Calibrating Video Capture Systems To Aid Automated Analysis and Expert Rating of Human Movement Performance

Sai Krishna Yeshala

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Computer Science and Applications

Aisling Kelliher, Chair
Thanassis Rikakis, Co-Chair
Denis Gracanin
Margaret Ellis

May 10, 2022
Blacksburg, Virginia

Keywords: Video capture systems, Multi-camera calibration, Activity space calibration.

Calibrating Video Capture Systems To Aid Automated Analysis and Expert Rating of Human Movement Performance

Sai Krishna Yeshala

(ABSTRACT)

We propose a methodology for calibrating the activity space and the cameras involved in video capture systems for upper extremity stroke rehabilitation. We discuss an in-home stroke rehabilitation system called Semi-Automated Rehabilitation At Home System (SARAH) and a clinic-based system called Action Research Arm Test (ARAT) developed by the Interactive Neuro-Rehabilitation Lab (INR) at Virginia Tech. We propose a calibration workflow for achieving invariant video capture across multiple therapy sessions. This ensures that the captured data is less noisy. In addition, there is prior knowledge of the captured activity space and patient location in the video frames provided to the Computer Vision algorithms analyzing the captured data. Such a standardized calibration approach improved machine learning analysis of patient movements and a higher rate of agreement across multiple therapists regarding the captured patient performance. We further propose a Multi-Camera Calibration approach to perform stereo camera calibration in SARAH and ARAT capture systems to help perform a 3D reconstruction of the activity space from 2D videos. The importance of the proposed activity space and camera calibration workflows, including new research paths opened as a result of our approach, are discussed in this thesis.

Calibrating Video Capture Systems To Aid Automated Analysis and Expert Rating of Human Movement Performance

Sai Krishna Yeshala

(GENERAL AUDIENCE ABSTRACT)

In this thesis, I describe the workflows I developed to perform calibration of stroke rehabilitation activity spaces, including the calibration of cameras involved in video capture systems for analyzing patient movements in stroke rehabilitation practices. The proposed workflows are designed to facilitate convenient user involvement in calibrating the video capture systems to provide invariant and consistent video captures, including the extraction of fine-grain information utilizing camera calibration results, to the therapists and computer vision-based automated systems for improved analysis of patient performance in stroke rehabilitation practices. The importance of human-in-the-loop systems, including future research paths to strengthen the symbiotic relationship between humans and Artificial Intelligence systems in stroke rehabilitation practices, is discussed. The quantitative and qualitative results generated from the workshops conducted to test and evaluate the calibration workflows align with the stakeholder's needs in stroke rehabilitation systems.

Dedication

To my family, mentors, and well wishers, thank you for supporting me in this incredible journey.

Acknowledgments

I would like to thank my advisors Dr. Rikakis and Dr. Kelliher for giving me the opportunity to work on challenging and interesting problems in the field of stroke rehabilitation. Their constant guidance and support made this work possible. I would also like to thank my amazing team members at INR for guiding and mentoring me throughout this journey. Finally, I thank my parents for supporting me throughout my education.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Stroke is one of the most commonly occurring neurological disorders as observed by the CDC [3]. There has been a significant amount of ongoing research in the field of stroke rehabilitation practices. Leveraging the latest advancements in Artificial Intelligence and Computer Vision, stroke rehabilitation systems that use cameras and wearable sensors have been on the rise in the past decade.

My work focuses on human-in-the-loop architecture employed to assess upper extremity stroke rehabilitation practices using video camera-based capture systems. The captured video data is utilized to provide insights to the therapist post-rehabilitation sessions to analyze and provide reasoning for their ratings of patient exercises and perform computer vision analysis of the patient movements for automated assessment of the exercises performed. My work was conducted as a part of the Interactive Neuro-Rehabilitation Lab (INR) at Virginia Tech. The INR team has developed an approach to collect expert therapist ratings and utilize them to provide insights into the Machine Learning models for a more informed automated analysis of patient activity performance and relating patient movement quality to functionality [4]. I propose a novel activity space and multi-camera calibration approach to set up, calibrate, and extract fine-grain information from the captured videos to improve the computer vision based automated analysis of stroke patient movements. As a result of this work, expert therapists can take advantage of the reconstructed granular information of stroke patient movement and transfer their expert knowledge to the machine learning models

for improved analysis of patient movements in stroke rehabilitation practices.

1.1 Background and Motivation

Telehealth and telemedicine unsurprisingly experienced rapidly accelerated growth in 2020. In the US alone, Medicare primary care visits via telehealth rose from 1% in February to almost 50% of visits in April [7]. But even pre-pandemic, the field of telehealth was already greatly rising in prominence and prevalence. With the loosening of insurance restrictions in the US concerning routine medical care, the field is expected to continue to develop at speed. As the pace of global population aging also accelerates, there is an increasing need for widespread physical and occupational rehabilitation services for common age-related debilitating illnesses such as arthritis, stroke, and osteoporosis. [26, 27]. Effective rehabilitation requires intensive training and the ability to adapt the training program based on patient progress, and therapeutic judgment [19]. Active participation by the patient is also critical for improving self-efficacy, and program adherence [20]. However, intensive and adaptive rehabilitation is challenging to administer in an accessible and affordable way as it necessitates frequent trips to the clinic (usually reliant on a caregiver) and significant face-to-face time with rehabilitation experts [22]. Ultimately, the biggest problem here is that there is a significant lack of available rehabilitation experts and therapists to cover the needs of a geographically dispersed and aging population [35].

Capturing videos of stroke patients performing rehabilitation practices in a meaningful way is a crucial aspect of developing stroke rehabilitation systems. There are great benefits of utilizing such systems for providing a way for therapists to track patient performance over time, analyze the recorded videos to obtain deeper insights into patient's rehabilitation progress, and allow automated computer vision based systems to analyze patient movement

in the captured videos.

Stroke rehabilitation systems can be implemented in patient homes as well as in hospitals or clinical settings. Standardized exercises and tests are regularly administered by occupational and clinical therapists for stroke rehabilitation purposes. Creating digital versions of such systems with the goal of automated analysis of patient movements requires the cooperation of the therapists as well as the automated systems. A cyber-human architecture that leverages expert knowledge for improved automation and therapist understanding of rehabilitation progress can be extremely beneficial to the overall performance of stroke rehabilitation systems.

1.2 Research Problems

The development of stroke rehabilitation video capture systems comes with specific challenges in camera and activity space standardization, camera calibration, and extraction of meaningful information from the captured videos for better overall assessment of patient performance. There are two main research questions that I aim to answer in this thesis: How do we improve the quality of video capture systems to provide less noisy data for Computer Vision analysis of human movement? and how do we offer better user understanding of specific domains by providing more granular data utilizing machine learning advancements? The two research problems are discussed in the context of the work I conducted for cyber-human stroke rehabilitation systems developed as a part of the INR lab.

1.3 Contributions

The INR lab has developed two rehabilitation systems utilizing low-cost cameras to capture patient movements. The in-home rehabilitation system called Semi-Automated Rehabilitation At Home (SARAH) is designed for remote therapy for upper extremity stroke recovery. The Action Research Arm Test (ARAT) system is developed for upper extremity stroke rehabilitation in clinics.

1.3.1 Semi-Automated Rehabilitation At Home System (SARAH)

The SARAH system utilizes low-cost video cameras, a set of therapy objects, an activity mat, and an automated video analysis engine, including interactive capture and rating interfaces for capturing and extracting expert therapist ratings to analyze patient movements. This system is intended to be deployed in patient homes and allows the therapists to administer highly effective stroke rehabilitation practices.

1.3.2 Action Research Arm Test (ARAT)

The Action Research Arm Test (ARAT) is a standardized assessment that focuses on measuring changes in the upper limb functionality of stroke survivors. This tool is a 19-item measurement divided into four sub-tests (grasp, grip, pinch, and gross arm movement) which assesses upper limb functioning using observational methods [25]. Performance on each item is rated on a 4-point ordinal scale ranging from: “0” Can perform no part of the test to “3” performs test typically. In addition, the ARAT is rated as having excellent test-retest reliability and inter-rater reliability, moderate burden overall to complete, and moderate construct validity and responsiveness [32].

