptable cost.

Table 3.1: Camera types in articles reviewed

Method	Articles
Depth camera	[18]–[29], [44]–[46]
Multi-view camera	[20], [31]–[35]
Monocular RGB Camera	[16], [36], [38]–[43], [47], [48]

Depth cameras: Real-time method [30] of gesture matching based on depth camera could provide a robust tracking method.In the article, the hand ROI detector is used to find the hand position then use Reinitializer combine particle swarm optimization (PSO) to renderer hand 3D model. It only supports single-handed 3D models. It can not be used for estimation of multi-hand interaction, while MediaPipe Hand [16] is able to achieve simultaneous tracking of multiple hands.Depth cameras, also used to build the hand tracking dataset, BigHand2.2M Benchmark [46] is a accuracy hand dataset for hand gesture recognizing.

With the popularity of deep learning, the use of neural networks and supervised learning can be used to get better results than traditional methods while having labeled data. A two CNNs method could produce 3D hand joint locations by RGB-D cameras in [44]. This paper presents a method for real-time, robust and accurate rate hand pose estimation by moving egocentric RGB-D cameras in a cluttered real environment. The method in the article uses two subsequently applied convolutional neural networks (CNNs) to localize the hand and regress 3D joint localization. By using the CNN to estimate the 2D position of the hand center in the input, hand localization can be achieved even in the presence of clutter and occlusion. To train the CNN, a new realistic dataset is introduced in the paper that uses a merged real approach to capture and synthesize annotated data from a large number of natural hand interactions in cluttered scenes. Later, fingertips can be localized by a trained Cycle-GAN model [48]. The Cycle-GAN [49] training materials do not need paired training data, which provide an easy way to find and train data by group category.

Multiview cameras: Deep learning could provide better key point estimation for optical marker-based tracking by 2D camera in [35]. The system in the article can efficiently and accurately collect a large amount of ground truth from the hand by using special gloves

¹https://en.wikipedia.org/wiki/Generative_adversarial_network

²https://en.wikipedia.org/wiki/Convolutional_neural_network

and multiple cameras for triangulation. This accurate data can help deep learning training to get better results. FreiHAND [43] uses multi camera to take different hand pose data in green space. It increases robust in real situation. FreiHAND also provides a 3D hand normalization method, which helps improve 3D space hand tracking accuracy.

Table 3.2: Hand tracking running platform and cost compare

Application	ContactPose [45]	Multiview Bootstrapping [34]	MediaPipe [16], [38]
Method	Depth camera	Multi-view camera	Single RGB Camrea
PC	Yes	Yes	Yes
Android	X	X	Yes
Apple IOS	X	X	Yes
Webpage	X	X	Yes
No extra cost	X	X	Yes

Monocular RGB Camer: InterHand2.6M [41] is a large 3D hand pose baseline dataset. It provides 3D hand joint coordinates by a single RGB camera.

In the path of single-camera hand tracking development, it will start to achieve the reconstruction of the hand 3D model with the help of some props. [47] uses a single camera to track gestures wearing colored gloves, and calculate the distance between the target and the database through Gaussian noise reduction and comparison with the model in the database. This article introduces a hand-tracking user input device composed of a single camera and a cloth glove. It demonstrated this device for several canonical 3-D manipulation and pose recognition tasks.

The method of hand tracking began to develop from the traditional method to the direction of deep learning. From the depth RGB camera, to the RGB camera, to the monochrome camera, the gesture 3D model can be tracked with less information with new theory development. Thanks to the development of neural network algorithms and mobile hardware performance increasing, mobile devices now is possible to have real-time gesture tracking capabilities.

4 Theory and Methodology

4.1 Digital twin

Digital twin(DT) becomes popular in industry application[50] and play a key role in satisfying customer needs and requirements[51]. Digital Twin Assistant (DTA) system is DT powered by different sensors and other new technologies, eg: Machine Learing(ML) and Deep Learning(DL) [52]. Our method is a DTA system designed for stroke or other disability patients based on ML and DL with a single RGB camera.

Successful rehabilitation is physically demanding as it involves testing the body's limits while attempting to improve physical conditions. For rehabilitation, the involvement of professional rehab trainers and doctors may be essential. This presents us with the same problem, sometimes there are not enough professionals to guide every person who needs rehabilitation, whether it is due to high demand, inaccessibility to rehab centers, or cost.

Papers that focus on rehabilitation target a different population than papers that focus on sports or welfare. The focus of this type of research is not on being competitive, as in sports, or on maintaining good health in everyday life, as in well-being. Rehabilitation papers focus on people who have lost their ability to exercise and how to help them recover. They may have suffered a stroke, have a movement disorder, or their age has taken its toll on their bodies. The DTA system is a way to track and compare historical data and at an acceptable cost by using the information collected as a quantitative assessment before and after rehabilitation training. Because the equipment for body tracking does not have to be purchased specifically, it reduces additional expenses for the patient.

4.1.1 Origins and evolution

Beginning in the early 2000s, the use of digital twins became popular in the digitization of machinery and production systems in manufacturing. For example, General Electric (GE) builds cloud-hosted digital twins for its machines, and these computers use artificial intelligence, physics-based models, and data analysis to process the information gathered from sensors to better manage these machines [4].

The concept of the digital twin can be applied to a wide range of technologies and therefore may disrupt industries beyond manufacturing. Therefore, it is critical to expand its definition. By seamlessly transferring data between the physical and virtual worlds, the digital twin will facilitate monitoring, understanding, and optimizing the functionality of all physical entities, living and nonliving.

For example, a biologist creating a tree's digital twin would be able to digitally examine its internal and external components; measure the amount of oxygen released, the sum of water, and the sunlight received; determine the age of the tree and track its growth from seedling to adult; and monitor and combat harmful pests or diseases before they spread to other parts of the tree.

The human digital twin can also collect and analyze physical, physiological, and contextual data to improve quality of life and enhance well-being. For example, strokes can be predicted before they happen so that preventive measures can be taken. Machine and deep learning techniques can also be used to detect lifestyles and predict potential health problems. In addition, contextual data, such as information about environment, age, emotional state, and preferences, can be collected and analyzed to fully understand

and characterize the user's overall situation. The purpose of using digital twins in this paper is to provide stroke patients with a platform that can collect physical movement data and medical data to help provide a system for sharing information between stroke patients and rehabilitation trainers and physicians.

4.1.2 Digital twin assistant system

The Digital twin assistant system is a cloud-based service that collects body movement data and biological data from stroke patients on a long-term basis, and then analyzes and stores the results and raw data at different time points, securely stored in a database using encrypted methods [Figure 4.1](#). Real-world physical characteristics and medical data are collected through 2D single RGB cameras on mobile devices (laptops, smartphones and tablets) and medical sensors (also smartwatches that can record basic biological information: e.g. Apple watch since Series 4 has heart monitoring electrocardiogram (EKG) and irregular heart rhythm, both are approved by US Food and Drug Administration at 2018). The digital twin of a stroke patient contains not only a replayable body landmark and bio-information. The system can be used to automatically compare models to see how the patient's body is performing. The system's feedback and viewing systems provide secure, authorized access to rehab trainers and physicians, who are kept informed of the patient's condition. The system can also push alerts directly to the patient and the relevant authorized viewers if something unexpected happens. Since the recording of user data can be done by the user, the frequency of recording can be significantly increased and reduced with the frequency of meeting with the doctor. Moreover, the results of the recorded body landmark can quantify the changes in the user's physical condition, which can help doctors better track and judge the patient's physical condition.

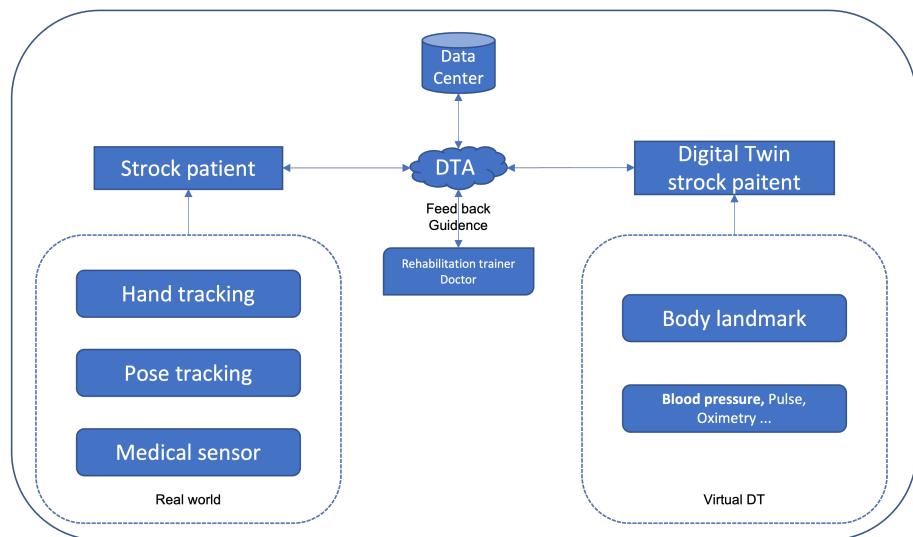


Figure 4.1: Digital twin assistant system

By storing the data over time, it is possible to analyze small changes in the physical function and health of stroke patients. Dedicated rehabilitation trainers and doctors can also access long-term exercise and biorecorded data of patients through DTA authorization, and by viewing the results and combining their expertise, they can help stroke patients develop more effective training and treatment plans.

This master thesis focuses on collecting users' body movement data through MediaPipe's Hand and Pose AI models, and comparing historical record data through landmark based

registration mathematical models to discover changes in stroke patients' physical conditions. It provides various deep learning based camera tracking test methods.

1. Hand peak-valley motion, which can assess hand dexterity by the change of hand motion area and speed.
2. By mirror transformation, continue to use the above-mentioned Hand peak-valley motion, which can compare the flexibility of the left and right hands.
3. By doing dumbbell exercises with the arms, the body landmark was recorded during the exercise and the number and speed of movements were calculated. These data can be used to analyze the arm muscle strength changes.

4.1.3 DTA characteristics

Under the definition of DT, DTA system could have these characteristics. They show on [Figure 4.2](#)

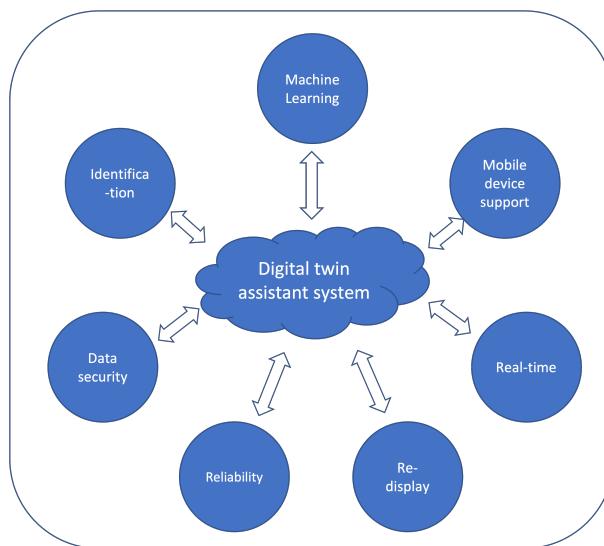


Figure 4.2: Digital twin assistant system characteristics

- **Identification.** Digital twins need to use each other's unique IDs to keep the information interaction between twins via DTA system. Each stroke patient's body mechanics are unique, such as the size of the palm of the hand, the length of the fingers and arm bones and characteristic behavior patterns.
- **Mobile device support.** The DTA system uses a monocular RGB camera from a mobile device, e.g., smart phone, laptop, computer, and iPad. Through the camera sensor, the patient's dynamic body movement data can be captured and stored in real time. Digital twin provides data to support the analysis of the patient's hand dexterity.
- **Machine Learning.** The DTA system and with the support of Artificial Intelligence and Deep Learning, the 3D motion information of the human body can be captured using only a monocular RGB camera. This can help patients to obtain 3D motion data of the body at a lower cost and can provide better support for decision making in the DTA system.
- **Real-time.** The DTA system can display 3D data of body movements of stroke patients in real time with low latency. The body movement data can also be trans-

mitted remotely to a rehabilitation physician or data storage center via 5G, 4G or other network transmission protocols.

- **Re-display.** The redisplay of the collected human 2.5D movement data by MediPipe can be used to compare the physical movement history of stroke patients and to know the effect of rehabilitation training.
- **Reliability.** The 3D human motion data recorded by the DTA system is based on real-time video streaming data, calculated by artificial intelligence and deep learning methods, which can authentically record the human motion data.
- **Data security.** Digital twin records body movement data and does not contain other personal and private data. This essential movement data can be stored and transferred using standard encrypted methods, such as SHA-265 and HTTPS, to ensure that only relevant rehabilitation physicians and researchers can view and use it to protect the privacy of patient users.

4.2 MediaPipe

MediaPipe [53] is a cross-platform platform that supports Android, iOS, C++, Python and webpage, and helps developers focus on implementing machine learning algorithms. The platform's input, output and internal computation module flow are optimized to improve the overall operational efficiency. The following Hands[16], Pose [38] and Holistic¹ are all based on the MediaPipe platform.

4.2.1 Hands

Real-time, low-latency gesture tracking with a single RGB camera via MediaPipe Hands project [16]. The input is a video stream, or a single image. The output is a video image at 30 frames per second, containing the coordinates and depth information of the 21 finger joint planes of multiple hands. The detail hand landmark and a real hand tracking are show in [Figure 4.3](#)

¹<https://ai.googleblog.com/2020/12/mediapipe-holistic-simultaneous-face.html>



(a) Hand landmarks



(b) Hand skeleton calculated by MediaPipe

Figure 4.3: Real time hand tracking(ref:MediaPipe Hands)

The MediaPipe[53] visualizer is a tool that helps users understand the topology and overall behavior of their pipelines, as shown in Figure 4.4. The tool consists of a visualization of a timeline view and graph view.

First, the MediaPipe system in Figure 4.4 receives the image stream through the camera and starts queuing each frame as it enters the Flow Limiter. When the landmark of the hand is detected for the first time, the palm of the hand is detected first. After the palm is detected, the landmark of the hand is then detected by image cropping. If the landmark of the hand is detected, the next step detects the landmark of the hand in the next image. If the landmark of the hand is not detected before, the detection of the palm of the hand will be performed again. Finally, the landmark and hand images are output.

MediaPipe Hand combines two machine learning(ML) models, a palm detector and a hand landmark model. The palm detector uses Single Shot MultiBox Detector(SSD)[54] for real time palm tracking. It helps crop the image before the hand landmark model is working.

4.2.2 Pose

MediaPipe Pose proposes a lightweight convolutional neural network architecture for human pose estimation. The network is able to track 33 key points of the human body Figure 4.5 in real time and can run smoothly on mobile phones. In Figure 4.6, the results based on this model are shown. In this thesis, it uses the output of MediaPipe Pose landmark to calculate the angle of arm movement. The speed of the dumbbell movement is analyzed to evaluate the flexibility of the arm in section 5.2.

4.3 Landmark based registration

4.3.1 Introduction

Image registration is the process of determining a geometrical transformation that aligns the points of an image with the corresponding points in a reference coordinate frame.