The ARAT system developed by the INR team is intended to be installed in clinical settings. This raises new challenges, such as requiring a quicker installation process, considering that clinical settings are fast-paced environments. Moreover, the exercises performed in the ARAT system require more granular information capture of the interaction between the patient's impaired hand and objects. Extracting such information from the ARAT capture system's video captures requires a three-dimensional reconstruction of the patient hand and object interaction. This information is utilized by both the therapists and the computer vision algorithms to perform a more informed assessment of patient performance in the rehabilitation exercises.

The current versions of both SARAH and ARAT systems involve a capture interface used for controlling video captures of rehabilitation exercises and a Rating/ Video Annotation tool for providing the captures to the therapists and allowing them to review and rate the movements following a rubric developed by the INR team [34] for activities involved in each system.

1.3.3 Capture Interfaces

The capture interfaces for SARAH and ARAT are developed to allow the respective users to start and stop the exercises for the duration of which videos are captured [9]. The SARAH capture interface involves an activity space calibration approach to place the cameras in space and computationally verify the patient and activity space locations. The ARAT capture interface involves a camera-calibration process that extracts the extrinsic and intrinsic parameters of the camera. The calibration steps are performed before beginning the assessments in both systems.

1.3.4 Rating/ Video annotation tools

As mentioned earlier, the SARAH and ARAT systems involve a Video annotation tool developed by the User Experience (UX) members of the INR team. This tool is utilized to provide the captured data to the therapist. It allows them to offer reasoning for their rating of the patient's performance by walking through a rubric developed by the INR team [34]. The ratings collected from the therapists through the Video annotation tool is supplied to the computer vision based engine for improved analysis of patient movements in the captured videos.

1.4 Thesis Outline

This thesis details my research efforts in calibrating capture systems to analyze the movements of upper extremity stroke patients during rehabilitation. Chapter 3 outlines my proposed approach and workflows to calibrate the activity space and the cameras of in-home and clinic-based capture systems developed at INR including a method to reconstruct the hand-object relationship of patients' hand with the therapy objects. The development of the intuitive calibration workflows is discussed in the context of the capture systems developed by the INR team at Virginia Tech. The results achieved show an improvement in automated analysis of patient movement and therapist ratings of the captured videos. In addition, the camera calibration approach opens up new paths for research in Computer Vision based 3D reconstruction techniques to extract and analyze granular information from the captured videos.

Figure 1.1 depicts a high-level flow diagram of the general architecture followed by the INR team for their stroke rehabilitation capture and analysis systems. The capture system

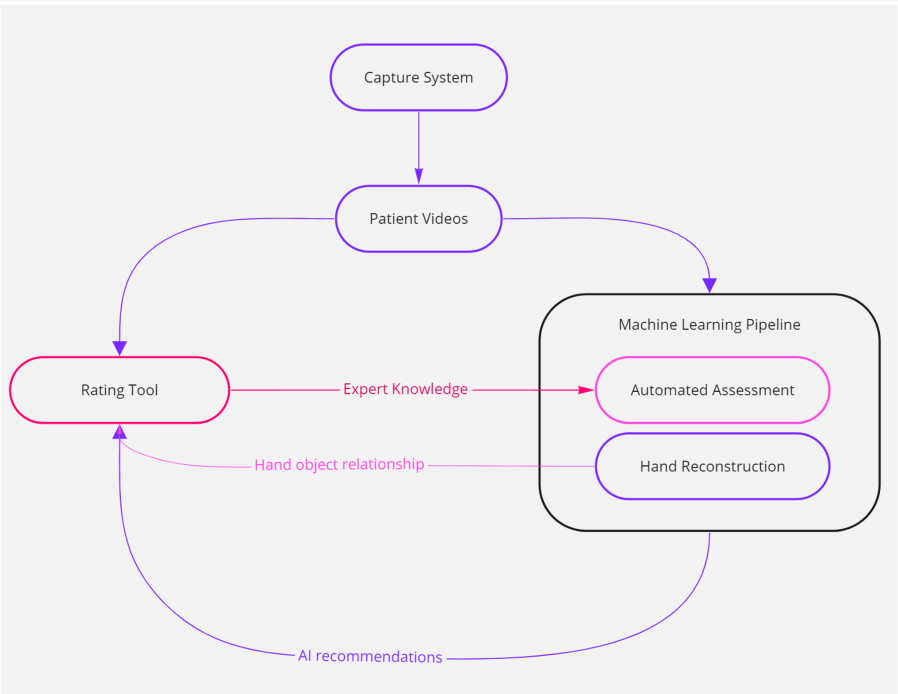

Figure 1.1: A high level flow diagram depicting the architecture used for Stroke Rehabilitation Systems at INR.

records patients performing specific exercises. The captured videos are utilized by both the Machine Learning pipeline and the rating tool, which allows the therapists to look back at the captured videos and provide their ratings based on their expertise. There is a transfer of knowledge between the machine learning pipeline and the rating tool where the experts help guide machine learning with their domain knowledge. Machine learning helps experts by providing granular information from the captured videos that they could not observe with the naked eye. I include another highlighted version of this diagram in the conclusions chapter to indicate my contributions and how they enhance the functioning of this system architecture.

Ultimately, this thesis depicts my contributions to the stroke rehabilitation capture systems developed at INR. My contributions to the INR team facilitate better Machine Learning

outcomes due to the reduction of noisy video data and improved quality of therapist ratings of the captured videos achieved by providing insights into intricate details of patient movements observed in the videos. The ultimate incentive for the therapists to rate the captured videos is the recommendations that the Machine Learning algorithms will generate in the future.

Chapter 2

Review of Literature

Our approach builds upon extensive related work on developing telerehabilitation systems and semi-automated rehabilitation systems for accurately capturing patient movement quality by incorporating intuitive user interfaces. In addition, there have been various computational techniques and the latest technologies applied to develop non-obtrusive stroke rehabilitation systems. Our approach involves an array of low-cost video cameras that capture patient movements during rehabilitation practices and perform computational evaluation and analysis of the captured videos.

2.1 Stroke Rehabilitation

Stroke is one of the most common neurological disorders around the world and is one of the leading causes in the United States. According to the CDC, 795,000 people experience a stroke in the United States [3]. There is excellent support for extended rehabilitation therapy in helping with recovery post-stroke [23, 33]. Standard stroke rehabilitation practices involve patients performing various movements to grasp and move objects in space in the presence of a therapist. The therapist rates the patients' exercises and generates an overall score per session. This information is tracked to analyze patients' recovery over time.

2.2 Stroke Rehabilitation Systems

Various systems leverage advancements in technology to develop rehabilitation systems for stroke. These systems come as Virtual Reality, Computer Vision, and Robot-Assisted systems that collect patient information to provide insights into the performance and improvements of stroke patients over time [23].

Virtual Reality based Stroke rehabilitation systems is not successful in providing an appropriate environment for stroke patients to perform stroke recovery effectively [23]. Various gaming-based rehabilitation systems have proven to be effective. However, they require much assistance from therapists [10, 28]. The goal of capturing the patient data is to analyze the collected data using computational methods to gain deeper insights into patients' performance and recovery over time. Several motion sensor-based systems can accurately capture patient movements, such as gloves and motion capture suits. Such systems are costly, complex, and obtrusive, which affects patient engagement [29, 30, 31].

2.3 Video Capture Systems for Stroke Assessments

Stroke rehabilitation systems involving video cameras must be calibrated to obtain fine-grain details of patient movements. In addition, the activity space where the patient performs the rehabilitation exercises is also required to be set up to facilitate better analysis of the captured videos. This section looks into some of the existing techniques to achieve activity space and camera calibration of video capture systems.