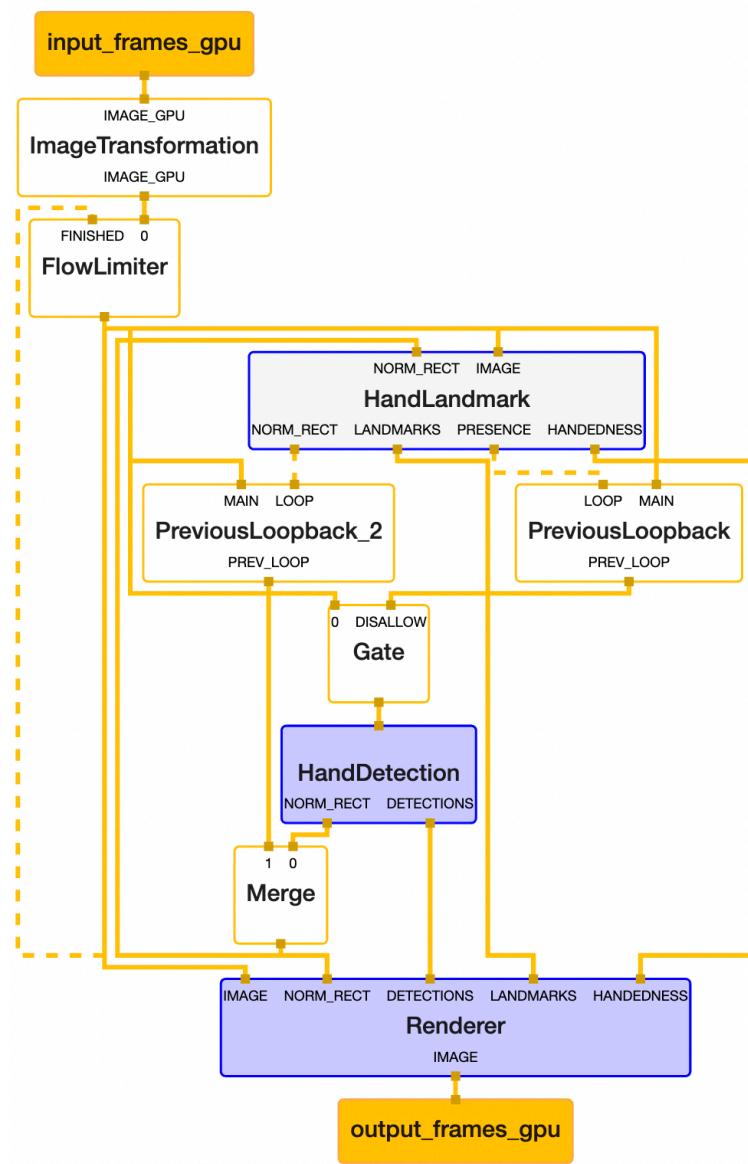


Figure 4.4: Machine learning inference pipelines of hand tracking([ref:MediaPipe Hands](#))

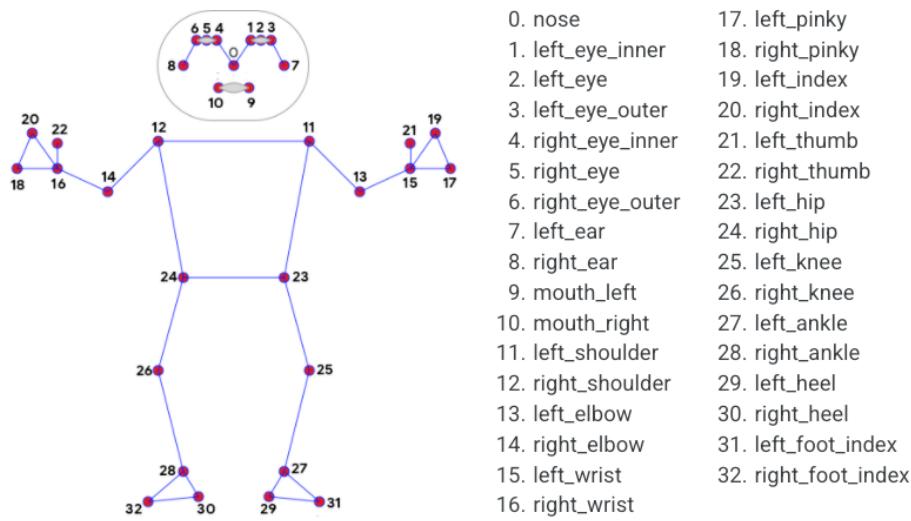


Figure 4.5: pose landmarks(ref:[MediaPipe Pose](#))

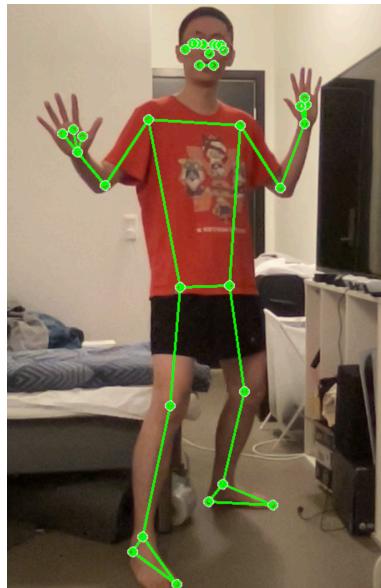


Figure 4.6: 33 pose landmarks in tester

This reference coordinate frame is often given as another image. However, it can be any arbitrary object coordinate system. In this topic, image registration will be applied in hand area tracking comparing.

Image registration adds value to hand landmark images by enabling

- Monitoring of changes in the individual hand
- comparison of one hand area with others hands in different frames
- comparison of groups' hand landmarks with others experiments

4.3.2 Elements of image registration

The task of image registration can be phrased as the following [55]: Given a so-called reference \mathcal{R} and a so-called template image \mathcal{T} , the basic idea is to find a reasonable geometrical transformation such that a transformed version of the template image becomes similar to the reference image.

Image registration consists of the following elements:

- **The geometrical transformation** required to transform the template to the reference image. In some cases a rigid transformation, i.e., a translation and rotation suffices, in other situations a nonlinear geometrical transformation is necessary.
- **The similarity** measure that describes the goodness of the registration.
- **The optimization** algorithm that controls (determines the parameters of) the geometrical transformations to maximize similarity.
- **The regularization** term securing that only reasonable transformations are obtained.

4.3.3 Landmark and label

A **landmark** is a point of correspondence for each object that matches between and within populations.

In the literature, landmarks have been known by various synonyms and are partitioned into various types. We will consider the three types defined by [56] in relation to biological shapes:

- **Scientific landmarks** are points assigned by an expert that correspond between objects in some scientifically meaningful way, for example, the corner of an eye or the meeting of two sutures on a skull. In biological applications, such landmarks are also known as **anatomical landmarks**.
- **Mathematical landmarks** are points located on an object according to some mathematical or geometrical property of the figure, for example, at a point of high curvature or at an extreme point. The use of mathematical landmarks is particularly useful in automated recognition and analysis. Geometrical transformations to maximize similarity.
- **Pseudo-landmarks** are constructed points on an object, located either around the outline or in between scientific or mathematical landmarks.

A **label** is a name or number associated with a landmark, and identifies which pairs of landmarks correspond when comparing two objects. Such landmarks are called labelled landmarks. In [Figure 4.3\(a\)](#), the hand has 21 joint point, each landmark has assigned a label name. The [Figure 4.3\(b\)](#) shows the hand joints in reality.

4.3.4 Image transformation and FRE

This part introduces the landmark based registration and the contents of the section are mainly adopted from [57]. If two image are similar, the optimal translation, rotation, and scale parameters are found by matching feature points, which is called landmark based registration. This transformation method is often used in computed tomography(CT) and magnetic resonance imaging (MRI) slice comparison. It is used to find physical changes in the patient's body over time, such as whether the size of the tumor has changed. But in this master thesis, the main purpose is compare hand landmarks in single and multiple experiments. Suppose there are two images with feature points $\mathbf{x}_i, \mathbf{y}_i \in \mathbb{R}^2$. The process of finding the transformation $\hat{\mathbf{w}} \in \mathbb{R}^m$ [Equation 4.1](#) such that the squared distances between \mathbf{x}_i and \mathbf{y}_i after geometric transformation minimized is called geometrical mapping.

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} \sum_{i=1}^N \|\hat{y}(x_i, \mathbf{w}) - y_i\|^2 \quad (4.1)$$

Translation. In exceptional cases, it is enough to consider only the translated movement between image captures. The translation formula for 2D space $\mathbf{t} = (t_1, t_2)^T$ in 2D is as follows [Equation 4.2](#)

$$y(x; \mathbf{t}) = x + \mathbf{t} \quad (4.2)$$

By differentiating the objective function of [Equation 4.1](#) get $\hat{\mathbf{t}}$ in [Equation 4.3](#)

$$\begin{aligned} \bar{x} &= \sum_{i=1}^N x_i / N \\ \bar{y} &= \sum_{i=1}^N y_i / N \\ \hat{\mathbf{t}} &= \bar{y} - \bar{x} \end{aligned} \quad (4.3)$$

Rigid transformation, including translation and rotation, can be used for comparison in the case where the participating user's hand is at a fixed distance from the camera. Translation parameters in 2D are $\mathbf{t} = (t_1, t_2)^T$. The rotation matrix R is an orthogonal matrix, which mean $R^T R = I$. If the angle in 2D space is θ , the rotation matrix is [Equation 4.5](#). In here the transformation shows in [Equation 4.4](#).

$$y(x; R, \mathbf{t}) = Rx + \mathbf{t} \quad (4.4)$$

$$R = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \quad (4.5)$$

Centered landmark is the average of all coordinates subtracted from each point coordinate, which is used to reduce the cost of align. By inserting [Equation 4.6](#) into the [Equation 4.1](#) the objective function becomes [Equation 4.7](#)

$$\begin{aligned} \tilde{x}_i &= x_i - \bar{x} \\ \tilde{y}_i &= y_i - \bar{y} \end{aligned} \quad (4.6)$$

$$L_{OBJ} = \sum_{i=1}^N \|R\tilde{x}_i - \tilde{y}_i + t - (\bar{y} - R\bar{x})\|^2 \quad (4.7)$$

Differentiation to t yields $t = \bar{y} - R\bar{x}$ and inserting this in [Equation 4.7](#), the result is [Equation 4.8](#)

$$\begin{aligned} L_{OBJ} &= \sum_{i=1}^N \|R\tilde{x}_i - \tilde{y}_i\|^2 \\ &= \sum_{i=1}^N (\tilde{x}_i^T R^T R \tilde{x}_i + \tilde{y}_i^T \tilde{y}_i - 2\tilde{x}_i^T R^T \tilde{y}_i) \end{aligned} \quad (4.8)$$

Through $R^T R = I$, it can be shown [58] that get minimize R by apply a SVD(singular value decomposition) of $H = \sum_{i=1}^N \tilde{x}_i \tilde{y}_i^T = UDV^T$, with $U^T U = V^T V = I$, then get

$$\begin{aligned} \hat{R} &= V \text{diag}(1, 1, \det(VU)) U^T \\ \hat{t} &= \bar{y} - \hat{R}\bar{x} \end{aligned} \quad (4.9)$$

The diagonal matrix inserted in between V and U ensures that we get a proper rotation.

Similarity transformation, including translation, rotation, and isotropic scaling, adds a scale parameter compared to the rigid transformation. The transformation equation is as follows [Equation 4.10](#) and $s > 0$. By similarity transformation, the area of the hand can be compared with the landmark of different users' hands at different distances. Because the area of the hand will shrink and enlarge with the distance from the camera lens, as in [Figure 4.16](#).

$$y(x; s, R, t) = sRx + t \quad (4.10)$$

By calculating similarity to [Equation 4.10](#), the result [Equation 4.11](#) is obtained .

$$\begin{aligned} \hat{R} &= V \text{diag}(1, 1, \det(VU)) U^T \\ \hat{s} &= \frac{\sum_{i=1}^N \tilde{x}_i^T \hat{R}^T \tilde{y}_i}{\sum_{i=1}^N \tilde{x}_i^T \tilde{x}_i} \\ \hat{t} &= \bar{y} - \hat{s}\hat{R}\bar{x} \end{aligned} \quad (4.11)$$

Fiducial registration error(FRE), is used to compare the difference between the two sets of points and is used to judge the good or bad result of the geometric transformation. When the FRE is smaller, it means that a better transformation result is achieved.

$$\sigma_{FRE}^2 = \frac{1}{2N} \sum_{i=1}^N \|(x_i; \hat{w}) - y_i\|^2 \quad (4.12)$$

and \hat{w} are the optimal parameters given by [Equation 4.1](#).

4.4 Hand area tracking procedures

4.4.1 Hand area calculation



Figure 4.7: The whole hand area are calculated by sum of 4 triangles

Heron's formula ², formula that is named after Hiero of Alexandria, the formula requires only three side lengths to calculate the area of a triangle. This differs from other triangle area formulas in that it is not necessary to first calculate the angles or any other distances in the triangle. In Heron's formula [Equation 4.13](#), (a,b,c) are three edges of triangle. In each frame hand landmark area are calculated by sum of 4 triangles. Five fingertips and wrist point show in [Figure 4.7](#).

$$s = \frac{a + b + c}{2}$$

$$Area = \sqrt{s(s - a)(s - b)(s - c)}$$
(4.13)

4.4.2 Data acquisition

The real-time video is captured by a monocular RGB camera and the landmark data of timeline and video frames are calculated by MediaPipe Hand model. The example of landmark in image shows on [Figure 4.3:b](#). These progress shows on graph [Figure 4.4](#).

²sometimes called Hero's formula, [wiki:https://en.wikipedia.org/wiki/Heron%27s_formula](https://en.wikipedia.org/wiki/Heron%27s_formula)

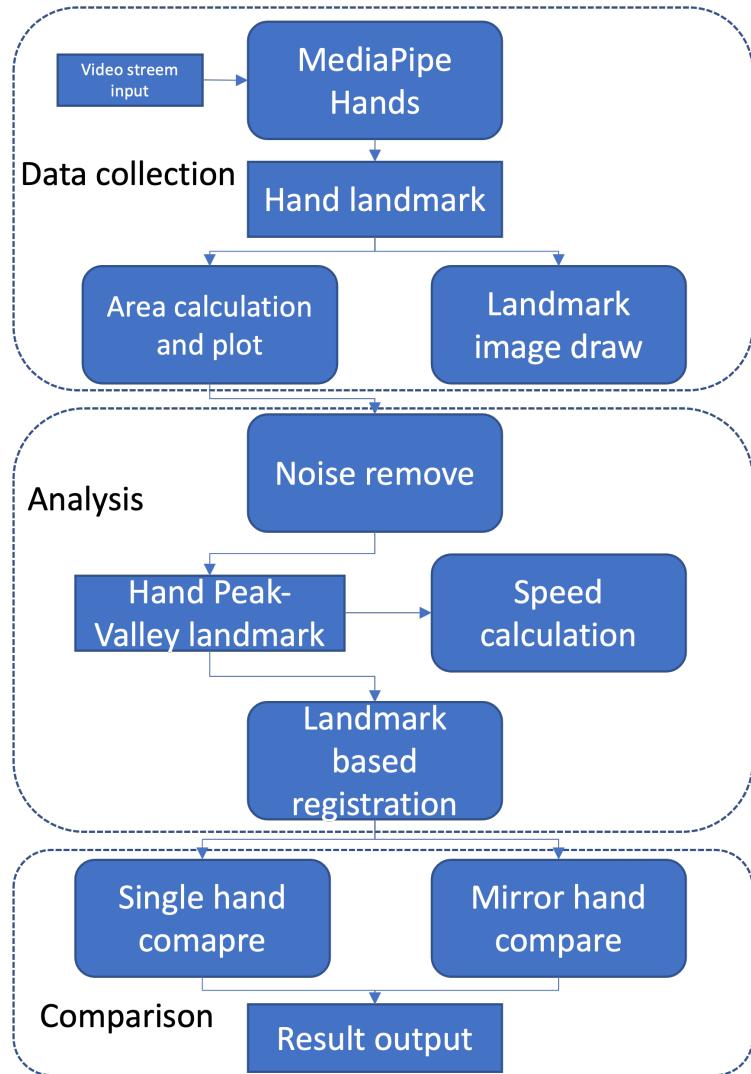


Figure 4.8: The whole process of hand tracking.

In Figure 4.8, after obtaining the landmark of the hand through MediaPipe, the data collection part in A will automatically save and calculate the area of the hand and display the landmark image result in real time, e.g. Figure 4.10. Analysis and comparison's part will do more calculations later.

4.4.3 Description of data set

In the experimental real-time video stream, each frame is computed by MeidaPipe Hand to obtain the landmark result, the result of each frame is as follows Figure 4.9. The calculated data for each frame of the video stream is saved along with the current timestamp, while the video data is not saved, which significantly reduces the amount of data that needs to be saved.

The storage method is **Hickle**. It is a convenient way of dumping python variables to HDF5 files that can be read in most programming languages, not just Python. Hickle is fast, and allows for transparent compression of the data, and only save landmark data also protects the user's privacy. To better identify the different experiments in the saved file names and to protect user privacy, the system uses the current timestamp to generate an **MD5** as the unique ID for this experiment. Different parts of the data in detail show

below.

MULTI_HAND_LANDMARKS: Collection of detected/tracked hands, where each hand is represented as a list of 21 hand landmarks and each landmark is composed of x, y and z. x and y are normalized to [0.0, 1.0] by the image width and height respectively. z represents the landmark depth with the depth at the wrist being the origin, and the smaller the value, the closer the landmark is to the camera. The magnitude of z uses roughly the same scale as x.

MULTI_HANDEDNESS: Collection of handedness of the detected/tracked hands (i.e., is it a left or right hand). Each hand is composed of a label and score. label is a string of values either "Left" or "Right". Score is the estimated probability of the predicted handedness and is always greater than or equal to 0.5 (and the opposite handedness has an estimated probability of 1 - score).

Note that handedness is determined assuming the input image is mirrored, i.e., taken with a front-facing/selfie camera with images flipped horizontally. If it is not the case, please swap the handedness output in the application.

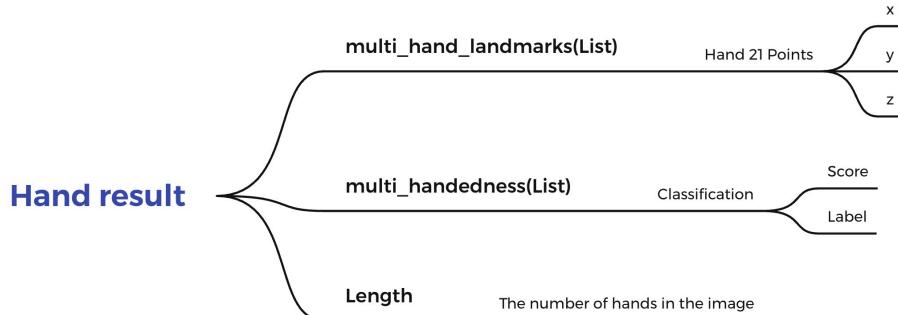


Figure 4.9: The hand landmark description in one image

4.4.4 Data Re-display

By reading the stored files and replotting the images of the hand joints according to the timeline, the necessary data collected from the previous experiment can be represented. In [Figure 4.10](#), the top red point with the largest hand area is called Peak, and the red point with the smallest hand area is called Valley. Additional experimental results are obtained later by analyzing the hand peak and valley.

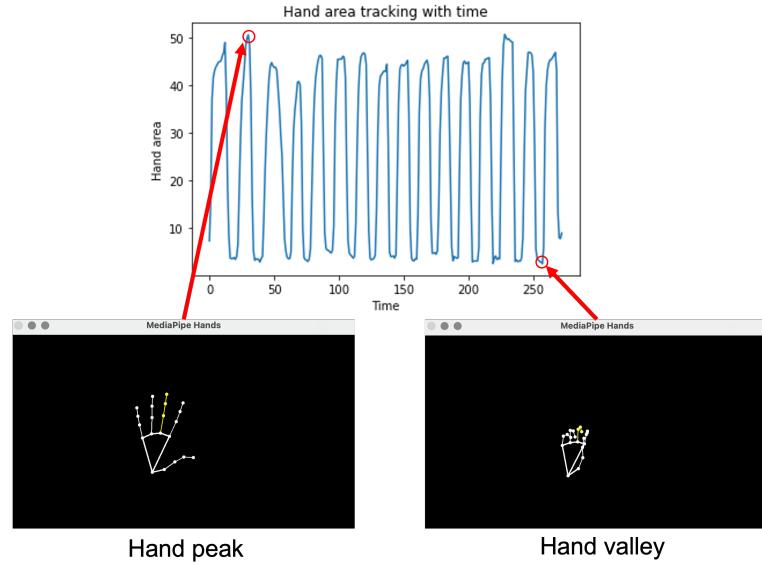


Figure 4.10: The hand landmark re-display by video with timeline

4.4.5 Head area peak/valley noise remove

In patient experiments recorded in real time, patients may shake their hands or make other gestures due to different situations, resulting in irregular changes in the hand area, which may generate noise like in [Figure 4.11](#). For example, at the beginning of the experiment, the patient's hand may shake when it starts to move, or the patient may receive other pathological factors that cause hand trembling during the experiment.

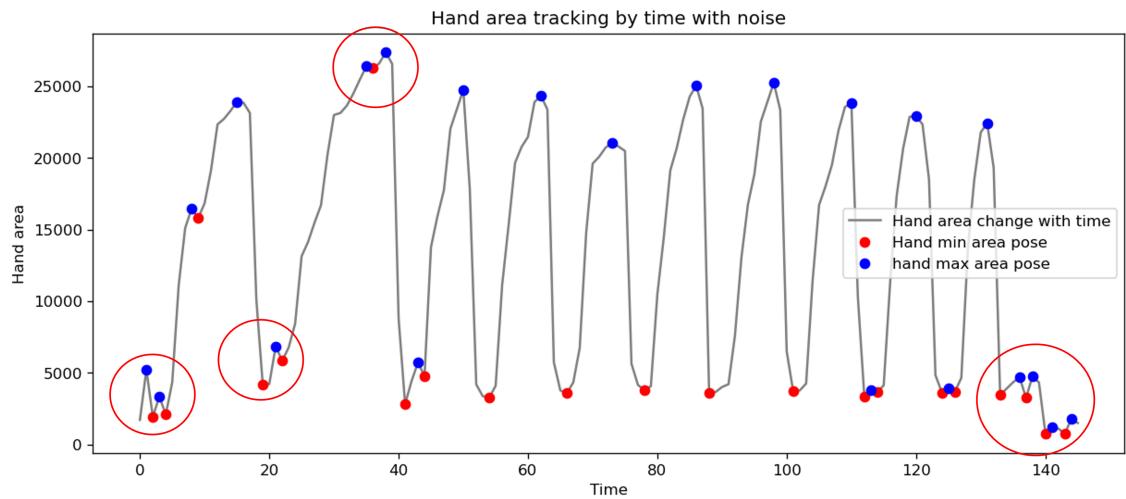


Figure 4.11: Hand area with noise, the red circles are noise in the experiment

In order to get Hand area **Local extremes** in [Figure 4.11](#), there are four steps. Step1: Calculating the first difference by [Equation 4.14](#), the result [Figure 4.12](#) shows the trend of increasing and decreasing area is indicated by positive and negative numbers, and the magnitude of the value indicates the speed of change..

$$out[i] = a[i + 1] - a[i] \quad (4.14)$$

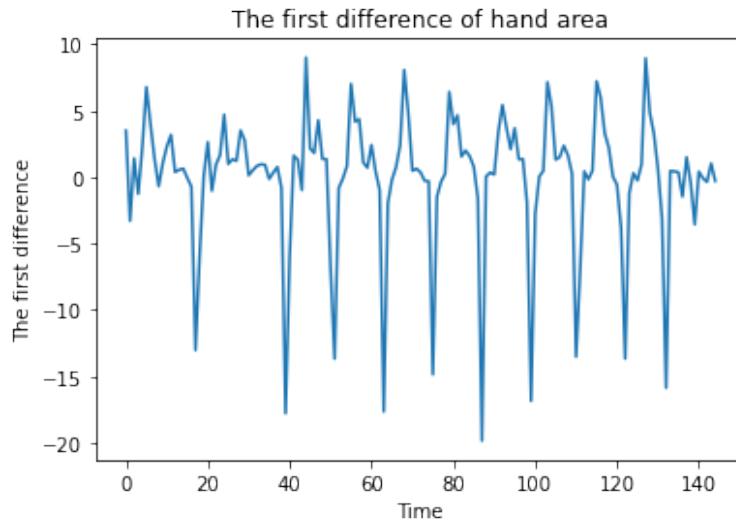


Figure 4.12: The first difference of hand area

Step2: The local extreme is focused in here, so the Sign function [Equation 4.15](#) will remove trends in [Figure 4.13](#).

$$\operatorname{sgn} x := \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases} \quad (4.15)$$

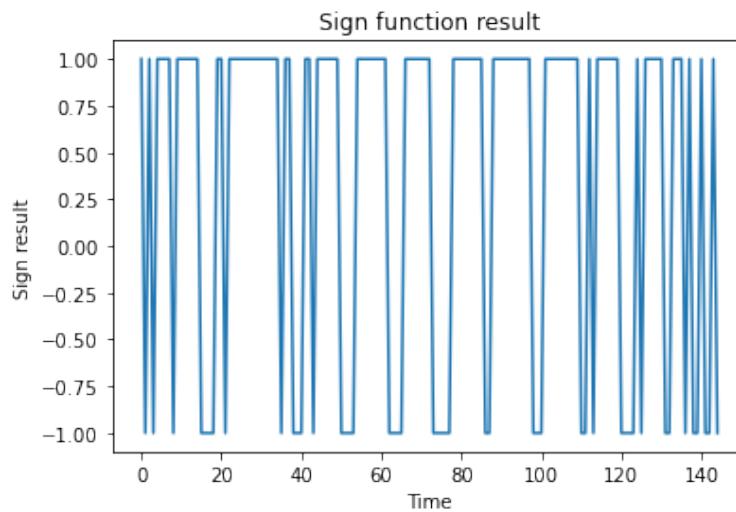


Figure 4.13: Sign function result after the first difference of hand area

Step3: after the Sign function, redo the [Equation 4.14](#) will get local extremes like [Figure 4.14](#). The peak and valley can be identified by [Equation 4.16](#), the final result shows on [Figure 4.11](#).

$$Index = \begin{cases} Peak & \text{if } x < 0 \\ Valley & \text{if } x < 0 \end{cases} \quad (4.16)$$

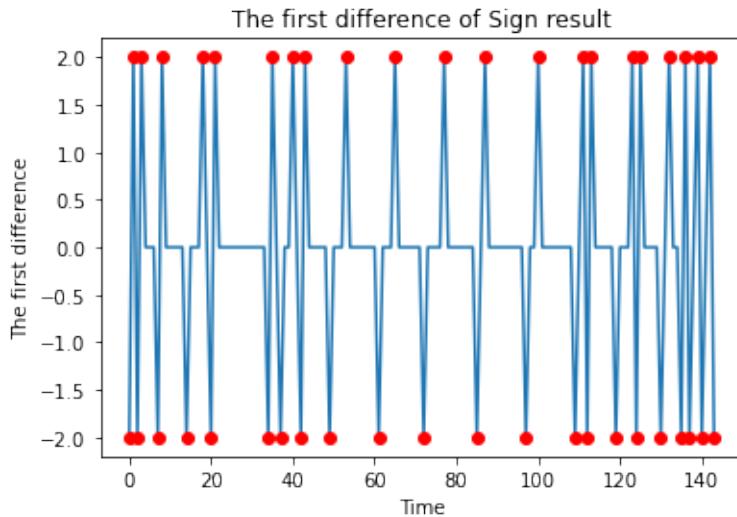


Figure 4.14: Peak and Valley with noise after second difference calculation

Due to the fluctuation of the gesture area, the wrong Peak and Valley are shown in [Figure 4.11](#). By setting α and β to control, the equation is as follows [Equation 4.17](#). Now the true hand area Peak and Valley are filtered in [Figure 4.15](#). In experiment test, the $\alpha = 0.4, \beta = 0.5$ in here.

$$\begin{aligned} Peak[i] &> \alpha \times Peak_{max} \\ Valley[i] &< \beta \times Valley \end{aligned} \quad (4.17)$$

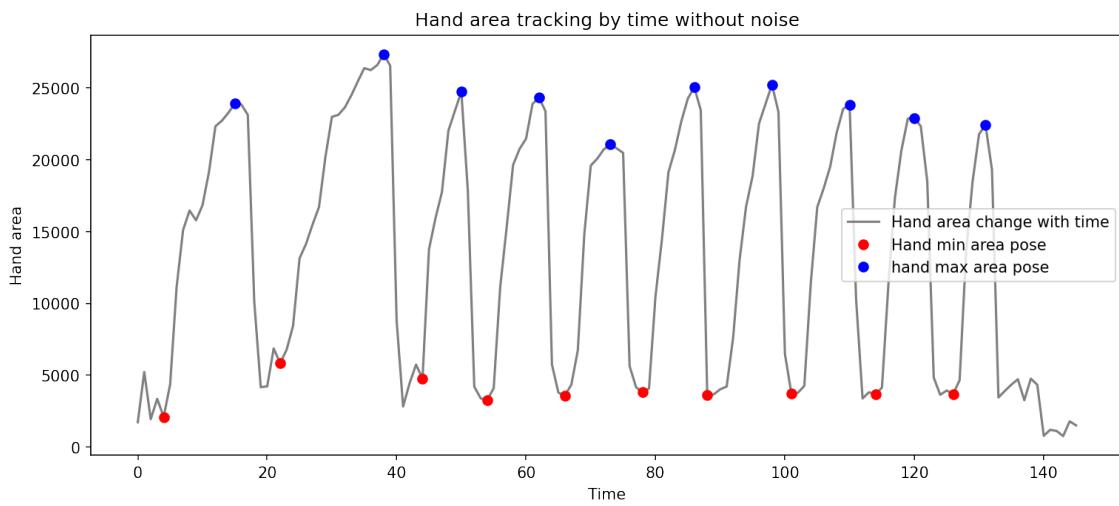


Figure 4.15: Hand area after removing noise, only the real peak and valley are kept

4.4.6 Hand peak/valley speed

After filtering by [Equation 4.17](#), the true peak and valley can be used to calculate the velocity of hand-area motion. **Definition of a complete motion:** The motion between two adjacent Peaks or two adjacent Valleys can be called one training cycle. To mitigate the miscalculation of Peak and Valley, the average speed (count/minute) of both is calculated by the [Equation 4.18](#) to represent the frequency of the overall training.

$$\begin{aligned}
 V_{Peak} &= \frac{\text{Length}(Peak)}{S_{peak[\text{end}]} - S_{peak[0]}} \\
 V_{valley} &= \frac{\text{Length}(Valley)}{S_{valley[\text{end}]} - S_{valley[0]}} \\
 \bar{V} &= \frac{V_{peak}}{V_{valley}}
 \end{aligned} \tag{4.18}$$

The opening and closing of the fingers is a basic form of hand movement, the amplitude and frequency of hand movement which can be used to assess and track hand health in real time and in the long term.

4.4.7 One group landmark based registration

When recording the user's hand landmark in real time, moving the hand closer and further away from the camera causes the measured hand area to expand or shrink in the equal proportion, but the area of the hand remains virtually unchanged. In [Figure 4.16](#), the object at the same height is shown on the left, and when it is close to the camera, the image becomes larger, and when it is far away, the image becomes smaller, eg: The height of the object is L , when $d_3 > d_1 > d_2$, the projection result is $L_2 > L_1 > L_3$. The photo on the right is the result of the same hand in the real camera, which validates the above statement.

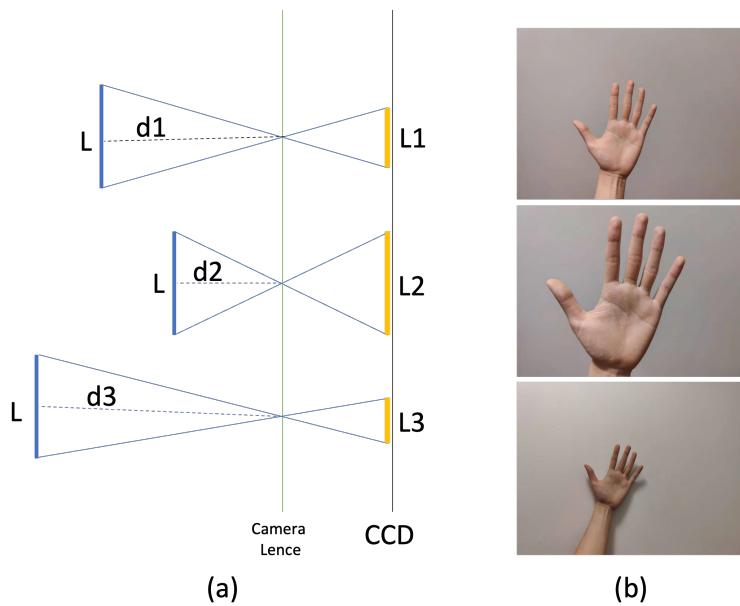


Figure 4.16: Hand area scaled by distance

In the experiment, the testers need to repeat the Peak => Valley => Peak loop exercise, and in different experiment groups, the testers need to take rubber bands for simulating hand stroke or injuring caused disability impact. The MediaPipe could calculate the hand joint position in real time. The whole test process shows on [Figure 4.17](#).

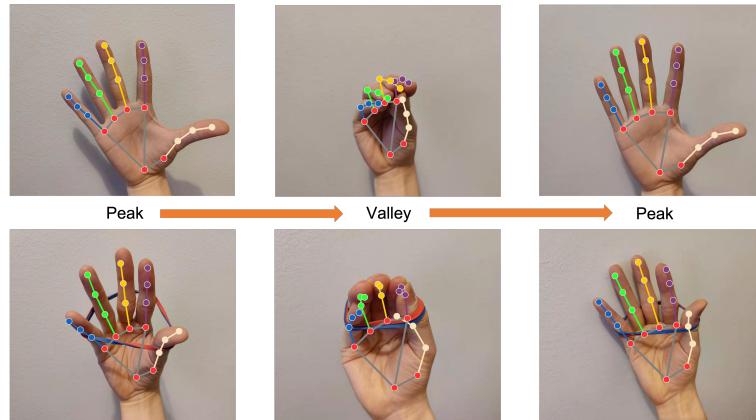


Figure 4.17: Hand peak => valley => peak hand area test progress by bare hand and restrained with rubber band

At the same time, when users do hand area testing, they also need to take into account the impact of hand rotation and translation on hand area. Because MediaPipe can only capture 2.5D data, the joint depth data is the depth of the joint relative to the wrist, which is negative near the camera and positive far from the camera. These depth data are not accurate spatial data, so the user's palm is required to face the camera for testing to ensure the accuracy of the experiment.

When the testers do the peak and valley exercise, the palm joint will not change too much in reality, because the bone in the palm has no extra joint. For this reason, the landmark based registration will base on the palm five joint. The palm joints show in [Figure 4.18](#). Through section 4.3: landmark based registration, the algorithm could auto-register the peak and valley landmark to the first Peak and Valley landmark. This method could remove the hand translation, rotation, and scale influence.

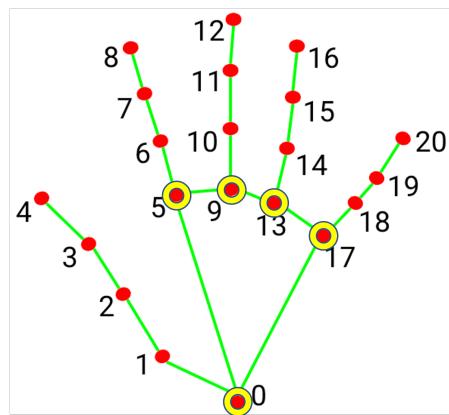


Figure 4.18: The yellow marker is hand palm registration 5 key joints

From [Figure 4.10](#), plot one group all Peak and Valley Landmark together in [Figure 4.19](#). The left part is an unregistered landmark, they are not aligned well. The right part is the landmark registered by five palm key joints, they are aligned well. In one group landmark based registration process, the later landmarks will register with the first Peak and Valley landmark, which will have reduced translation, rotation, and scale influence on the hand area.

The FRE calculated by [Equation 4.12](#) explains the result of landmark based registration. The smaller FRE means the register result is better, the larger FRE means the register is not good.