2.3.1 Monocular and Stereo-Camera Calibration

Calibration of a camera is the process of obtaining extrinsic and intrinsic parameters of the camera [21]. The extrinsic parameters, also known as external parameters, of a camera describe a transformation between the camera and its external world. Intrinsic parameters of a camera are specific to every camera, and they represent the optical center and focal length of the camera. The world points are transformed to camera coordinates using extrinsic parameters, and the camera coordinates are mapped into the image plane using intrinsic parameters [21].

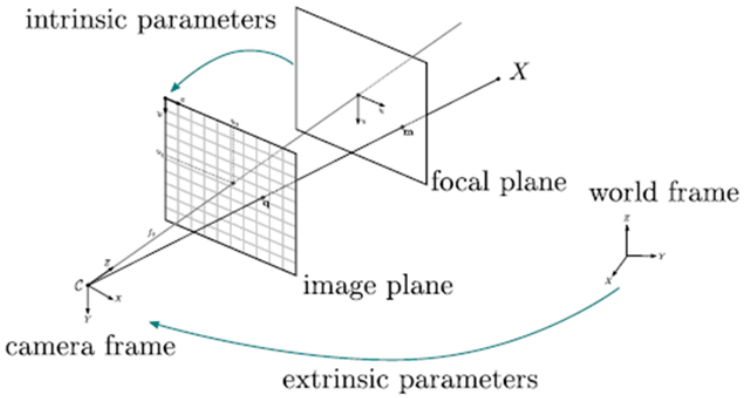

Figure 2.1: Use of intrinsic and extrinsic parameters of cameras [1].

A stereo camera setup involves two different cameras to capture the environment [12]. One of the cameras can be assumed to be the primary camera and the other as the secondary camera. The calibration of stereo cameras involves two steps: extracting camera intrinsic and extrinsic parameters of both cameras and computing the rotation and translation of the secondary camera to the first camera [12].

A calibration grid is a tool utilized to allow the detection of various points on the grid as identified by the cameras to perform the required computation to extract intrinsic and extrinsic parameters of the cameras [14]. In most cases, a checkerboard print is used as

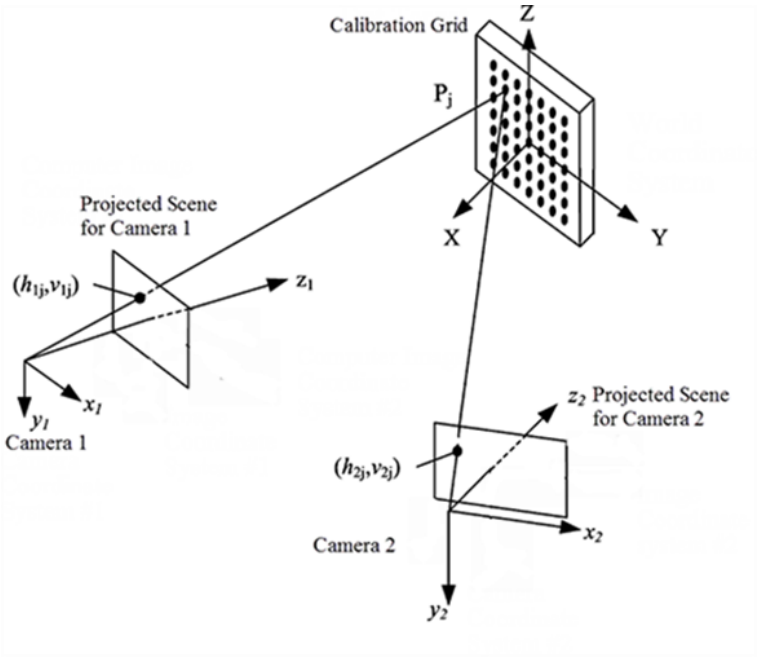

Figure 2.2: A stereo camera setup and calibration using a calibration grid [1].

the calibration grid, where the corners of the checkerboard are detected by the calibration applications which are utilized to perform calibration.

2.3.2 Calibrating activity space for invariance across captures

To maintain consistency in capture systems involving video cameras, it is ideal for standardizing the cameras relative to the activity space or the target that is being captured. We have proposed a workflow to standardize the camera placement concerning the rehabilitation activity space (discussed in Chapter 3) to achieve invariance across video captures of multiple patients. The goal is to ensure that the captured activity space and location of the patient in the video frame are consistent across multiple captures. This provides the Computer Vision algorithms with prior knowledge of activity space and patient locations in the video frames for a smoother analysis. The ultimate video data will be less noisy.

The camera placement problem has been studied previously [13, 15] while prior work in this area suggests the productive use of motion sensors and omnidirectional cameras [13]. We propose that all the requirements be met using two low-cost video cameras and fiducial markers in combination with a calibration interface. We utilize fiducial markers to track the markers in video frames to allow the users to use the tracked markers as a reference while placing the cameras in space. According to [6] one of the significant advantages of using fiducial markers for calibration is the gain in accuracy in feature localization and correspondence. The use of proper fiducial markers ensures precise localization by sub-pixel accurate line intersection [6]. The paper also asserts that the fiducial marker-based targets allow us to process only reliably located points. Hence, the users do not have to take care of the visibility of the calibration target. This means that the fiducial markers are less sensitive to partial occlusions and the libraries have excellent accuracy at identifying the markers. One of the main reasons we chose ArUco fiducial markers was also because of the high accuracy at detecting the markers even in conditions of uneven illumination [24].

Chapter 3

Problem Description

The major goal of video capture systems for analyzing patient performance in rehabilitation exercises is to capture videos that facilitate improved post-analysis of the captured videos by therapists and automated video analysis systems. Furthermore, extracting granular information from the captured videos can help provide a better overall analysis of patient performance in rehabilitation practices. As discussed in Chapter 2, video capture systems involving an array of video cameras must be calibrated before utilizing them to capture human movements. The process should include calibration of activity space and the cameras to achieve a better overall analysis of the captured videos.

Calibrating the activity space being captured requires user involvement with intuitive and thorough guidance of the setup process. The instructions can be provided through interactive interfaces with real-time feedback to calibrate the captured activity space. In stroke rehabilitation systems, we have established that the activity space and the cameras involved must be calibrated for better analysis of patient movement. How can we achieve an intuitive and time-efficient activity space and camera calibration processes involving therapists to calibrate the video capture systems appropriately? And how we can extract meaningful granular information from the captured videos to help the therapists and the machine learning pipelines are the main problems we focus on.

Rehabilitation for stroke can be administered broadly in two settings: In-home rehabilitation, which involves occupational therapists visiting patients' homes to administer stroke

rehabilitation exercises, and stroke rehabilitation in hospitals or clinical settings. Both areas have their own challenges in terms of deploying stroke rehabilitation systems.

3.1 In-Home Stroke Rehabilitation Systems

Therapists often visit patients' homes to administer rehabilitation practices and keep track of patient progress in stroke recovery. To deploy video capture systems to capture patient movements requires the system to be non-obtrusive, low-cost, and efficient use of space in patient homes. Moreover, the process of setting up the system has to be intuitive for the therapists with proper guidance to ensure the accurate placement of cameras and the activity space. The interfaces utilized to facilitate the setup process are also required to be simple and time-efficient.

3.2 Clinic Based Stroke Rehabilitation Systems

Clinical environments have various other challenges in terms of setting up the video capture systems in a dedicated space. Clinical therapists administering stroke rehabilitation practices require a non-obtrusive, cost-efficient, and fast-paced setup process of the capture systems compared to in-home settings. Moreover, the hospitals have specific requirements in dedicated clinical spaces to ensure patient safety and convenience. Setting up video capture systems should not cost the therapists more than five minutes pre-assessment of stroke practices.

Chapter 4

Proposed Approach

This chapter discusses the workflows and methods used to solve the problems described in Chapter 3. Section 4.2 discusses the activity space calibration approach for setting up the activity space and ensuring invariance in captured videos in the in-home-based SARAH system. Section 4.3 details the activity space calibration pre-incorporated in the capture system design for ARAT. Section 4.4 discusses the multi-camera calibration approach in the context of the ARAT capture system developed by the INR team with initial experiments conducted on the SARAH system. This section also details how we tackled some of the challenges encountered in the ARAT system due to a clinic-based setup involved. Finally, in section 4.5, An approach to reconstructing the hand object relationship from the captured videos in three dimensions and a way to deliver that information to the therapists is discussed.

4.1 Activity Space Calibration

I describe my approach to developing a hybrid workflow process to achieve invariance in captured videos of stroke rehabilitation training across multiple capture sessions. The approach builds upon prior computer vision based work [4, 8]. The importance of standardizing the activity space and the capture system relative to the activity space was realized through the results achieved by the automated analysis of upper-extremity stroke rehabilitation performed on the data collected during a pilot study conducted by the INR team at the Emory

Rehabilitation Hospital [4]. The video data captured in this study was noisy. Considering that there was no standardization enforced in camera placement, every therapist session had varying locations of activity space and patients in the captured video frames.

4.2 Activity Space Calibration for SARAH

We considered the in-home rehabilitation system developed by the INR team over a decade - The Semi-Automated Rehabilitation At Home (SARAH) system for testing and evaluating our approach. This work integrated the tools developed by the industrial design and User Interface teams in the lab and constitutes crucial steps before beginning a rehabilitation session and while installing the SARAH system. The results discussed in Chapter 5 indicate the advantages of standardizing camera placement and relative location of activity space in captured videos. In addition, there was a considerable improvement in the standard deviation of segment accuracy [5] produced by the automated patient movement analysis engine developed previously by the computer vision team at INR.

The results showed a great room for improvement in the design of automated analysis of stroke patient videos. A physical modification of the capture system was required to facilitate this.

Considering the involvement of healthcare providers in installing the cyber-human systems developed at INR, it was crucial to loop in the feedback of expert clinicians and physiotherapists throughout the design process. The goal was to create a simple and usable calibration process that allows the therapists to set up the system in a reasonable amount of time. We conducted a pilot study in Roanoke, Virginia, with six expert occupational therapists tasked with setting up and calibrating the SARAH system. We engaged the expert therapists in a brief debrief session with the development team, where they answered survey questions re-

garding the calibration process. This allowed us to evaluate the hybrid calibration workflow process from a Human-Computer Interaction perspective.

The capture tool was designed and developed by the UX members of the team and described in [9]. I collaborated with them to integrate the calibration workflow into the capture tool. As mentioned earlier, the therapists are required to perform a set of pre-calibration steps (described in [8]) to ensure that the equipment needed for the capture session is available. The three major steps involved in setting up the capture system are as follows:

Step 1: The therapists are provided with real-time instructions on placing the camera at the accurate position in space. They first start with the side camera, and once the requirements are met, they are instructed to repeat the process with the front camera.

Step 2: Once the cameras are set up in space, the therapist is instructed to bring the patient to the activity space and have them seated on a chair. The capture tool performs the patient detection step, and necessary instructions are provided to meet the patient positioning requirements.

Step 3: Finally, once the patient is in the required position relative to the captured frame and the activity space, the therapists are moved to a loading screen that performs automatic circle detection on the side and front camera frames to segment the activity space.

Our movement capture workflow for setting up, calibrating, and capturing videos combines screen-based interfaces for the patient and the therapist. The goal of our system is to support remote monitoring of patient activities by the therapist as part of a regular therapy protocol (typically two visits in person by the therapist per week). On non-visit days, the patient will complete their therapy activities using the system with summary information sent to the therapist. A vital issue is minimizing the technical burden on both the therapist and patient. Our workflow process is designed to assist the therapist with initial system setup

and calibration, while patient-specific software ensures the efficient delivery of the therapy protocol. We focus here primarily on the therapist interface for assisting with system setup, and calibration [8, 9].

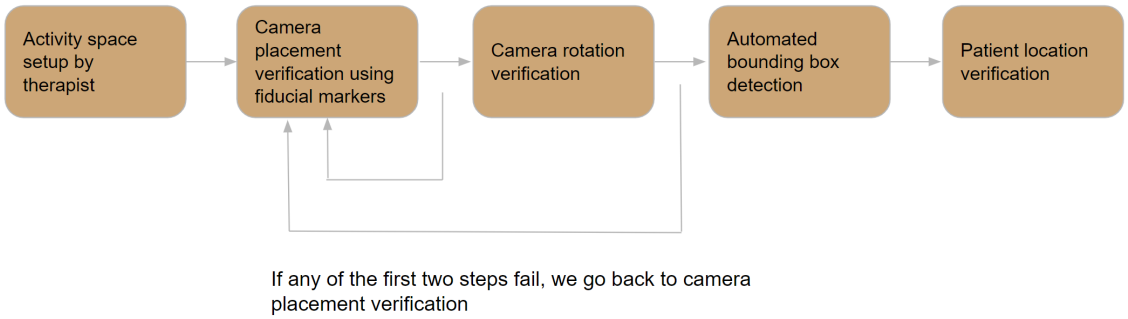

Figure 4.1: Activity space calibration workflow.

The SARAH system consists of two Logitech Brio cameras (1080 x 720 pixels resolution) mounted on tripods, a table, a chair, a custom-designed tabletop mat to cater both to the needs of the computer vision team and the patient, a set of therapy objects, and an intuitive tablet interface for therapists and patients to facilitate the rehabilitation exercises [8].

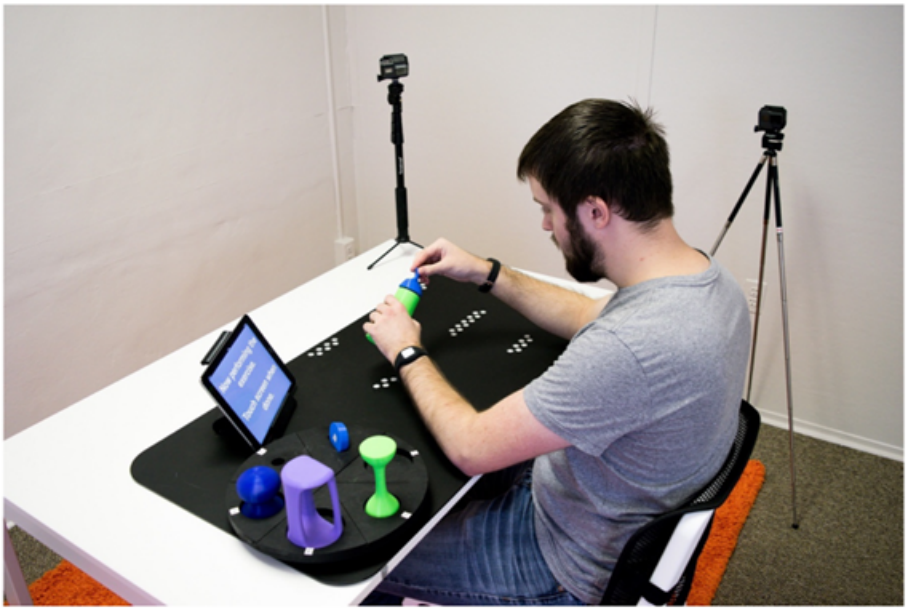

Figure 4.2: The custom objects, cameras, interface, and activity mat

Developing a movement capture workflow process for diverse stakeholders requires committed collaborations between all involved parties. Our team includes UI/UX designers, industrial designers, occupational therapists, computer vision/machine learning experts, and computer scientists. The success of our approach required considerable trade-offs and handshakes between physical and virtual components in our system, in conjunction with the need to create a straightforward interaction experience for the ultimate users of our system. For example, the unique design of the rehabilitation objects [18] is tailored to maximize the patient's perception of the possible affordances [11] of the artifacts while also supporting the computer vision detection algorithm through the different sizes, shapes, and colors of the objects. Figure 4.2 depicts the seven modular rehabilitation objects in our system that can be grasped, manipulated, and combined in various ways to support therapeutic activities related to activities of daily living.

Similarly, the screen-printed staging mat used in our tabletop system (seen in Figure 4.4) assists the patient in determining where to place the objects, while the markings also function to help automatically calibrate the correct location of the patient and the staging mat itself.