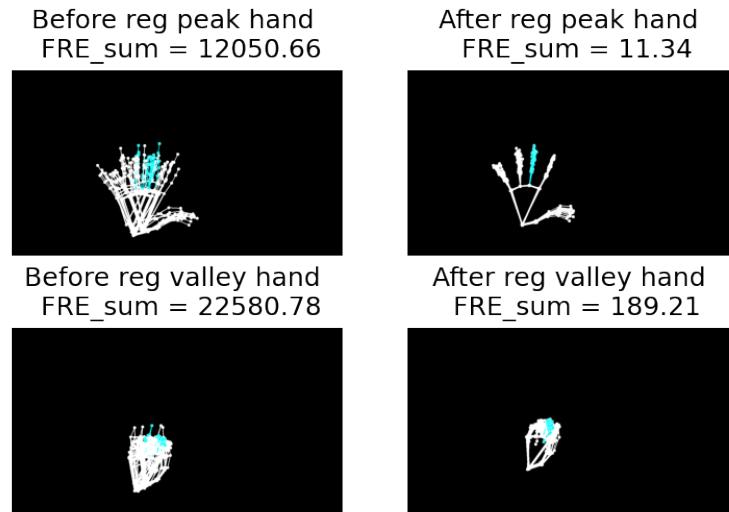


Figure 4.19: One group landmark registration

4.4.8 Multi-group landmark based registration

Comparing different groups, landmarks need register the second group Peak and Valley landmarks to the previous group first Peak and Valley landmark. The second group experiment is a hand with two rubber bands. The registration result shows on [Figure 4.19](#). The second column is the result after landmark based registration, which is in same position with the result in [Figure 4.19](#). The FRE of peak hand landmark decreases a lot from 75256.9 to 205.71, which means the registration is success. The FRE of Valley hand landmark has the same conclusion.

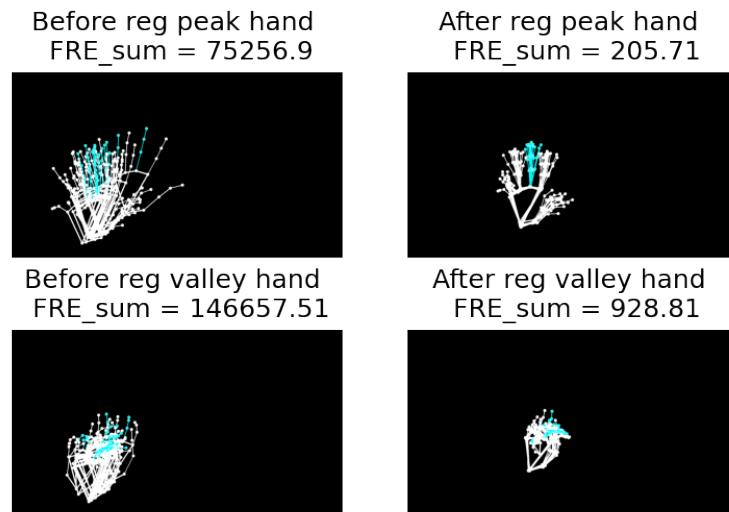


Figure 4.20: Two group landmark compared registration

5 Experiment

5.1 Hand area tracking experiment

In this experiment, MediaPipe hand model is used to capture the hand landmark in real time video, the DTA system will calculate the hand area in real time and save the landmark data. Then, the peak and valley values of the hand area are obtained by data analysis. By removing noise, then find these local extremes, the frequency and number of hand movements are obtained. Subsequently, the hand landmark at different time points was matched and transformed by landmark based registration to provide methodological support for comparing the average change of hand area between one experiment and multiple experiments.

5.1.1 Single hand tracking with rubber band constraints

Before starting to design the experiment, it is necessary to explain why different rubber bands are used to limit the hand motion, and then observe the changes in hand area and velocity. The ultimate goal of the experiment is to detect changes in hand dexterity in elderly stroke patients or patients with hand injuries in time to give physicians or rehabilitation physiotherapists a tool to quantitatively track and assess changes in hand dexterity. In order to simulate the dexterity of the user's hand in different situations, different numbers of rubber bands were used to achieve this result. Multiple experiments under the same conditions and repetitive experiments at different distances are performed to enable statistical experiments to increase the reliability of the results.

Four different repeated experiments were designed with the same participant's left hand. The detail shows in below:

- **Bare hand:** left bare hand doing Peak-Valley exercise 20 times.
- **Hand with 1 rubber band:** 20 times of Peak-Valley exercise with the left hand.
- **Hand with 2 rubber bands:** The left hand did 20 times Peak-Valley exercise under the restriction of two rubber bands.
- **Hand with 3 rubber bands:** The left hand did 20 times Peak-Valley exercise under the restriction of three rubber bands.

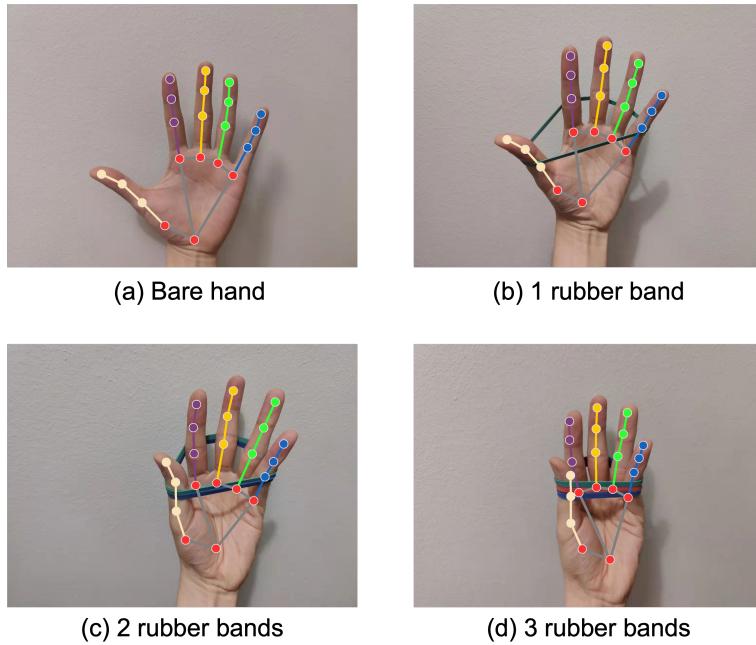


Figure 5.1: Four Control test groups compare hand area and speed with none, one, two and three rubber bands

There are three conditions in the later compared analysis. **Condition:Mix**, in each experiment will repeat 20 times Peak-Valley exercises; **Condition:Stable**, the first 10 times hand Peak-Valley exercise, the hand is fixed at same place; **Condition:Fluctuate**, the later 10 times, the hand will do the exercise at difference place so that the hand's area will be influenced by the distance of camera. Each exercise is required for more than 30 seconds, when the hand wearing rubber bands need to be placed at the two joints of each finger, the position of reference [Figure 5.1](#) in 3: THUMB_IP, 6: INDEX_FINGER_PIP, 10: MIDDLE_FINGER_PIP, 14: RING_FINGER_PIP, 19: PINKY_PIP, e.g., B. For the analysis of experimental results, see C.

Table 5.1: Single hand tracking conditions and experiments design

Condition/Experiments NO	Bare hand	1 rubber band	2 rubber bands	3 rubber bands
Mix	1-20	21-40	41-60	61-80
Stable	1-10	21-30	41-50	61-70
Fluctuate	11-20	31-40	51-60	71-80

In each repeat experiment, the hand landmark is found through the noised removed process to all Hand Peak and Valley landmarks and timestamps, and then the speed of hand movements, and the hand area at each Hand Peak and Valley are calculated. By comparing the changes in hand area and velocity for different numbers of rubber bands, the following questions are addressed. The result and analysis show on section [6.1](#).

- First: Can the MediaPipe Hand model find significant changes in hand Peak and Valley area under different finger limitation by wearing different rubber bands after landmark based registration, and does the result can be verified by statistical experiments?

- Second: Does different rubber bands have a negative impact on the hand movement speed, i.e., the more rubber bands there are, the lower the hand movement speed will be?

5.1.2 Mirror hand compare

The above method is mainly used to compare one hand in one group of experiments and one hand in different groups of experiments, however, most people have two hands. Although humans have two similar hands, it is practically impossible to compare them directly in an experiment. The reason they cannot be compared is that the two hands are mirror symmetric. To compare the two hands experimentally by image, one would need to perform the experiment on both hands separately. Then use a mirroring technique, such as in [Figure 5.2](#), to perform the transformation before comparing.

By comparing the speed of left hand and right hand and the Hand Peak-Valley landmark over time, the difference in dexterity between the two hands can be known. This is important to assess the overall physical coordination of stroke patients, as the left and right hands need to work together to perform various tasks in daily life more efficiently, such as: putting on clothes, carrying dishes, lifting heavy objects, clapping, hugs etc. Assessing the dexterity of both hands experimentally can also help detect stroke earlier if a large difference is found in the dexterity of one hand compared to the other.

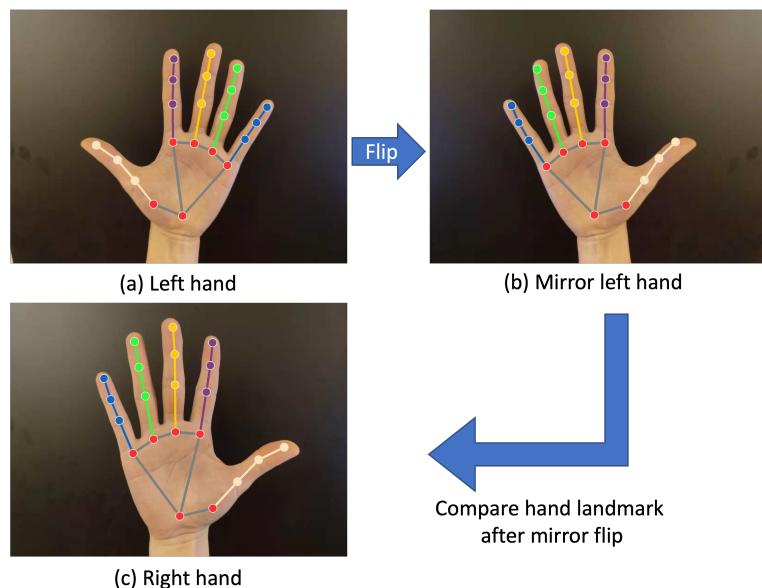


Figure 5.2: Left hand flip as mirror hand to compare with right hand

When the landmarks of the left and right hands are compared, special comparison processing is required. The reasons and steps of the mirroring experiment are described in detail in the next.

- **Step1:**[Figure 5.3:\(a\)](#) is the landmark of the left hand, and the Mirror transform in 1 will mirror the left hand into the right hand. The specific method, is the left hand original picture [Figure 5.2:\(a\)](#) after doing the mirror transformation, and then use MediaPipe Hand model to generate the landmark of the hand.
- **Step2:** By registering the right hand landmark [Figure 5.3:\(b\)](#) to the left hand landmark [Figure 5.3:\(a\)](#), we get the right hand transformation result.

- **Step3:** Putting the landmark of the right hand transformation result and the left and right landmark on the same image [Figure 5.3:\(d\)](#), it is obvious that the registration process is a failure. The reason is that the registration points are based on the five points of the palm as the standard, while the left and right hands cannot be matched directly due to the mirror symmetry problem, resulting in the failure of the registration results.
- **Step4:** By registering the right hand landmark [Figure 5.3: \(b\)](#) to the left hand mirror landmark [Figure 5.3: \(c\)](#), the new transformation result of the right hand is obtained.
- **Step5:** Once again, put the transformed landmark of the right hand and the mirror [Figure 5.3:\(c\)](#) of the left hand on the same image [Figure 5.3:\(e\)](#), and you can see that the left and right hands almost perfectly match each other. This shows that the registration is successful, and this method of postregistration mirroring can be used to compare the landmarks of both left and right hands.

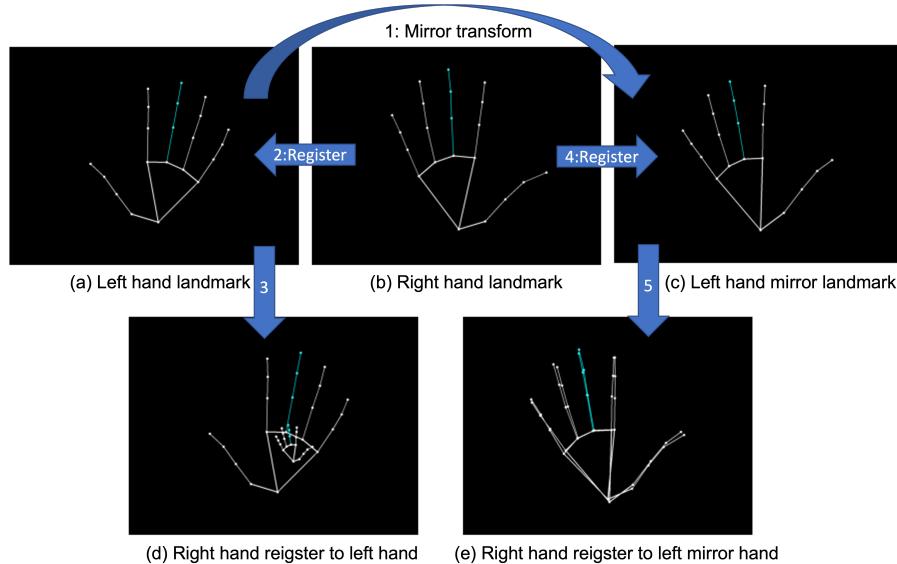


Figure 5.3: This figure show why left and right hand landmark can not register directly. After the left hand mirrored to right, the landmark based registration work well

To verify that the left-right hand dexterity comparison is valid in the MediaPipe Hand model, the following experiments were designed in this section to verify that the left-right hand model can be used to compare the difference between left and right hand dexterity.

Three experimental groups were designed with the same participant.

- **Left bare hand group:** left bare hand doing Peak-Valley exercise 20 times.
- **Left bare hand with 2 rubber bands group:** The left hand did 20 times Peak-Valley exercise under the restriction of two rubber bands.
- **Right bare hand group:** 20 times of Peak-Valley exercise with the right hand.
- **Right bare hand with 2 rubber bands group:** The right hand did 20 times Peak-Valley exercise under the restriction of two rubber bands, so that it could compare with the left hand.

Each exercise is required for more than 30 seconds, when the hand wearing rubber bands

need to be placed at the two joints of each finger, the position of reference [Figure 4.7](#) in 3: THUMB_IP, 6: INDEX_FINGER_PIP, 10: MIDDLE_FINGER_PIP, 14: RING_FINGER_PIP, 19: PINKY_PIP, e.g., B. For the analysis of experimental results, see C.



Figure 5.4: Put two rubber bands on the right hand at second joint of fingers

Two questions need to be explored here based on the experimental results

- **First:** Is there any difference in the comparison of hand area between the left hand and the right hand of the same tester's bare hand or hand with two rubber bands?
- **Second:** compared with the left and right hand speed test results, is there a significant difference in the comparison?

The null hypothesis given for both cases is that there is no significant difference between the area and speed of the left hand and the right hand. The result and analysis show on section [6.1.3](#).

5.2 Dumb-bell exercise arm angle tracking Experiment

The experiment was conducted with MediaPipe Pose [38] to track 2.5D data of major joints in the human body. Since it is for detecting the angle of arm movement, only 2D data is used in this experiment to calculate the angle of arm movement using the three points right_shoulder(12), right_elbow(14), right_wrist(16) in [Figure 4.5](#). The purpose of the experiment is to explore whether the force change of the user's arm can be detected by the speed through the MediaPipe Pose model. This will help stroke patients to better focus on the changes in arm strength. All data are recorded in real time and eventually stored as landmarks . The stored data does not include the user's picture, which reduces the storage size and protects the user's privacy.

In this experiment of dumbbell movement, three experimental groups will be used: no load, light load and heavy load in [Figure 5.5](#). Because the angle does not change depending on the distance away from the camera, each experimental group is repeated 10 times in this experiment and the data is collected for more than 30 seconds each time. In order to simulate the variation of the user's arm muscles, different weights of load were used to simulate the variation of arm force. The null hypothesis was that the average speed of the dumbbell was the same for the three different loads. The results of the analysis of the experiment are presented in section [6.2](#).

- **No load:** No load was used as the control group for the experiment.
- **Light load (683g):** A cup with water was used as the load.
- **Heavy load (1060 g):** A large plastic bottle with water was used as the weight.