Intuitively guided instructions are provided to the therapists setting up the system through the capture interface. In addition, pre-calibration requirements are visually depicted on the interface for the therapists to prepare for the calibration process.

Figure 4.3: SARAH objects.

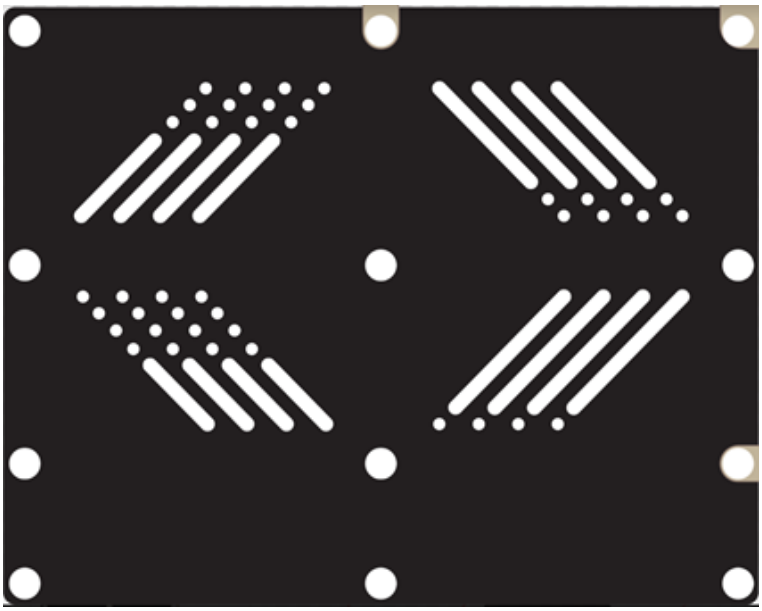

Figure 4.4: SARAH screen printed tabletop activity mat.

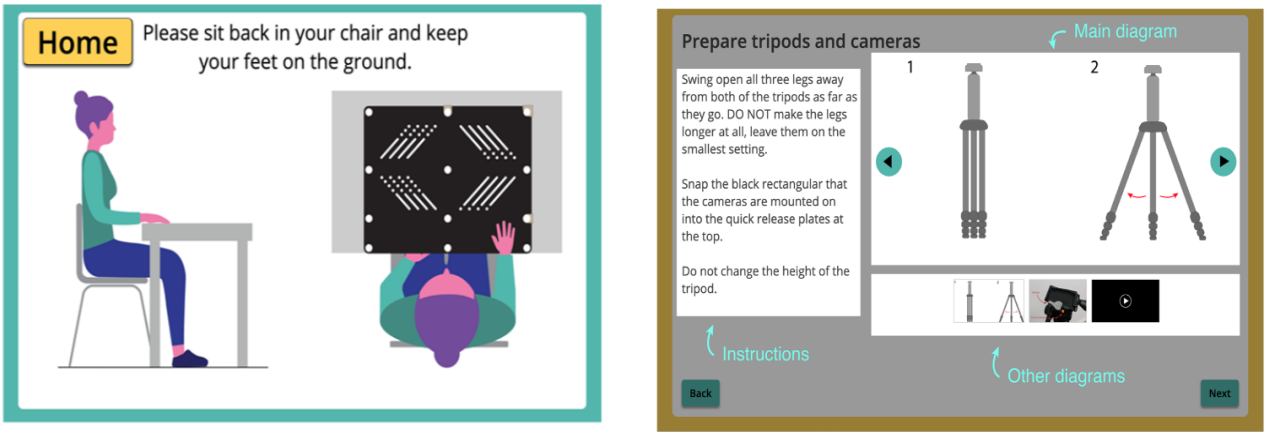

Figure 4.5: Examples of pre-calibration steps [9].

The computational verification of the calibration process aims to verify two primary aspects: Tripod placement at standardized locations relative to the tabletop activity mat; and automated patient detection to ensure the right placement of patient in the image frame and automated segmentation of activity space.

The two steps of the verification process are further broken down into two computational sub-

steps. The primary goal of the calibration process is to incorporate specific standardization processes to ensure high-quality capture of consistent data provided to the therapist and the machine learning algorithms. The setup and calibration process also needs to be intuitive and efficient for patients and therapists to use in a reasonable amount of time.

4.2.1 Standardized Camera Placement

The first step towards achieving invariance in video captures is to standardize the placement of cameras relative to the activity space. This ensures that the position of the activity space (tabletop mat), the patient, and the objects relative to the video frame is constant across multiple captures and multiple SARAH sessions. This provides a strong prior to the computer vision team by narrowing down the search space of the activity mat and patient location in the video frame, which is analyzed during the automated assessment of patient movement quality and functionality [4]. After thorough experimentation with various camera locations and tripod heights, the camera tripod locations are standardized to the following measurements: The side camera is 3.5 feet away from the edge of the mat, and the front camera is 2.5 feet from the edge of the mat.

Through pilot testing we performed, we realized that there is a great deal of precision required to place the camera-mounted tripods at their respective standardized locations. A computational approach was needed to ensure that the requirements to achieve invariance across captures are met for a smoother analysis of the captured video data.

Step 1 - Locating tripods in space

We utilized ArUco fiducial markers [24] which are detected in the image frames, and real-time instructions are generated to guide the therapists in finding the optimal location of cameras in space. The ArUco markers are the fiducial markers that are heavily used for various purposes such as Augmented Reality, Virtual Reality, and Robotics [24]. They are also used mainly to perform pose estimation in many different kinds of applications. The original library of ArUco fiducial markers is developed in C++. There is detailed documentation for this library that is very straightforward and easy to understand. The ArUco markers can detect the markers with great accuracy, considering that the contrast difference between the white and the black blobs is exceptionally high. The detection algorithms mainly look for the corners of the marker and return the corners and the ids of the marker when detected. We took advantage of this particular functionality of the ArUco markers and used the ArUco library as one of the key elements in the calibration code.

The first step of the calibration process involved the camera setup that provides a view of the patient from the right side of the table. The goal of this camera is to capture the therapy session from an angle that allows the computer vision team to perform the best possible analysis of the patient's hand movements and the movements of the objects involved in the therapy session. The manual calibration performed at the INR lab before developing the calibration workflow acted as the reference video input we needed to achieve from the side camera. The first step of the process was to ensure that the camera was 3.5 feet away from the edge of the mat that was used for the therapy sessions. This was achieved by using the following formula to compute the distance:

F = (P x D) / W

F - Perceived focal length of the camera in pixels

P - Apparent width of the marker in pixels

W - Width of the marker in inches.

D - Distance of the marker from the camera in inches.

The width of the marker in pixels is measured by placing the 1-inch marker three feet away from the camera and calculating the reference focal length value, which for the SARAH system is 756 pixels. This reference value is then used to calculate the unknown variable D. The width of the marker is 1 inch, and the width of the marker in pixels is calculated when the program detects the marker. This approach gave an accurate measurement of the distance of the marker from the camera.

The second part of the calibration process is to align the camera such that the first ArUco marker is aligned at the center of the frame along the X-axis and slightly below the center on the Y-axis. This ensures that the camera successfully captures enough visual information of the activity space of the mat for further analysis by the computer vision team. The therapist is thoroughly guided during this process with both visual and textual display of instructions to ensure that the camera is being moved as required. Instructions such as “move the camera to the left”, “move the camera closer to the table”, etc., are used.

Step 2 - Verification of mat placement

The second ArUco marker on the mat is placed at the top right corner of the mat. This marker is detected by performing a triangulation process to ensure that the mat is not crooked and is facing the camera directly. The angle is calculated between the reference line perpendicular to the Y-axis going through the center of the frame and the line from the

center of marker1 to the center of marker2. This angle has to be validated to be around 90 degrees to pass the second step of the calibration process. A leeway of 2 degrees higher and lower than 90 degrees is allowed to ensure a smoother verification of the mat placement. If the angle calculated crosses the upper or lower boundaries, the therapist is instructed to check the positioning of the mat on the table. Both step1 and step2 are constantly reviewed by the program performing the calibration, and any movements that move the camera away from the ideal spots instruct the therapist to fix the respective step again.

The two substeps for standardized tripod placement are repeated for the front camera with a standardized distance of 2.5 feet away from the edge of the mat.

4.2.2 Patient detection and automated segmentation of activity space

One of the crucial steps of automated assessment of upper-extremity stroke activities in the SARAH system is the body keypoint detection performed by OpenPose [4]. OpenPose is a state-of-the-art pose detection library for detecting body and face key points in images of humans. A standardized position of the patient in the video frames facilitates a smoother analysis of the movement quality and functionality of the patients. Therefore, we designed this step of the calibration process to ensure that the patient is positioned within the bounds of the patient box in the frames captured. Then, we analyze the activity space to perform automated segmentation using a circle detection algorithm I developed.

Patient location in the video frame

After locating the cameras in space and verifying the placement of the activity space in relation to the frames captured, the therapist brings the patient to the activity space and instructs them to sit on the chair. Utilizing Microsoft Azure's face detection API, we detect the bounding box of the patient's face and verify the bounding box locations against a standardized reference bounding box that we defined. We provide real-time feedback on whether the patient's location in the frame satisfies our requirements.

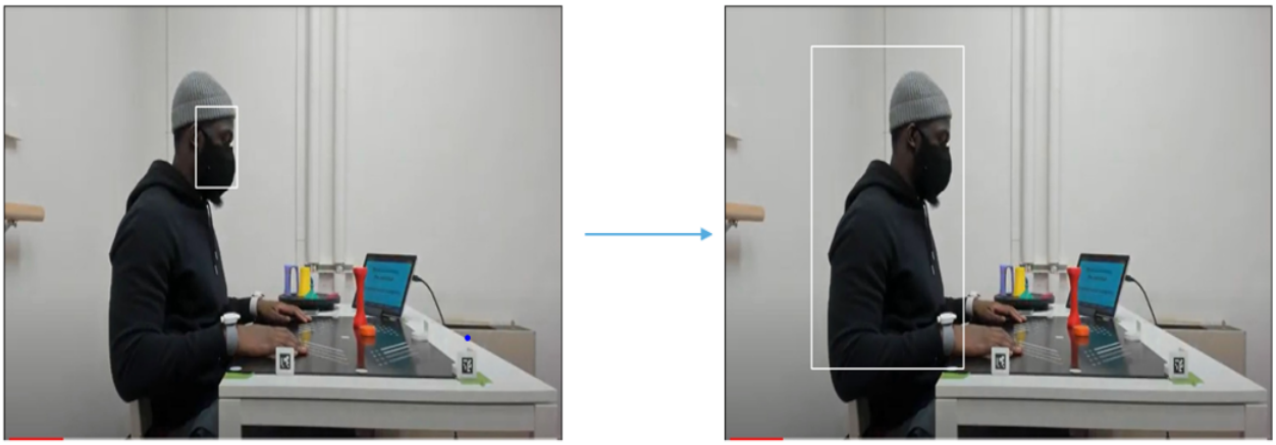

Figure 4.6: Patient bounding box detection.

Automated segmentation of activity space

The updated mat design (figure 4.7) ensures that we can automatically segment the activity space (area of the mat) into five segments - A, B, C, D, and a box indicating the hand of the patient at rest position. The computer vision team identifies the location of the object in the video frame at a given time in the captured activity video to analyze the completion and accuracy of the exercise performed by the patient. Providing an automated approach to divide the activity space into the five segments given a video frame from the calibrated set of side and front cameras is crucial for the analysis of exercise accuracy and completion.

Therefore, to achieve this, we screen-printed four rows of circles depicted on the activity mat (Figure 4.7). Each segment is effectively formed by using the circles as the corner vertices of each segment in the frame. The distance between the rows is designed to account for the foreshortening issue we faced during testing. When the rows are equidistant, the farthest row and the row in front of the most distant row do not accurately represent the area of the segments that lie between those two rows. By increasing the distance between the two farthest rows, we were able to capture the area of the respective segments accurately.

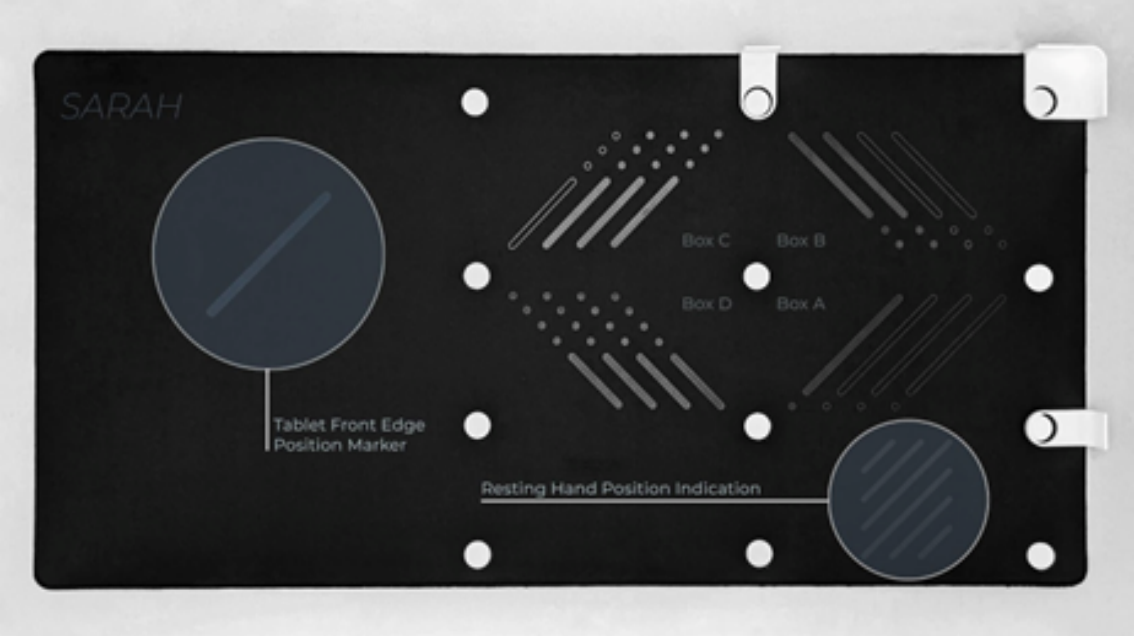

Figure 4.7: Depiction of four segments on mat area including the rest position of the hand.

We utilize a connected component detection algorithm to detect the circles on the mat after applying a threshold of "cv2.threshold(image, 127, 255, cv2.THRESH_BINARY)" (OpenCV function for greyscale thresholding of images). Considering the known location of the mat in the frame as a result of standardized camera placement relative to the activity space, we are able to crop out each row of the mat and apply the connected component algorithm. This helps us reduce noise in the image and detect the circles with higher accuracy on both side

and front cameras.

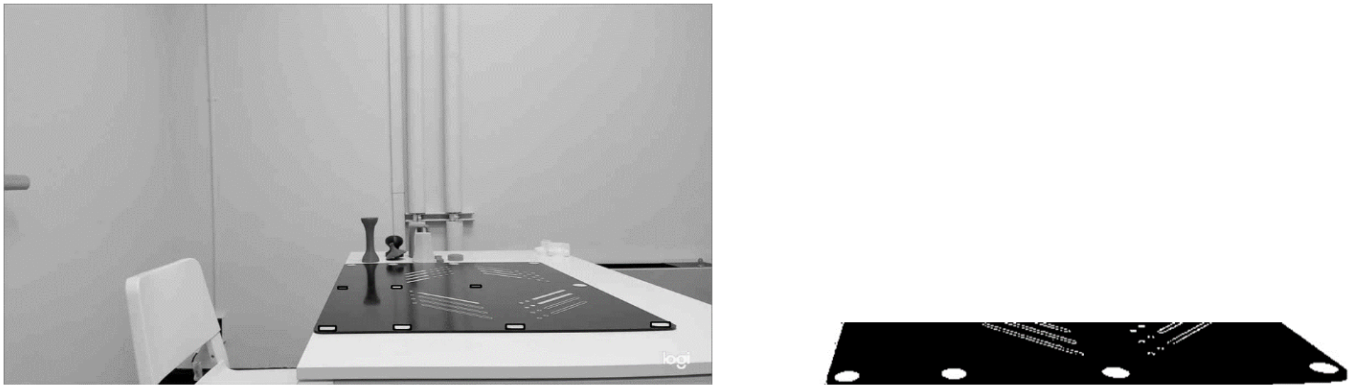

Figure 4.8: calibrated side camera setup (left) with circle detection algorithm applied on the last row (cropped image on the right).