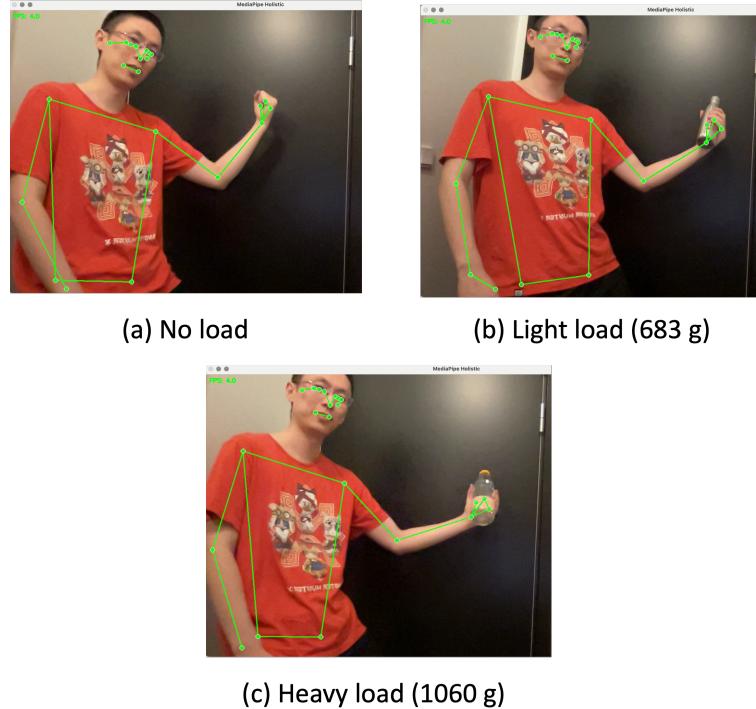


Figure 5.5: Dumb-bell three conditions: no load, light load (683 g) and heavy load (1060 g).

An image of the change in arm angle with time during a dumbbell movement is shown in [Figure 5.7](#). The blue dots and red dots are the positions of arm extension and contraction, respectively. When the noise was not removed in [Figure 5.6](#), the final local extremes of the plot clearly did not meet the experimental requirements, resulting from the disturbance generated at the end of the experiment when the arm ended its motion. After noised removal, most the correct local extremes are preserved in [Figure 5.7](#) to ensure the accuracy of the speed calculation. The average speed in each experiment was calculated by using the same noise method section [4.4.5](#) and equation [Equation 4.18](#) as in the hand area analysis. The mean speed results in different conditions are shown in [Table 6.27](#).

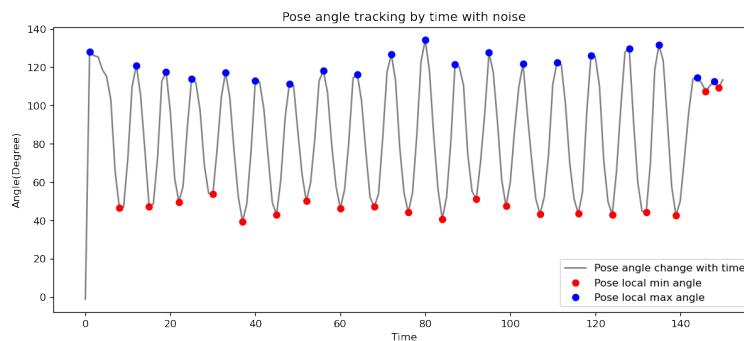


Figure 5.6: Dumb-bell angle tracking plot with noise in one experiment.

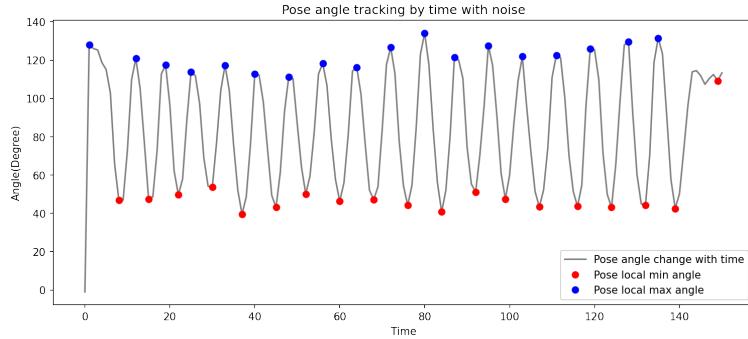


Figure 5.7: Dumb-bell angle tracking plot after noise removed in one experiment.

5.3 MeidaPipe model performance and stability

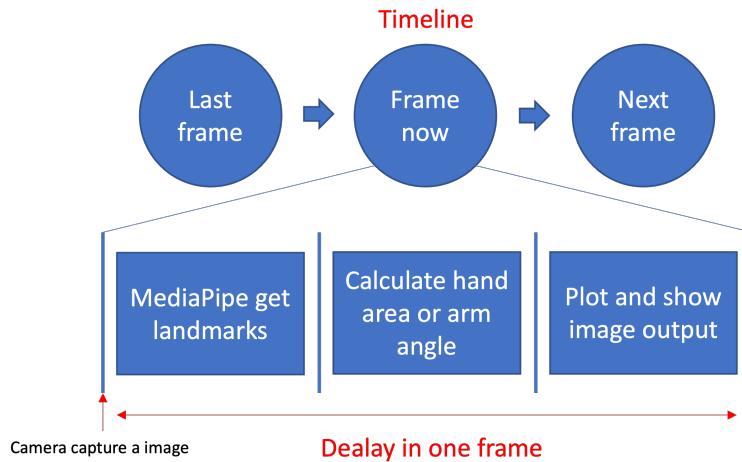


Figure 5.8: Delay test methods used in both experiments.

In this experiment, by using MediaPipe hand and pose time delay analysis, after capturing a frame with the camera, the complete processing time of the current image is calculated after the end of processing [Figure 5.8](#). By recording the delay for each frame of the image, the results of the two experiments recorded in [Figure 5.9](#) are shown. The average delay time and FPS are calculated in [Table 5.2](#). The data shows that the dumbbell motion has a higher delay time and a lower FPS. However, both methods have in stronger stability and the delay does not fluctuate with the change of environment. Even if the user wears the exoskeleton device [Figure 5.10](#), the MediaPipe Hand and Pose models can still find the hand and body key points correctly.

Table 5.2: Experiment delay data and FPS

Experiment	Mean delay(ms)	FPS
Hand area tracking	83	10
Dumb-bell angle tracking	178	4

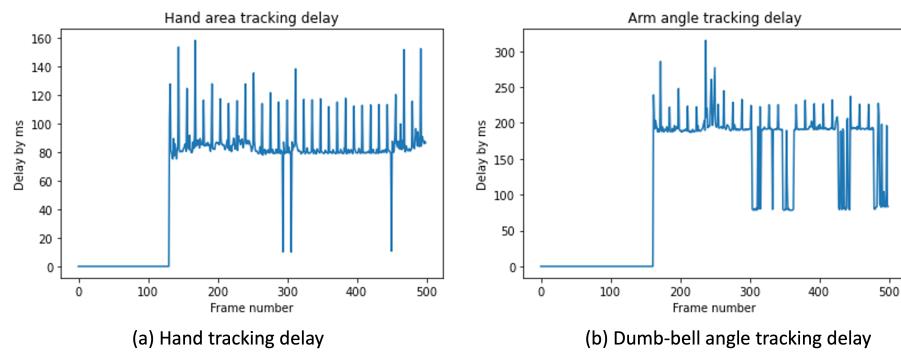


Figure 5.9: Delay plot for both experiments, the test equipment is Macbook air with M1 chip.

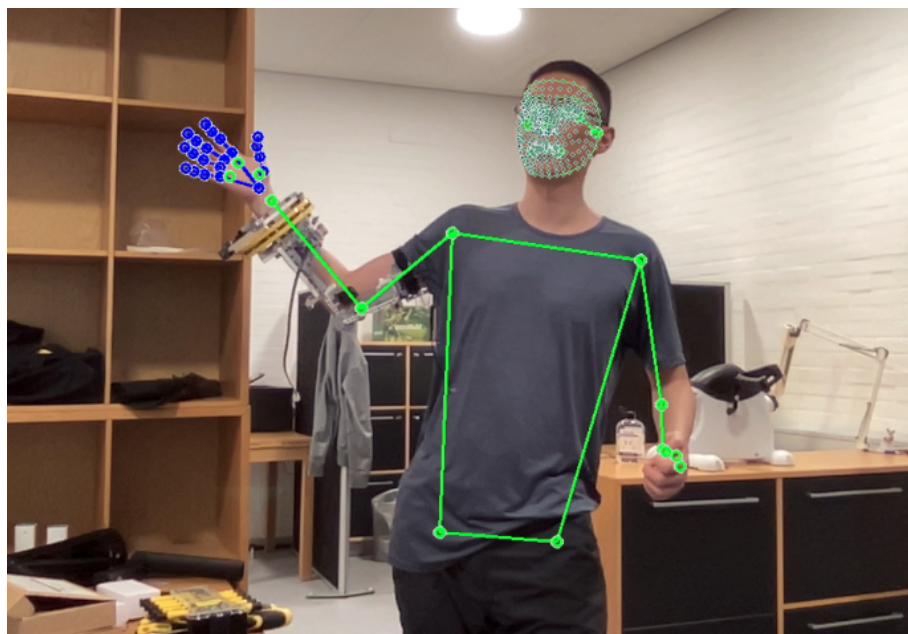


Figure 5.10: MediaPipe model still could recognize pose landmark when participant wearing LEGO exoskeleton equipment.

6 Numerical Result

6.1 Hand area tracking experiment analysis

The hand tracking system based on landmark registration section 4.3 is design for long-term hand dexterity and comparison. The next two experiments explore the hand tracking working on different situations and people. The assumption is that the hand tracking system could detect hand dexterity changed by rubber band and work with different people.

6.1.1 Different rubber band data analysis

In the first experiment, it will test one user putting different number rubber bands around the finger or bare hand without any rubber band for testing flexibility of the hand under different fingers' pressures. Because the hand tracking system will project the hand landmark of different groups to the first peak and valley landmarks of the first group, to eliminate the effect of projecting different groups to each other. The experiment will also compare whether the different order of projection groups will affect the final results of the experiment. The detail design of the four groups shows on [Figure 5.1](#).

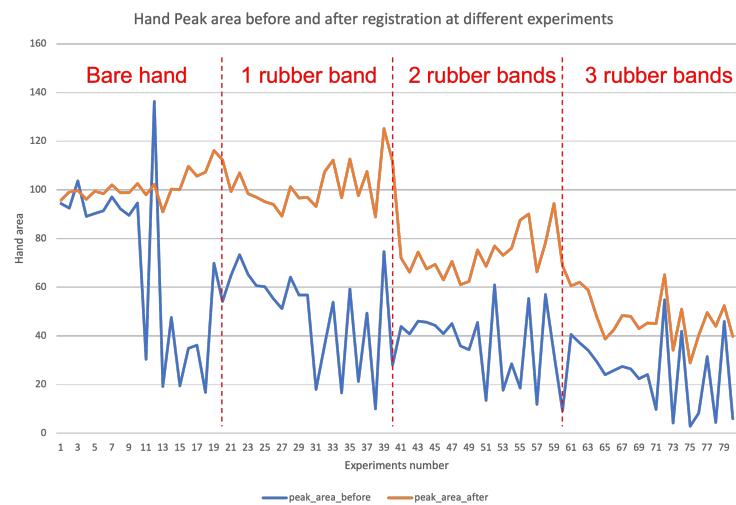


Figure 6.1: Hand Peak area compare by bare hand and hand with 1 to 3 rubber bands in repeat 80 times experiments

The experiments of Peak-Valley area tracking were repeated 20 times for each of the four different cases, and the total number of experiments was 80. The area of Hand Peak in each experiment is shown in [Figure 6.1](#). The horizontal coordinates of the axes are the numbers of the different experiments, each belonging to experiment group reference [Table 6.1](#); the vertical coordinates are the average relative area of the hand's landmark in each experiment. The blue line represents the average area of the hand in each experiment from the raw recorded landmark data; the orange line represents the average area of the hand after landmark based registration.

The assumption for [Figure 6.1](#) Hand peak area is that the average area of the peak hand decreases as the number of rubber bands increases. The assumption is that the average area of Hand Peak is the same for the same person with bare hand and 1-3 rubber bands. This prerequisite is satisfied by using two-sample t-test to test whether the hand peak mean values are equal between different experimental groups.

6.1.1.1 Mix condition compare

In [Figure 6.1](#), the mean area of the original hand landmark fluctuated widely from one experiment to another, so this conclusion could not be obtained directly by visual observation between different experimental groups.

Table 6.1: Experiment mix conditions and Hand Peak area mean

Experiment NO	Condition(mix)	Peak Area original mean	Peak Area mean after registration
1-20	Bare hand	70.00	101.71
21-40	1 rubber band	48.78	101.42
41-60	2 rubber bands	36.39	73.12
61-80	3 rubber bands	25.09	47.27

Table 6.2: P-value of different experiments original Peak mean area compare by Two-Sample t-Test at mix condition

Two-Sample t-Test	Bare hand	1 rubber band	2 rubber bands	3 rubber bands
Bare hand	-	0.02	<0.001	<0.001
1 rubber band	-	-	0.03	<0.001
2 rubber bands	-	-	-	0.02

Table 6.3: P-value of different experiments after registration Peak mean area compare by Two-Sample t-Test at mix condition

Two-Sample t-Test	Bare hand	1 rubber band	2 rubber bands	3 rubber bands
Bare hand	-	0.91	<0.001	<0.001
1 rubber band	-	-	<0.001	<0.001
2 rubber bands	-	-	-	<0.001

In [Table 6.1](#), it can be seen that the landmark mean area decrease in Mix condition, and in [Table 6.2](#), it can be seen that the comparison of the mean area between different experiments passed the Two Sample t-Test, i.e., the P-value of all the experiments for two-by-two comparison was less than the confidence level of 0.05, which means that the mean area between different groups The mean area of Hand Peak decreased with the increase of rubber bands.

The mean area of different experiments tends to be more stable in [Figure 6.1](#) after landmark based registration. The area in [Table 6.1](#) found Bare hand and hand with 1 rubber band were found to be very close to each other. By Two Samples t-Test D, except for this group of data, the P-value of all groups is close to 0, which is much smaller than the confidence level of 0.05, indicating that the mean values between different groups are significantly different. It can be tentatively concluded that the area of the hand after landmark based registration provides better comparative results when rubber bands are increased to a certain number. However, since there is no more significant difference in

the results of non-landmark based registration, the robustness of different methods under different segmentation cases will be compared later.

6.1.1.2 Stable condition compare

When the experimental hand was recorded in a fixed position on the landmark of the hand, the comparison of the average area of the hand after landmark based registration of the bare hand and hand with 1 rubber band could not pass Two Sample t-Test in [Table 6.6](#). However, all other cases could pass the statistical test and the experimental results were consistent with the hypothesis from [Table 6.4](#) and [Table 6.4](#).

Table 6.4: Experiment stable conditions and Hand Peak area mean

Experiment NO	Condition(stable)	Peak Area original mean	Peak Area mean after registration
1-10	Bare hand	93.50	99.11
21-30	1 rubber band	60.84	97.51
41-50	2 rubber bands	42.25	68.21
61-70	3 rubber bands	29.18	49.52

Table 6.5: P-value of different experiments original Peak mean area compare by Two-Sample t-Test at stable condition

Two-Sample t-Test	Bare hand	1 rubber band	2 rubber bands	3 rubber bands
Bare hand	-	<0.001	<0.001	<0.001
1 rubber band	-	-	<0.001	<0.001
2 rubber bands	-	-	-	<0.001

Table 6.6: P-value of different experiments after registration Peak mean area compare by Two-Sample t-Test at stable condition

Two-Sample t-Test	Bare hand	1 rubber band	2 rubber bands	3 rubber bands
Bare hand	-	0.34	<0.001	<0.001
1 rubber band	-	-	<0.001	<0.001
2 rubber bands	-	-	-	<0.001

The only reason why the hand could not pass the test was that one rubber maybe not affect the change in the area of the Peak-Valley exercise, but when the number of rubber bands increased to two, the movement of the hand received greater restriction and the area of the hand decreased again. Another situation is that the palm of the hand becomes smaller when doing the hand Peak-Valley exercise, and the area of the hand becomes larger after landmark based registration, and only when the difference in hand area is large, this method can steadily distinguish the area between different landmarks.

6.1.1.3 Fluctuate condition compare

When the distance of the user's hand from the camera is random for each test. Although the original landmark area in [Table 6.7](#) still decreased with rubber bands increased, their

average area results [Table 6.7](#) could not pass the Two Sample t-Test, because their hand mean compare P-value all above Confidence level 0.05. However, most of the average area using landmark based registration still passed the Two Sample t-Test in [Table 6.8](#), and confidence level $\alpha \approx 0 < 0.05$. The only group that did not pass the test here was still the bare hand and hand with one rubber band case, which confidence level $\alpha = 0.81 > 0.05$. The reason for this situation is the same as the hypothesis analyzed above.

Table 6.7: Experiment fluctuate conditions and Hand Peak area mean

Experiment NO	Condition (fluctuate)	Peak Area original mean	Peak Area mean after registration
11-20	Bare hand	46.51	104.30
31-40	1 rubber band	36.73	105.34
51-60	2 rubber bands	30.54	78.03
71-80	3 rubber bands	21.00	45.03

Table 6.8: P-value of different experiments original Peak mean area compare by Two-Sample t-Test at fluctuate condition

Two-Sample t-Test	Bare hand	1 rubber band	2 rubber bands	3 rubber bands
Bare hand	-	0.47	0.23	0.07
1 rubber band	-	-	0.52	0.11
2 rubber bands	-	-	-	0.31

Table 6.9: P-value of different experiments after registration Peak mean area compare by Two-Sample t-Test at fluctuate condition

Two-Sample t-Test	Bare hand	1 rubber band	2 rubber bands	3 rubber bands
Bare hand	-	0.81	<0.001	<0.001
1 rubber band	-	-	<0.001	<0.001
2 rubber bands	-	-	-	<0.001

The experimental comparison of the fluctuating condition, combined with the results of the mix and stable conditions above, shows that after using the landmark based registration method, the hand tracking method used in this paper is more stable than the direct calculation of the hand landmark area. The hand tracking method is more robust than the direct calculation of hand landmark area. Using this method, it is possible to more accurately detect the hand's status change when the Hand Peak area changes significantly. Thus, we know the change of hand dexterity.