4.3 Activity Space Calibration for ARAT

The ARAT capture system is deployed in fast paced clinical settings where the therapists cannot afford to spend time setting up the activity space and computationally verify the activity space setup as performed in the SARAH system. The INR team developed a rigged system involving three cameras and a fourth camera independent of the rigged system. The rigged system is depicted in Figure 4.9, and Section 4.2.1 describes the design of this setup. The ARAT setup leverages this design to achieve invariance in captures across multiple therapy sessions considering that the cameras are always in a fixed position with respect to the activity table. This ensures that the video captures across patients are consistent and no additional effort is required by the therapists to setup the cameras in the ARAT system. There is no requirement for computational verification as steps implemented in SARAH either. The rigged ARAT setup proved to be a crucial componenet in not only reducing the setup time spent by the therapists, but also provided the computer vision team with

consistent invariant captures of patient movements.

4.4 Multi-Camera Calibration

The importance of achieving invariance in captured video data and how it can improve the results of automated analysis of patient movement quality and task completion is established in Section 4.1. In this chapter, I discuss the workflow we developed to calibrate the cameras involved in the SARAH and the ARAT capture systems.To obtain three-dimensional information from two-dimensional videos, we require at least two cameras that are calibrated. Considering that capture systems developed at the INR lab contain two or more cameras, we had to come up with a way to calibrate multiple cameras at once to obtain the camera parameters that help with the 3D reconstruction of the activity space.

One of the major challenges we encountered was the overhead/transverse camera in the ARAT capture system as depicted in Figure 4.9. Considering that this camera is orthogonally located to both the side (sagittal left and sagittal right) cameras, the process of calibrating all three cameras at once was not straightforward. I developed a pair-wise calibration approach utilizing MATLAB's stereo camera calibrator app to solve this problem. Stereo camera calibration consists of two additional parameters called rotation and translation to define the relation between the two cameras in space [12]. The pair-wise approach performs a stereo-calibration of the transverse and sagittal left camera pair and a stereo-calibration of the transverse and sagittal right camera pair. This approach stores the stereo calibration parameters of the two camera pairs separately, which will later be used for obtaining 3D information from the activity space.

4.4.1 ARAT Capture System

The INR team, including myself, developed a capture system to record patients performing the ARAT test under clinical conditions. The Industrial Design member of the team, developed and fabricated a rigged system consisting of three cameras that mount to the activity table. The concept of achieving invariance across multiple therapy sessions was at the core of ARAT capture system design.

In collaboration with expert therapists, we designed the rigged system to consist of three cameras. One on each side of the activity table (sagittal left and sagittal right) and an overhead/transverse camera to capture the patients performing ARAT exercises. The goal was to ensure that the cameras capturing the activities are as unobtrusive as possible to the therapists administering the test and the patients attempting it. The rigged system accomplishes the same while ensuring that the cameras are positioned at standardized locations relative to the activity space to maintain invariance across captures. The captured videos are later utilized by the Computer Vision team to perform automated analysis of movement quality, functionality, and task completion by the patients [4].

4.4.2 Development of Camera Calibration approach

I implemented the preliminary version of the camera calibration approach at the INR lab on the SARAH capture system post activity space calibration (described in Section 4.1). The tools I incorporated to accomplish a stereo calibration on the SARAH capture system are a checkerboard image printed on a letter-sized paper, a python script to capture images from the two cameras, and the MATLAB stereo camera calibrator application.

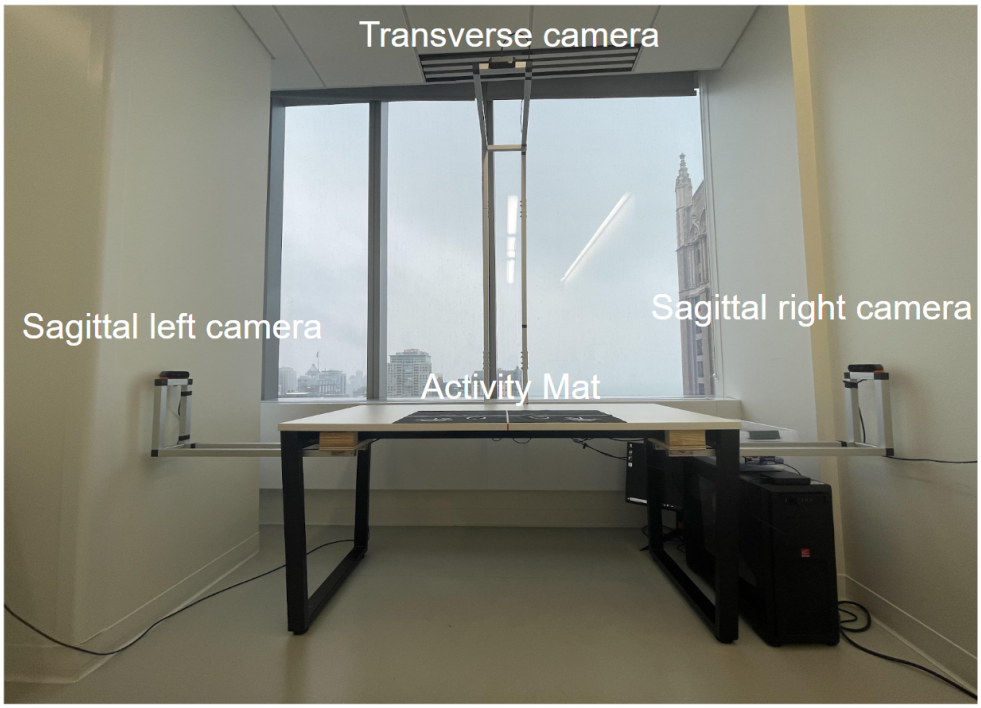

Figure 4.9: ARAT capture system.

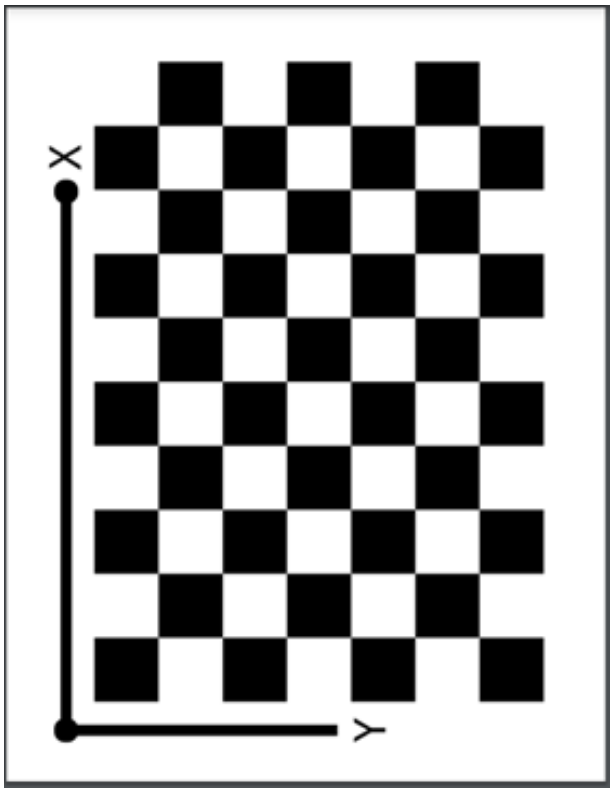

Figure 4.10: Checkerboard pattern used for camera calibration obtained from MATLAB.