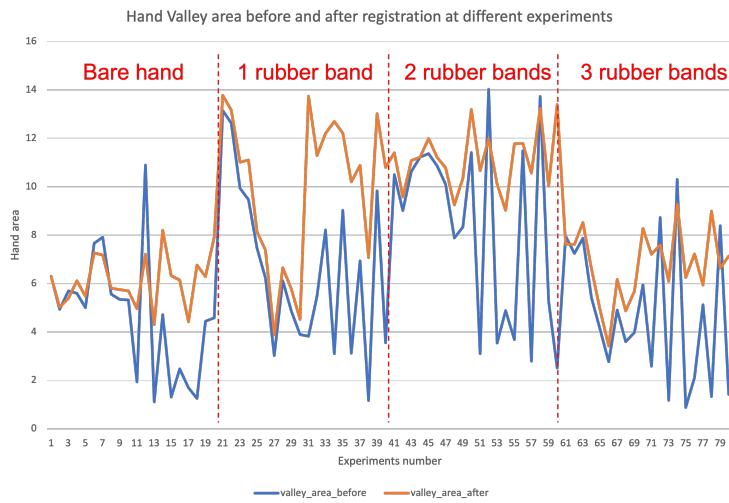


Figure 6.2: Hand Valley area compare by bare hand and hand with 1 to 3 rubber bands in repeat 80 times experiments

Table 6.10: Experiment mix conditions and Hand Valley area mean

Experiment NO	Condition(mix)	Valley Area original mean	Valley mean after registration
1-20	Bare hand	3.44	6.25
21-40	1 rubber band	5.43	11.41
41-60	2 rubber bands	6.50	11.25
61-80	3 rubber bands	4.21	7.24

Table 6.11: P-value of different experiments original Valley mean area compare by Two-Sample t-Test at mix condition

Two-Sample t-Test	Bare hand	1 rubber band	2 rubber bands	3 rubber bands
Bare hand	-	0.15	0.10	0.61
1 rubber band	-	-	0.54	0.42
2 rubber bands	-	-	-	0.24

Table 6.12: P-value of different experiments after registration Valley mean area compare by Two-Sample t-Test at mix condition

Two-Sample t-Test	Bare hand	1 rubber band	2 rubber bands	3 rubber bands
Bare hand	-	<0.001	<0.001	0.10
1 rubber band	-	-	0.83	<0.001
2 rubber bands	-	-	-	<0.001

From [Figure 6.2](#), the area changes of the four experiments, the average area of the original hand landmark and the average area after landmark based registration have no obvious characteristics. By analyzing the mean area of different groups through [Table 6.10](#) analysis, the mean area after landmark based registration only met the expectation at 1-3 rubber bands, which is more rubber bands smaller Hand valley area. In [Table 6.11](#), as all the p values are greater than the confidence level $\alpha = 0.05$, which means the different experiments handed the original mean area are equal, it can not be rejected.

Although in [Table 6.12](#), the P-values of the four comparisons were much smaller than the confidence level $\alpha = 0.05$, indicating that the stability of the average hand area improved after landmark based registration. However, the original hypothesis cannot be rejected because the average area of bare hand is lower than the other three groups, and the confidence level $\alpha = 0.83 > 0.05$ for the comparison of hand area of 1 rubber band and 2 rubber bands. Therefore, the original hypothesis that there is no significant difference between the two mean values should also be accepted.

In conclusion, Hand valley area could not be used to directly judge the dexterity of the hand in this experiment. The reason may be that there was only one tester and it was a person with normal hand movement, so the area of hand closure was small causing a large error. If the elderly stroke patient's hand could not be completely closed and participated in the test, different conclusions might be derived.

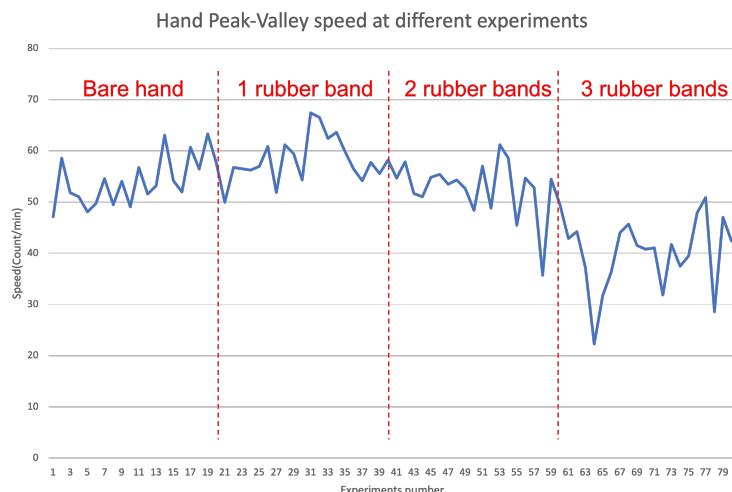


Figure 6.3: Hand Peak-Valley speed compare by bare hand and hand with 1 to 3 rubber bands in repeat 80 times experiments

Table 6.13: Hand Peak-Valley different rubber bands mean speed compare

Experiment NO	Condition	Speed(count/min)
1-20	Bare hand	54.08
21-40	1 rubber band	56.55
41-60	2 rubber bands	51.52
61-80	3 rubber bands	37.13

Table 6.14: P-value of different experiments speed compare by Two-Sample t-Test

Two-Sample t-Test	Bare hand	1 rubber band	2 rubber bands	3 rubber bands
Bare hand	-	0.10	0.11	<0.001
1 rubber band	-	-	<0.01	<0.001
2 rubber bands	-	-	-	<0.001

In different groups of speed tests, the null hypothesis is that the speed of the Peak-Valley exercise is the same for different conditions of the hand. The expected result is that the speed of the Peak-Valley exercise of the hand decreases as the rubber band increases. Two Sample t-test was done to verify whether the means were the same between the two groups by comparing the means in different conditions.

From the observation of [Figure 6.3](#), it can be seen that the average speed tends to decrease with the increase of rubber bands. In [Table 6.13](#), the average speed of Bare hand and 1 rubber band are close to each other. By Two Sample t-test, the mean values between different experiments were compared, and only when bare hand was compared with 1-2 rubber bands respectively, the confidence level was $\alpha = 0.10, 0.11 < 0.05$ in [Table 6.14](#), and the original hypothesis could not be rejected. However, when comparing bare hand with 3 rubber bands, the confidence level is $\alpha = 0.00$, and the original hypothesis is rejected, indicating that the mean values of the two are significantly different. Also by [Table 6.14](#), it can be seen that when the rubber bands are above 1, the confidence level $\alpha < 0.05$, rejecting the original hypothesis, indicating that the speed decreases with the increase of rubber bands at 1-3 rubber bands.

6.1.1.4 Registration based group compare

Based on the conclusions of the above repeated experiments under different conditions, it is concluded that the average area will have better stability after landmark based registration. Now what is explored here is the impact of choosing different reference groups for landmark based registration on the stability of experimental results under different rubber bands in one experiment.

After registration with group1, the peak result In [Table 6.15](#) shows FRE decrease a lot, e.g. Group1's FRE sum decrease from 12050.7 to 11.34, which means the 21 hand joints aligned well after 5 palm joints landmark based registration; the standard deviation(std) keep at same level or decrease,e.g. the group2 hand area std decrease from 5.44 to 3.68. Due to the translation, rotation and scale by registration, the hand average area changed. The [Figure 4.19](#) and [Figure 4.20](#) display the progress of group landmark based registration.

In [Figure 6.4](#), each plot contains the whole group hand area of peak or valley. The registration progress will reduce the influence of hand move close or far way scale problem like [Figure 4.16](#) by find optimized translation, rotation, and scale parameters for each peak and valley landmark.

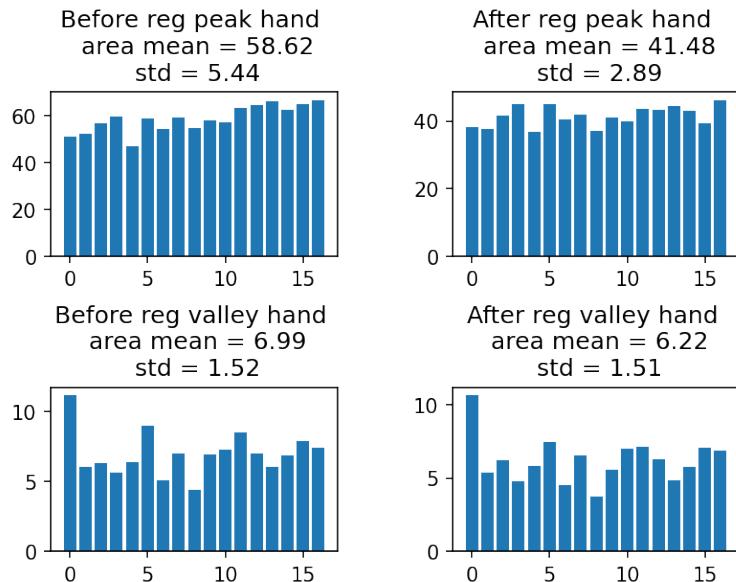


Figure 6.4: Group3 hand with one rubber band register to hand with three rubber band data result. The left column is result before landmark based registration, the right column is after

Table 6.15: Landmark based registration by 5 palm joint based on Bare Hand

Parameter	Peak					
	Before registration			After registration		
	Avg area	std	FRE sum	Avg area	std	FRE sum
Bare hand	53.52	3.01	12050.7	51.36	3.49	11.34
1 rubber band	58.62	5.44	26692	52.46	3.68	47.63
2 rubber bands	46.34	6.14	75256.9	40.59	4.19	205.71
3 rubber bands	25.37	4.27	46701.9	35.85	4.34	106.47

In Table 6.16, despite the decrease in FRE, the small area of the Valley-shaped landmark cannot be used to analyze the change in hand area of the average user in this case. However, there is still a potential value in analyzing the area change of the Valley for special users whose hands cannot be closed. In this experiment, the tester owns a normal function hand, the later analysis will focus on the hand peak area.

Table 6.16: Landmark based registration by 5 palm joint based on Group1

Parameter	Valley					
	Before registration			After registration		
	Avg area	std	FRE sum	Avg area	std	FRE sum
Bare hand	4	1.25	22580.78	4.24	1.3	189.21
1 rubber band	6.99	1.52	61523.37	5.76	1.42	162.35
2 rubber bands	7.03	2.61	146657.51	4.97	2.35	928.81
3 rubber bands	4.92	1.7	60686.98	5.89	2.36	137.57

Table 6.17: Hand peak area compare based on different register group

Based group	Bare hand	1 rubber band	2 rubber bands	3 rubber bands	Speed (count/min)
Bare hand	51.36	54.01	60.34	40.55	38.11
1 rubber band	52.46	55.31	62.13	41.48	34.99
2 rubber bands	40.59	42.9	48.35	32.33	31.96
3 rubber bands	35.85	37.73	42.22	28.62	33.84

To test that the different selection of base projection groups does not affect the final results of the experiment, the results of the above four experimental groups [Figure 5.1](#) were projected onto different groups as shown in the [Table 6.17](#). Although there is a significant change in the hand average area when the landmark of the same group is projected to different groups; however, when projected to the same group, the hand area can still be obviously distinguished as the average area of the Peak of the hand decreases gradually with the increase of the rubber band. Nevertheless, group1 and Group2 showed different results from the hypothesis, and the expected result was that the average area of the hand in Group2: the hand with one rubber band should have a smaller hand average area, but in reality it was slightly larger than Group1 without rubber band restriction. The possible reason for this observation is that one rubber band is not enough to restrict a normal user to do simple hand movements of peak and valley. When there is more than one rubber band, the limitation of hand movement increases significantly and the average area of the hand peak decreases as expected.

According to the comparison in [Table 6.17](#), each column is the average area of one experiment registered with different experiments. The calculated average area data are multiple hand peak landmarks for each experiment. To compare the consistency of experimental results across different average areas after different registrations, two sample t-tests were performed for each column.

By comparing [Table 6.18](#), [Table 6.19](#), [Table 6.20](#) and [Table 6.21](#), it is observed that the confidence levels of bare hand and hand with 1 rubber band accept the null hypoth-

esis, which is their mean peak areas are equal. Because the confidence level $\alpha = 0.39, 0.33, 0.24, 0.36 > 0.05$ in four different registrations.

In the rest of cases, the confidence level $\alpha = 0.00 < 0.05$, and the null hypothesis is rejected, i.e., the mean values between the two groups are not the same. According to the comparison of the mean area size in [Table 6.17](#), it can be concluded that when the rubber bands are increased to more than 2, the peak area of the hand decreases significantly regardless of the group of data to which the registration is applied.

[Table 6.18: P-value of two sample t-test on Landmark based bare hand registration](#)

Two-Sample t-Test	Bare hand	1 rubber band	2 rubber bands	3 rubber bands
Bare hand	-	0.39	<0.001	<0.001
1 rubber band	-	-	<0.001	<0.001
2 rubber bands	-	-	-	<0.01

[Table 6.19: P-value of two sample t-test on landmark based 1 rubber band registration](#)

Two-Sample t-Test	Bare hand	1 rubber band	2 rubber bands	3 rubber bands
Bare hand	-	0.33	<0.001	<0.001
1 rubber band	-	-	<0.001	<0.001
2 rubber bands	-	-	-	<0.01

[Table 6.20: P-value of two sample t-test on landmark based 2 rubber bands registration](#)

Two-Sample t-Test	Bare hand	1 rubber band	2 rubber bands	3 rubber bands
Bare hand	-	0.24	<0.001	<0.001
1 rubber band	-	-	<0.001	<0.001
2 rubber bands	-	-	-	<0.01

[Table 6.21: P-value of two sample t-test on landmark based 3 rubber bands registration](#)

Two-Sample t-Test	Bare hand	1 rubber band	2 rubber bands	3 rubber bands
Bare hand	-	0.36	<0.001	<0.001
1 rubber band	-	-	<0.001	<0.001
2 rubber bands	-	-	-	<0.01

The speed of hand movement decreases, which to some extent satisfies the hypothesis that the speed decreases with the increase in resistance due to the number of rubber bands. However, in Group4, the speed did increase relative to Group3. There are two hypotheses, one is that the speed of hand movement is not only hand, the external force limit increases and then decreases, the other case is that as the external force limit increases to a certain level, because the hand does not change significantly resulting in an

increase in the speed of movement, but the overall amplitude of movement is smaller. As for which the hypothesis is correct, more experimental data are needed to prove it in the future.

6.1.2 Different people tracking analysis

The second experiment will compare different users' wear with one or more rubber bands compared to the bare hand with no any rubber band. The main purpose of this test is to verify that the hand tracking system can work properly with different users.

There are some testers attending this experiment in Health & Rehabilitation Scandinavia ¹ at Copenhagen Bella Center on 2021 November 16th. Health & Rehab Scandinavia focused on professional knowledge sharing, inspiration and networking in the industry for assistive technology and welfare technology. My project as part of DTU management rehabilitation projects attend this exhibition. The main purpose of attending the show was to get feedback from different user groups on hand dexterity assessments that were tracked and recorded over time. [Figure 6.5](#) is on site and different participants attended in the experimental tests. The following is the feedback from different testers.



Figure 6.5: Participants of different types and ages. Left: I am introducing an older person to the theory of how Hand tracking works. Right top: I am giving a presentation to a young woman in a wheelchair on how to participate in the experiment and how to understand the results. Right bottom: A healthy young woman showing a keen interest in Hand tracking as I am introducing how to participate in the experiment.

The elderly([Figure 6.5:Left](#)): As individuals grow older, there is an urgent need for daily body monitoring systems to help them understand changes in their physical status. The probability of stroke increases with age, and stroke detection has always been complex and requires visits to professional physicians. However, as an older adult, frequent visits to the doctor can increase the risk and cost of travel. The older adults in A showed strong interest in MediaPipe hand tracking. The reasons are: First, although the test was conducted using a computer at the time, it could be migrated to a smartphone in the future. Convenient, high-frequency and independent daily health tracking is important for the daily use of older people. Second, compared with a smartwatch that can detect heartbeat and oxygen saturation, our method can detect hand dexterity by external motion, which is

¹<https://www.health-rehab.dk/>

very attractive for elderly people with self-examine, and does not require the purchase of additional testing equipment, reducing the cost of use. Thirdly, the experimental results detected in this experiment can be compared to make the elderly users understand the purpose of the test visually. At present, there is no relevant, simple and effective testing method for the elderly to use in daily life.

The lady in the wheelchair([Figure 6.5](#):Right top): also expressed great interest in hand tracking because of the importance of the hand in life. Despite being in a wheelchair and participating in this activity, the results of the experiment showed her that hand dexterity can be recorded and compared by capturing the position of joints through artificial intelligence methods, and that data related to hand movement can be calculated and output through computers and mobile phones. This approach could be of great help in assisting those in need of rehabilitation to independently determine the health status of their hands. If this technology is widely used in mobile phone software in the future, it will be more convenient for different people to use.

Lady([Figure 6.5](#):Right bottom): Because what was presented in the table caught her attention, despite having healthy hands now, in the future each person will have decreased hand dexterity as they grow older or suffer from disease. There are no existing methods that can easily and directly measure hand dexterity quantitatively and track it over time. After her introduction to the program, she was able to understand that when hand stretch is limited by the rubber band, the speed of hand movement and the area of the hand will be affected. For normal people, the topic of concern for physical health still occupies an important place in life. She said she would increase her attention to the development of artificial intelligence body motion capture technology and its impact on the lives of ordinary people.

In this experiment, the process of user participation throughout the experiment is divided into four main steps.

1. explain to the test the reason why hand tracking is needed: to achieve a low-cost real-time tracking solution to help elderly stroke patients track the health status of their hands on a long-term scale.
2. introduce to the user how to do the Peak of hand and Valley's exercise, and then start the project to record the landmark of hand. after the recording, introduce to the user, the fundamental information of the recording: recording duration, data size and anonymous storage method.
3. Let the user conduct a second experiment, this time to simulate the restricted hand movement, and give the user one or more rubber bands to simulate this situation. After the user gets used to it, give the user to start recording the landmark of the hand for the second time.
4. Explain to the user the results of the change in hand speed and hand area between the two exercises, and the theoretical process of how to get the results. Then re-present the stored landmark data to the user, informing that the data is stored anonymously and does not contain any user photos. After obtaining the user's consent, an agreement is signed for the donation of the data for scientific use.

Data in the [Table 6.22](#) were collected at [Health & Rehab Scandinavia](#) from users of different ages and genders. In total, ten data were collected from participant donations, five of which could be used for experimental analysis. The other reasons for not being able to be used for analysis are: first, the users could not proficiently perform the Hand Peak and Val-

Table 6.22: Different people two group experiments hand Peak compare

Experiment condition		Bare hand (Peak)			Hand with rubber band (Peak)		
Group		Area before	Area after	Speed	Area before	Area after	Speed
1		25.97	25.77	12.48	8.86	14.93	14.9
2		71.89	50.71	40.4	73.57	47.79	16.02
3		49.4	40.41	45.33	22.71	24.2	65.39
4		24.18	25.38	49.56	9.81	9.75	39.89
5		30.8	37.45	50.73	24.1	33.37	48.71

Table 6.23: P-value of different participants two group experiments compare

Participants	Before registration	After registration
1	<0.001	<0.01
2	0.81	0.71
3	<0.001	<0.001
4	<0.001	<0.001
5	<0.01	0.01

key exercise in their first experience; second, there were complex lights and objects in the background, which affected the judgment of the model; third, the occasional passerby's hand would replace the tester's, which had a considerable impact on the experimental results. The improvement measure is to introduce the project to the user and then let the user try hand tracking first, and then let the user participate in the formal experimental analysis when the user is adept.

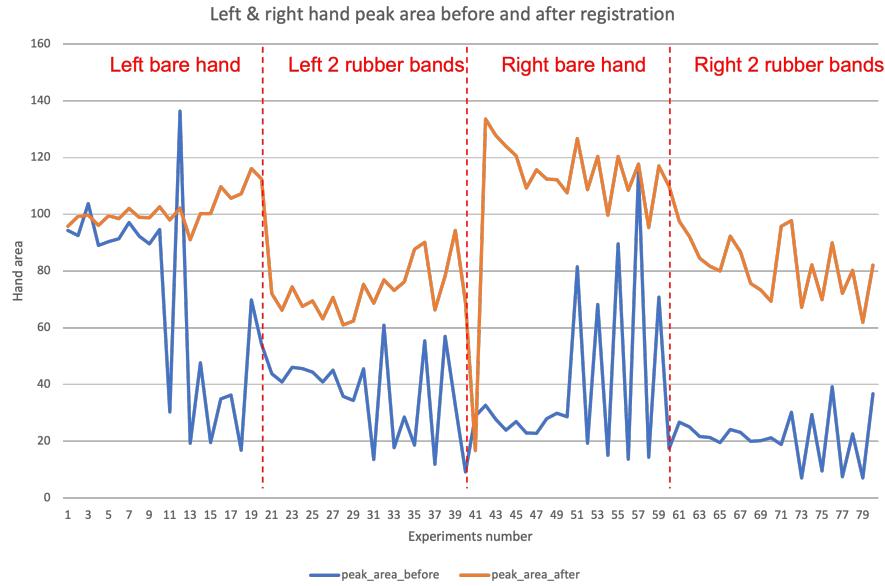
The results of five different participating users were analyzed to detect whether the average landmark area recorded by MediaPipe before and after landmark based registration is satisfied by the reduction of the average landmark area after wearing rubber bands.

The null hypothesis was tested by conducting two sample t-tests for each participant, for both before and after registration, to see if they had the same mean value in Table 6.22. Only the null hypothesis for participant 2 was accepted, with a confidence level of $\alpha = 0.81 > 0.05$ in Table 6.23. Although the area of the hand in the second experiment with rubber bands was smaller than the first group of bare hand, the area was closer, $50.71 > 57.79$ from Table 6.22. The other four participants had a confidence level of $\alpha = 0.00 < 0.05$ in Table 6.23, and the area difference was larger, that means MediaPipe hand landmark tracking can measure the decrease of hand dexterity change when the area difference is larger.

6.1.3 Mirror hand data analysis

In the 80 replicate experiments, the hand valley area of this experiment was not analyzed primarily because only the hand peak area had significant data differences. In Figure 6.6, the blue line before landmark based registration has large fluctuations; after landmark

based registration, the stability of the experimental area under different conditions is improved. By observing [Figure 6.6](#), it can be concluded that the peak area of the right hand is higher than that of the left hand under the same conditions. The same conclusion can be drawn in [Table 6.24](#) by comparing the data directly. That is, the bare hand peak area of the right hand is larger than that of the left hand; the peak area of the right hand with 2 rubber bands is also larger than that of the left hand.



[Figure 6.6](#): Left and right hand repeat experiments' peak area compare, each condition repeat 20 times, the total experiments is 80 times.

In [Table 6.25](#), although the average area of the bare hand peak area of the right hand is larger than that of the left hand, the P-value of this data set is $\alpha = 0.13 > 0.05$, and the confidence level cannot pass the two-sample t-test experiment. However, the confidence level of $\alpha = 0.00 < 0.05$ for the right hand with 2 rubber bands indicates that the original hypothesis is not true and there is a difference in the average area of the left hand and right hand. The Peak area of the right hand was greater, indicating that the dexterity of the right hand extension is greater than that of the left hand.

[Table 6.24](#): Left and right hand Peak before and after landmark based registration and speed of 20 times repeat experiments, the total experiments number is 80 times.

Experiment	Hand Peak mean data		
	Condition	Area before	Area after
Left bare hand	70.00	101.71	54.08
left 2 rubber bands	36.40	73.12	51.52
Right bare hand	38.84	110.20	65.72
Right 2 rubber bands	21.55	81.63	50.80

Through the observation of the average speed in [Figure 6.7](#) and the actual speed in [Figure 6.6](#), the basic conclusion is that the speed of the right bare hand is the highest. Mean-

Table 6.25: P-value of two sample t-test on left and right hand area compare after landmark based registration

Two-Sample t-Test	Left bare hand	left 2 rubber bands	Right bare hand	Right 2 rubber bands
Left bare hand	-	<0.001	0.13	<0.001
left 2 rubber bands	-	-	<0.001	<0.01
Right bare hand	-	-	-	<0.001

while, the p-value of the right bare hand compared with the left hand was $\alpha = 0.00 < 0.05$ by the two-sample t-test in Table 6.26, indicating a significant difference between the mean values of the two. Although there was no significant difference in the p-value of the 2 rubber bands, the movement speed of the right hand was higher than that of the left hand when combined with the actual situation without the limitation of rubber bands.

This comparison experiment of the peak area and speed of the right and left hands can help the rehabilitation physiotherapist or physician to better determine the changes in the motion of the right and left hands of stroke patients and give better rehabilitation advice in time.

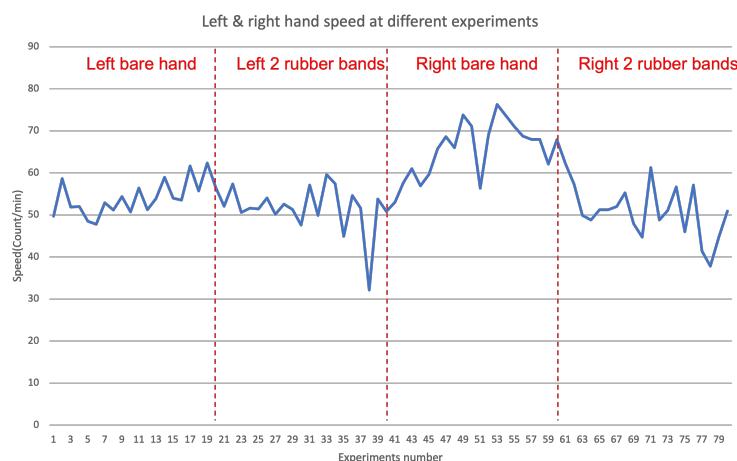


Figure 6.7: Left and right hand speed compare at different conditions, each condition repeat 20 times, the total experiments is 80 times.

Table 6.26: P-value of two sample t-test on left and right hand speed compare

Two-Sample t-Test	Left bare hand	left 2 rubber bands	Right bare hand	Right 2 rubber bands
Left bare hand	-	0.11	<0.001	0.06
left 2 rubber bands	-	-	<0.001	0.71
Right bare hand	-	-	-	<0.001

6.2 Dumb-bell exercise analysis

In this experiment, by conducting 10 replicate experiments for each of the three load conditions, the total number of experiments is 30. The average speed of each individual

experiment was recorded separately, and the average speed of all experiments is shown in [Table 6.27](#). The average speed fluctuates under different conditions, and combined with the data in [Table 6.27](#), it is known that the speed of the exercise decreases as the load increases. According to the P-value analysis in [Table 6.28](#), when heavy load, the confidence water product P-value $\alpha = 0.00, 0.047 < 0.05$. The original hypothesis is not true, which means no load, light load, and heavy load have different mean values. However, the average speed of no load and light load, despite $36.23 > 35.65$, has the same mean speed due to the close values because of P-value $0.74 > 0.05$. The original hypothesis is true and both have the same mean value.

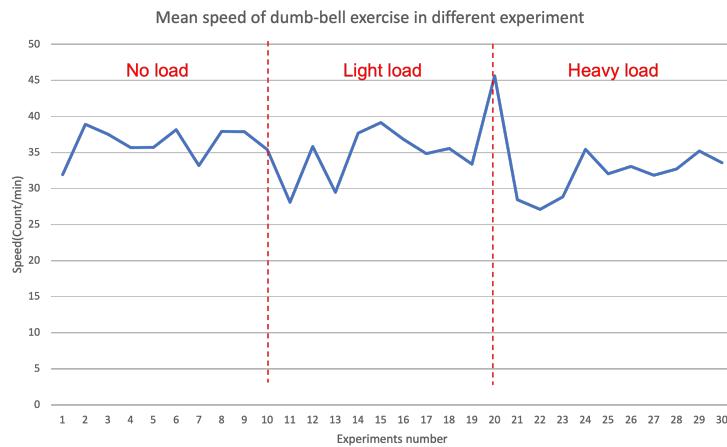


Figure 6.8: Mean speed of dumb-bell experiments in three different conditions: no load, light load and heavy load.

Table 6.27: Mean speed data in dumb-bell experiment

Condition	Experiments number	Speed
No load	1-10	36.23
Light load	11-20	35.65
Heavy load	21-30	31.83

Table 6.28: P-value of two sample t-test for mean speed compare in dumb-bell experiment

Two-Sample t-Test	No load	Light load	Heavy load
No load	-	0.74	<0.01
Light load	-	-	0.047

In summary, the MediaPipe Pose model can track the change in arm muscle strength by calculating the mean speed of dumb-bell movement only when the load is increased to a certain weight.

7 Discussion

This thesis explores different methods of hand tracking and finally settles on using a MediaPipe based deep learning approach to track hand and body movement data. Because it works smoothly on mobile devices such as laptops, smartphones, and tablets, users do not need to spend extra money on the device and the tracking results are real-time and accurate. This provides convenience and reduces the burden of daily use for users. By creating a digital twin for stroke patients, the landmarks of the user's hand and body can be collected in real time. The concept of Digital twin assistant sysmten is used to manage the interaction of data between virtual and real. The core contribution of this thesis is to provide a method that enables the quantification of body movement data that can be collected independently for stroke patients to help motivate and understand the changes in their physical condition. It also provides rehabilitation trainers and physicians with a better overview of changes in the patient's physical condition over time, avoiding the influence of subjective factors through a quantitative approach.

The Landmark base registration method compares the hand area of different experiments to analyze changes in hand dexterity. Three different experiments were conducted to examine the stability of the method under different conditions.

In Experiment 1, 80 sets of experimental data were analyzed by comparing four sets of 0-3 rubber bands. The data was taken from the same hand of the same user and repeated 20 times in each condition, the first 10 times the hand was fixed in the same position, and the other 10 times the hand would do peak-valley exercise at different distances from the camera to simulate the uncertainty of the distance of each experiment in reality. The mean area and velocity of the hands were cross-compared between different groups, combined with two sample t-tests. It was concluded that the hand mean area based on landmark based registration has better stability under different conditions. The change of hand dexterity when wearing different rubber bands can be distinguished more accurately. In this experiment, the landmark based registration was also performed on four groups wearing different numbers of rubber bands from 0 to 3. The experimental results showed that the same conclusion was reached regardless of the group to which the registration was applied. More than two rubber bands significantly reduced the average peak area of the hand. Similarly, for the analysis of speed, there is a similar conclusion that more than one rubber band will weaken the speed of the hand doing peak-valley exercise.

In Experiment 2, the tests of bare hand and hand with rubber bands were conducted with five different users. It can be concluded that although the mean hand of one of the five different users failed the two sample t-test test, the remaining four users satisfied the statistical test. It shows that the method provided in this thesis is applicable to different people in most cases.

In Experiment 3, for the same user's left and right hand, 20 experiments were repeated in the bare hand and hand with 2 rubber bands cases for the left and right hand, respectively, for a total of 80 experiments. Although the value of the peak area right hand was greater in the case of bare hand, the average value failed to pass the test experiment. That does not indicate that the flexibility in the right hand is greater than that in the left hand. However, in the case of the hand with 2 rubber bands, the peak area of the right hand was significantly larger than that of the left hand. This experiment shows that the mirror hand can be used

to detect the difference between the user's left and right hand dexterity. It can help stroke patients to compare the changes of left and right hand dexterity.

In the dumb-bell exercise experiment, 30 repetitions of no load, light load and heavy load were compared with cross-averages for the same test participant. Then after two sample t-tests, it was concluded that the MediaPipe Pose model can capture the landmark of the body and calculate the movement speed to infer the arm muscle movement state when the arm movement is limited by large external forces.

In the experiments of performance testing, hand area tracking and dumbbell motion delay were tested. Although it can be concluded that both methods are stable in real life and can get stable body and hand landmarks while wearing LEGO exoskeleton equipment and normal clothes, the latency was 80-180 ms and the frame rate was 4-10 FPS. Although the running speed does not exceed the ideal of 30 FPS, it is still possible to record and analyze the experimental results on mobile devices.

The digital twin assistant system can be combined with the three experimental methods mentioned in this thesis to obtain the basic data of the body landmark. This system can analyze changes in the flexibility of one hand and both hands over time by hand tracking. The user monitors the usual dumbbell movement through a mobile device, which can be used to compare changes in arm muscle strength. Changes in the physical condition of stroke patients can be detected in time by the data provided by the DTA system. It can also be used to provide feedback on the status of physical flexibility improvement after rehabilitation training. The data accumulated over time also provides rehabilitation trainers and physicians with quantitative data references on the physical movement status of stroke patients.

The method explored in this paper has some weaknesses. In both the experiments of hand tracking one-handed and left- and right-handed comparisons, it takes more than 2 rubber bands for the average area and movement speed of the hand to be significantly different. This indicates that the method cannot accurately detect the change in hand dexterity, but only when there is a large change in hand dexterity. In the dumbbell exercise experiment, there was no significant difference in speed between the no load and small load comparison experiments. When using a big load, the system detects a significant change in speed. The method used in this experiment can provide a basic reference for rehabilitation. For more advanced testing, it is still necessary to conduct other tests with the guidance of professionals.

The thesis uses these data to provide digital twin data support for stroke patients by establishing different methods to obtain different body data. The hand landmark data is used to compare the hand area using landmark based registration, combined with the peek-valley movement speed to evaluate hand dexterity. The velocity test of the dumbbell movement also allows the stroke patient to understand the strength changes of the arm muscles. Because these are quantitative data, direct comparisons can be made directly through numerical magnitudes rather than individual subjective perceptions. Thus, it helps stroke patients to develop better scientific rehabilitation training.

8 Conclusion

This thesis measures hand dexterity and arm strength changes through three different experiments based on the MediaPipe hand and pose models. The aim is to help stroke patients to better assess their rehabilitation training. In the hand tracking experiment, it was concluded that when there was a substantial change in hand dexterity, it was known by observing the resultant change in hand area and speed. In the dumbbell exercise experiment, the change in movement speed was obtained by observing different load situations. It can be learned that when the load is large enough, the movement speed of the arm decreases significantly. Although the performance analysis experiments were performed on a laptop with high latency, the change in body flexibility could still be analyzed. Although the accuracy is not as high as professional equipment, it is still a cheap and convenient way to test the flexibility of the body for patients at home using mobile devices.

Bibliography

- [1] W. H. Organization, *The world health report 2003: shaping the future*. World Health Organization, 2003.
- [2] R. Bonita, S. Mendis, T. Truelson, J. Bogousslavsky, J. Toole, and F. Yatsu, "The global stroke initiative," *The Lancet Neurology*, vol. 3, no. 7, pp. 391–393, 2004.
- [3] V. Parker, D. Wade, and R. L. Hewer, "Loss of arm function after stroke: Measurement, frequency, and recovery," *International rehabilitation medicine*, vol. 8, no. 2, pp. 69–73, 1986.
- [4] A. El Saddik, "Digital twins: The convergence of multimedia technologies," *IEEE multimedia*, vol. 25, no. 2, pp. 87–92, 2018.
- [5] R. I. Hartley and P. Sturm, "Triangulation," *Computer vision and image understanding*, vol. 68, no. 2, pp. 146–157, 1997.
- [6] M. Kolsch and M. Turk, "Fast 2d hand tracking with flocks of features and multi-cue integration," in *2004 Conference on Computer Vision and Pattern Recognition Workshop*, IEEE, 2004, pp. 158–158.
- [7] M. Kolsch and M. A. Turk, "Robust hand detection.," in *FGR*, vol. 4, 2004, pp. 614–619.
- [8] W. Hürst and C. Van Wezel, "Gesture-based interaction via finger tracking for mobile augmented reality," *Multimedia Tools and Applications*, vol. 62, no. 1, pp. 233–258, 2013.
- [9] Y. Jang, S.-T. Noh, H. J. Chang, T.-K. Kim, and W. Woo, "3d finger cape: Clicking action and position estimation under self-occlusions in egocentric viewpoint," *IEEE Transactions on Visualization and Computer Graphics*, vol. 21, no. 4, pp. 501–510, 2015.
- [10] R.-H. Liang and M. Ouhyoung, "A real-time continuous gesture recognition system for sign language," in *Proceedings third IEEE international conference on automatic face and gesture recognition*, IEEE, 1998, pp. 558–567.
- [11] A. Markussen, M. R. Jakobsen, and K. Hornbæk, "Vulture: A mid-air word-gesture keyboard," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2014, pp. 1073–1082.
- [12] S. Sridhar, A. M. Feit, C. Theobalt, and A. Oulasvirta, "Investigating the dexterity of multi-finger input for mid-air text entry," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2015, pp. 3643–3652.
- [13] T. Starner, J. Weaver, and A. Pentland, "Real-time american sign language recognition using desk and wearable computer based video," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 20, no. 12, pp. 1371–1375, 1998.
- [14] M. Kurien, M.-K. Kim, M. Kopsida, and I. Brilakis, "Real-time simulation of construction workers using combined human body and hand tracking for robotic construction worker system," *Automation in Construction*, vol. 86, pp. 125–137, 2018.
- [15] L. T. Taverne, M. Cognolato, T. Bützer, R. Gassert, and O. Hilliges, "Video-based prediction of hand-grasp preshaping with application to prosthesis control," in *2019 International Conference on Robotics and Automation (ICRA)*, IEEE, 2019, pp. 4975–4982.
- [16] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C.-L. Chang, and M. Grundmann, "MediaPipe Hands: On-device Real-time Hand Tracking," 2020. arXiv: 2006.10214. [Online]. Available: <http://arxiv.org/abs/2006.10214>.

- [17] N. Omarkulov, K. Telegenov, M. Zeinullin, I. Tursynbek, and A. Shintemirov, "Preliminary mechanical design of nu-wrist: A 3-dof self-aligning wrist rehabilitation robot," in *2016 6th IEEE international conference on biomedical robotics and biomechatronics (BioRob)*, IEEE, 2016, pp. 962–967.
- [18] I. Oikonomidis, N. Kyriazis, and A. A. Argyros, "Efficient model-based 3d tracking of hand articulations using kinect.," in *BmVC*, vol. 1, 2011, p. 3.
- [19] C. Xu and L. Cheng, "Efficient hand pose estimation from a single depth image," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 3456–3462.
- [20] S. Sridhar, A. Oulasvirta, and C. Theobalt, "Interactive markerless articulated hand motion tracking using rgb and depth data," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 2456–2463.
- [21] C. Qian, X. Sun, Y. Wei, X. Tang, and J. Sun, "Realtime and robust hand tracking from depth," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1106–1113.
- [22] S. Khamis, J. Taylor, J. Shotton, C. Keskin, S. Izadi, and A. Fitzgibbon, "Learning an efficient model of hand shape variation from depth images," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 2540–2548.
- [23] L. Ge, H. Liang, J. Yuan, and D. Thalmann, "Robust 3d hand pose estimation in single depth images: From single-view cnn to multi-view cnns," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3593–3601.
- [24] S. Yuan, G. Garcia-Hernando, B. Stenger, G. Moon, J. Y. Chang, K. M. Lee, P. Molchanov, J. Kautz, S. Honari, L. Ge, et al., "Depth-based 3d hand pose estimation: From current achievements to future goals," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2636–2645.
- [25] S. Li and D. Lee, "Point-to-pose voting based hand pose estimation using residual permutation equivariant layer," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11 927–11 936.
- [26] J. Malik, I. Abdelaziz, A. Elhayek, S. Shimada, S. A. Ali, V. Golyanik, C. Theobalt, and D. Stricker, "Handvoxnet: Deep voxel-based network for 3d hand shape and pose estimation from a single depth map," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 7113–7122.
- [27] L. Fang, X. Liu, L. Liu, H. Xu, and W. Kang, "Jgr-p2o: Joint graph reasoning based pixel-to-offset prediction network for 3d hand pose estimation from a single depth image," in *European Conference on Computer Vision*, Springer, 2020, pp. 120–137.
- [28] L. Huang, J. Tan, J. Meng, J. Liu, and J. Yuan, "Hot-net: Non-autoregressive transformer for 3d hand-object pose estimation," in *Proceedings of the 28th ACM International Conference on Multimedia*, 2020, pp. 3136–3145.
- [29] L. Huang, J. Tan, J. Liu, and J. Yuan, "Hand-transformer: Non-autoregressive structured modeling for 3d hand pose estimation," in *European Conference on Computer Vision*, Springer, 2020, pp. 17–33.
- [30] "Accurate, robust, and flexible realtime hand tracking," *Conf. Hum. Factors Comput. Syst. - Proc.*, vol. 2015-April, pp. 3633–3642, 2015. DOI: [10.1145/2702123.2702179](https://doi.org/10.1145/2702123.2702179).
- [31] L. Ballan, A. Taneja, J. Gall, L. Van Gool, and M. Pollefeys, "Motion capture of hands in action using discriminative salient points," in *European Conference on Computer Vision*, Springer, 2012, pp. 640–653.
- [32] F. Gomez-Donoso, S. Orts-Escalano, and M. Cazorla, "Large-scale multiview 3d hand pose dataset," *Image and Vision Computing*, vol. 81, pp. 25–33, 2019.

- [33] R. Wang, S. Paris, and J. Popović, “6d hands: Markerless hand-tracking for computer aided design,” in *Proceedings of the 24th annual ACM symposium on User interface software and technology*, 2011, pp. 549–558.
- [34] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, “Hand keypoint detection in single images using multiview bootstrapping,” in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2017, pp. 1145–1153.
- [35] M. Kurien, M.-K. Kim, M. Kopsida, and I. Brilakis, “Real-time simulation of construction workers using combined human body and hand tracking for robotic construction worker system,” *Automation in Construction*, vol. 86, pp. 125–137, 2018.
- [36] F. Mueller, F. Bernard, O. Sotnychenko, D. Mehta, S. Sridhar, D. Casas, and C. Theobalt, “GANerated Hands for Real-Time 3D Hand Tracking from Monocular RGB,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 49–59, 2018, ISSN: 10636919. DOI: [10.1109/CVPR.2018.00013](https://doi.org/10.1109/CVPR.2018.00013). arXiv: [1712.01057](https://arxiv.org/abs/1712.01057).
- [37] S. Han, B. Liu, R. Cabezas, C. D. Twigg, P. Zhang, J. Petkau, T. H. Yu, C. J. Tai, M. Akbay, Z. Wang, A. Nitzan, G. Dong, Y. Ye, L. Tao, C. Wan, and R. Wang, “MEgATrack: Monochrome Egocentric Articulated Hand-Tracking for Virtual Reality,” *ACM Trans. Graph.*, vol. 39, no. 4, 2020, ISSN: 15577368. DOI: [10.1145/3386569.3392452](https://doi.org/10.1145/3386569.3392452).
- [38] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, “Blazepose: On-device real-time body pose tracking,” *arXiv preprint arXiv:2006.10204*, 2020.
- [39] Y. Wang and Y. Liu, “IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY 1 Mask-pose Cascaded CNN for 2D Hand Pose Estimation from Single Color Image,” pp. 1–11, [Online]. Available: <http://www.yangangwang.com..>
- [40] D. Kulon, R. A. Güler, I. Kokkinos, M. Bronstein, and S. Zafeiriou, “Weakly-supervised mesh-convolutional hand reconstruction in the wild,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 4989–4999, 2020, ISSN: 10636919. DOI: [10.1109/CVPR42600.2020.00504](https://doi.org/10.1109/CVPR42600.2020.00504). arXiv: [2004.01946](https://arxiv.org/abs/2004.01946).
- [41] G. Moon, S. I. Yu, H. Wen, T. Shiratori, and K. M. Lee, “InterHand2.6M: A Dataset and Baseline for 3D Interacting Hand Pose Estimation from a Single RGB Image,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12365 LNCS, pp. 548–564, 2020, ISSN: 16113349. DOI: [10.1007/978-3-030-58565-5_33](https://doi.org/10.1007/978-3-030-58565-5_33). arXiv: [2008.09309](https://arxiv.org/abs/2008.09309).
- [42] X. Zhang, H. Huang, J. Tan, H. Xu, C. Yang, G. Peng, L. Wang, and J. Liu, “Hand image understanding via deep multi-task learning,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 11 281–11 292.
- [43] C. Zimmermann, D. Ceylan, J. Yang, B. Russell, M. J. Argus, and T. Brox, “FreiHAND: A dataset for markerless capture of hand pose and shape from single rgb images,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2019-Octob, pp. 813–822, 2019, ISSN: 15505499. DOI: [10.1109/ICCV.2019.00090](https://doi.org/10.1109/ICCV.2019.00090). arXiv: [1909.04349](https://arxiv.org/abs/1909.04349).
- [44] F. Mueller, D. Mehta, O. Sotnychenko, S. Sridhar, D. Casas, and C. Theobalt, “Real-Time Hand Tracking Under Occlusion from an Egocentric RGB-D Sensor,” *Proc. - 2017 IEEE Int. Conf. Comput. Vis. Work. ICCVW 2017*, vol. 2018-Janua, pp. 1284–1293, 2017. DOI: [10.1109/ICCVW.2017.82](https://doi.org/10.1109/ICCVW.2017.82).
- [45] S. Brahmbhatt, C. Tang, C. D. Twigg, C. C. Kemp, and J. Hays, “Contactpose: A dataset of grasps with object contact and hand pose,” in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16*, Springer, 2020, pp. 361–378.

- [46] S. Yuan, Q. Ye, B. Stenger, S. Jain, and T.-K. Kim, “Bighand2. 2m benchmark: Hand pose dataset and state of the art analysis,” pp. 4866–4874, 2017.
- [47] R. Y. Wang and J. Popović, “Real-time hand-tracking with a color glove,” *ACM Trans. Graph.*, vol. 28, no. 3, pp. 1–8, 2009, ISSN: 07300301. DOI: [10.1145 / 1531326.1531369](https://doi.org/10.1145/1531326.1531369).
- [48] F. Farahanipad, M. Rezaei, A. Dillhoff, F. Kamangar, and V. Athitsos, “A pipeline for hand 2-d keypoint localization using unpaired image to image translation,” pp. 226–233, 2021.
- [49] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017.
- [50] F. Tao, H. Zhang, A. Liu, and A. Y. Nee, “Digital Twin in Industry: State-of-the-Art,” *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2405–2415, 2019, ISSN: 15513203. DOI: [10.1109/TII.2018.2873186](https://doi.org/10.1109/TII.2018.2873186).
- [51] P. Hehenberger and D. Bradley, “Mechatronic futures: Challenges and solutions for mechatronic systems and their designers,” *Mechatronic Futures: Challenges and Solutions for Mechatronic Systems and Their Designers*, pp. 1–259, 2016. DOI: [10.1007 / 978-3-319-32156-1](https://doi.org/10.1007/978-3-319-32156-1).
- [52] R. Gámez Díaz, Q. Yu, Y. Ding, F. Laamarti, and A. El Saddik, “Digital twin coaching for physical activities: A survey,” *Sensors*, vol. 20, no. 20, p. 5936, 2020.
- [53] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee, *et al.*, “Mediapipe: A framework for building perception pipelines,” *arXiv preprint arXiv:1906.08172*, 2019.
- [54] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, “SSD: Single shot multibox detector,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9905 LNCS, pp. 21–37, 2016, ISSN: 16113349. DOI: [10.1007 / 978-3-319-46448-0_2](https://doi.org/10.1007/978-3-319-46448-0_2). arXiv: [1512.02325](https://arxiv.org/abs/1512.02325).
- [55] E. Haber and J. Modersitzki, “A multilevel method for image registration,” *SIAM journal on scientific computing*, vol. 27, no. 5, pp. 1594–1607, 2006.
- [56] I. L. Dryden and K. V. Mardia, *Statistical shape analysis: with applications in R*. John Wiley & Sons, 2016, vol. 995.
- [57] K. V. Leemput and R. Larsen, *22525 Course Note Medical Image Analysis*. 2020, pp. 17–21.
- [58] P. H. Schönemann, “A generalized solution of the orthogonal procrustes problem,” *Psychometrika*, vol. 31, no. 1, pp. 1–10, 1966.

A Appendices

A.1 Hand area data

Table A.1: Landmark based registration by 5 palm joint based on Group2

Landmark Type	Peak					
	Before registration			After registration		
Parameter	Avg area	std	FRE sum	Avg area	std	FRE sum
Bare hand	53.52	3.01	47681.6	54.01	3.68	39.67
1 rubber band	58.62	5.44	13880	55.31	3.88	19.2
2 rubber bands	46.34	6.14	71920.1	42.9	4.46	146.25
3 rubber bands	25.37	4.27	93945.9	37.73	4.58	123.59

Table A.2: Landmark based registration by 5 palm joint based on Group4

Landmark Type	Peak					
	Before registration			After registration		
Parameter	Avg area	std	FRE sum	Avg area	std	FRE sum
Bare hand	53.52	3.01	35329.9	60.34	4.16	194.53
1 rubber band	58.62	5.44	31354.5	62.13	4.37	48.88
2 rubber bands	46.34	6.14	103902	48.35	5.11	120.04
3 rubber bands	25.37	4.27	102859	42.22	5.15	258.11

Table A.3: Landmark based registration by 5 palm joint based on Group2

Landmark Type	Peak					
	Before registration			After registration		
Parameter	Avg area	std	FRE sum	Avg area	std	FRE sum
Bare hand	53.52	3.01	4459.24	40.55	2.75	128.4
1 rubber band	58.62	5.44	56633.6	41.48	2.89	134.5
2 rubber bands	46.34	6.14	106501	32.33	3.33	146.95
3 rubber bands	25.37	4.27	43033.7	28.62	3.52	32.59

Table A.4: Landmark based registration by 5 palm joint based on Group3

Parameter	Valley					
	Before registration			After registration		
	Avg area	std	FRE sum	Avg area	std	FRE sum
Bare hand	4	1.25	79489.24	4.76	1.44	118.59
1 rubber band	6.99	1.52	20085.37	6.49	1.59	60.9
2 rubber bands	7.03	2.61	86174.3	5.64	2.58	816.07
3 rubber bands	4.92	1.7	82105.11	6.64	2.67	52.74

Table A.5: Landmark based registration by 5 palm joint based on Group4

Parameter	Valley					
	Before registration			After registration		
	Avg area	std	FRE sum	Avg area	std	FRE sum
Bare hand	4	1.25	70303.91	6.8	2.04	91.98
1 rubber band	6.99	1.52	5280.12	9.24	2.25	64.1
2 rubber bands	7.03	2.61	42217.63	8.01	3.7	1206.75
3 rubber bands	4.92	1.7	49916.95	9.47	3.82	33.65

Table A.6: Four group hand speed analysis

Landmark Type	Meanspeed	Peak				Valley			
		Speed (count/min)	Real	Original	noise	Speed (count/min)	Real	Original	noise
Group1 bare band	38.11	38.41	17	38	21	37.8	17	37	20
Group2 single rubber band	34.99	35.27	17	35	18	34.71	17	34	17
Group3 two rubber band	31.96	31.95	16	47	31	31.96	15	46	31
Group4 three rubber band	33.84	37.09	18	54	36	30.59	15	53	38

Technical
University of
Denmark

Bygning 324
2800 Kgs. Lyngby
Tlf. 4525 1700