The stereo camera calibrator app on MATLAB requires the user to supply image pairs of the checkerboard visible in both cameras. This process required me to run the python script to capture around 20 image pairs (from side and front cameras), wherein the checkerboard is at different locations in each capture. The crucial requirement for this process is to ensure that the checkerboard is clearly visible in both the camera frames in every image pair captured. I then supplied the images to the stereo calibrator app, which rejects pairs in which the checkerboard is not visible or in cases where all the checkerboard corners are not detected.

Figure 4.11: Side (left) and front (right) camera image pairs used to detect checkerboard

The app utilizes the supplied image pairs to perform stereo calibration of the involved cameras and extracts the extrinsic and intrinsic parameters of the cameras along with the rotation and translation vectors between the cameras. This information is visualized on the app as the location of the cameras in space, along with reprojection errors calculated for each image pair. A reprojection error is a geometric error corresponding to the image distance between a projected point, and a measured one [17]. In stereo-calibration, one of the cameras is considered to be a primary camera, and the other is a secondary camera. The secondary camera's detected checkerboard corner coordinates are projected onto the checkerboard corner coordinates detected in the primary camera to calculate the mean distance between the points to produce the mean re-projection error.

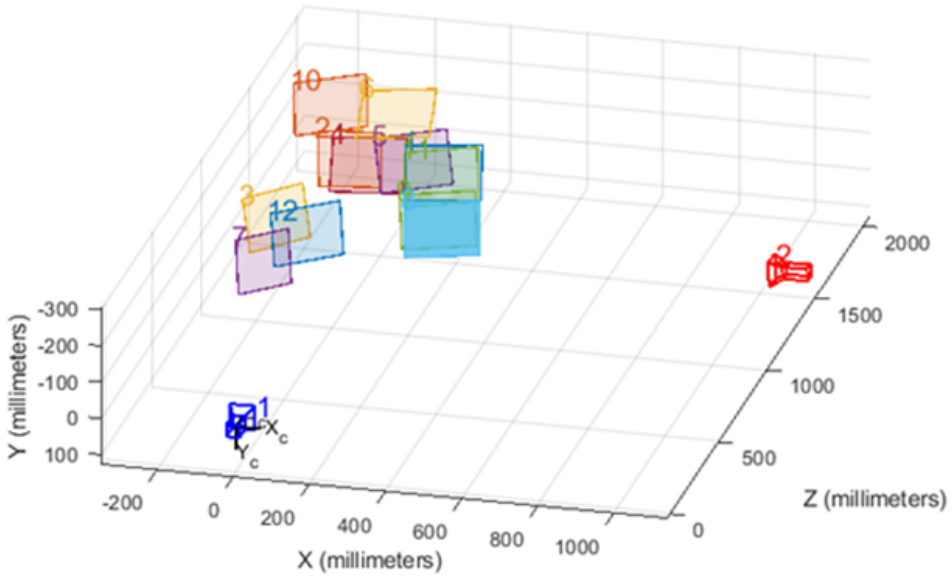

Figure 4.12: Visualization of camera placement detected by the calibrator app.

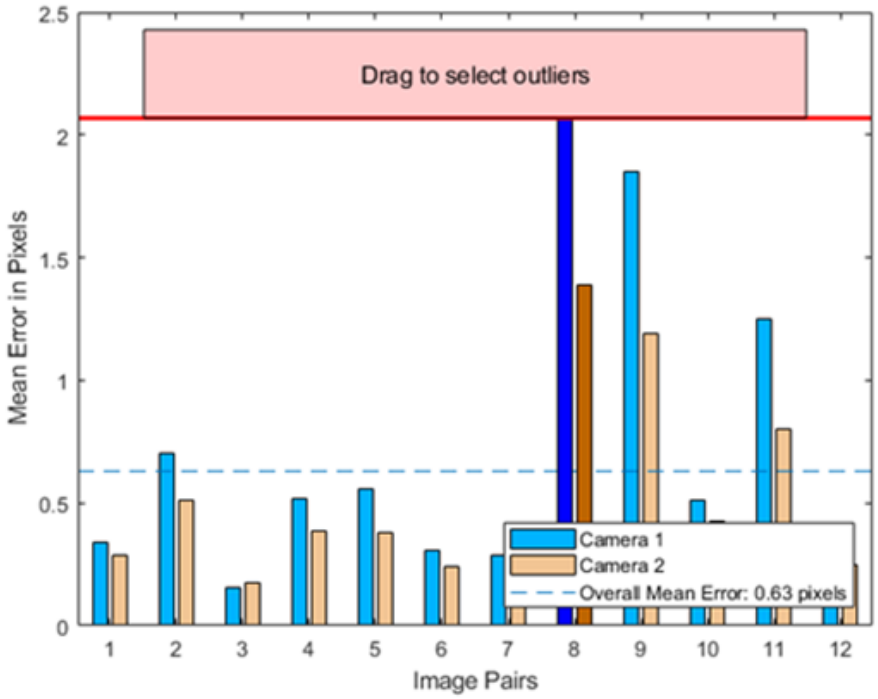

Figure 4.13: Re-projection errors of detected image pairs.

Considering the promising results obtained from the calibration of cameras in the SARAH capture system (accurate visualization of cameras in space and a low overall mean error of 0.63 pixels), I applied the same approach to the ARAT capture system. In the case of ARAT, I employed the pair-wise stereo calibration approach described in Section 4.2 to extract the stereo parameters of the camera pairs. The same checkerboard, python script, and calibrator app were used in the preliminary testing phase for both ARAT and SARAH capture systems.

4.4.3 Optimizing Camera Calibration process

Considering the high number of image pairs required (approximately 20) for a successful stereo calibration, the entire process of capturing multiple images for both camera pairs, i.e. transverse and sagittal left; and transverse and sagittal right cameras running the calibrator app, took around 25 minutes to complete for the ARAT capture system. In a clinical setting, the clinician is expected to set up and calibrate the cameras of the ARAT capture system before bringing the patient in and administering the test. Setting aside 30 minutes before every ARAT session and performing a highly time-intensive camera calibration process is not feasible in a clinical environment. Considering that an entire ARAT assessment takes only 15-20 minutes to complete, the therapists cannot afford to spend 30 minutes just to set up the system.

To tackle this challenge, we set a goal of accomplishing a pair-wise multi-camera calibration of the cameras in under two minutes. To achieve this, we considered the use of a one-minute video recording involving the movement of a checkerboard in space such that the checkerboard is facing the sagittal right and transverse cameras for the first half of the video and facing the sagittal left and transverse cameras for the second half of the video. All three cameras are recording during the checkerboard movement performed by the therapist.

The therapist's instructions for moving the checkerboard in space are designed such that they perform the checkerboard movement facing the sagittal right and transverse cameras for the first 30 seconds of the video and then rotating 180 degrees to face the sagittal left and transverse cameras for the next 30 seconds of the video. We then extracted the image frames from the synchronized video pairs (sagittal right - transverse: first half of right and transverse camera recordings; sagittal left - transverse: second half of left and transverse camera recordings). We supplied the extracted frames to the MATLAB stereo camera calibrator application. The extracted stereo camera parameters for both pairs are locally stored on the machine for future use.

The checkerboard movement that the therapist is instructed to perform is designed after testing with various possible checkerboard movements. The goal was to achieve a smooth extraction of frames (extracted frames cannot be blurred) while ensuring that we could extract a sufficient number of frames for both camera pairs to achieve successful calibration results. The final movement we decided on was instructed to the therapists as follows: "Hold the calibration object using the handles and move in a semi-circular motion starting from the shoulder and ending at the hip. Turn around to face the other side camera and repeat the process." Each semi-circular motion is expected to take approximately 12 seconds, and the therapists are provided with a timer on the camera calibration screen.

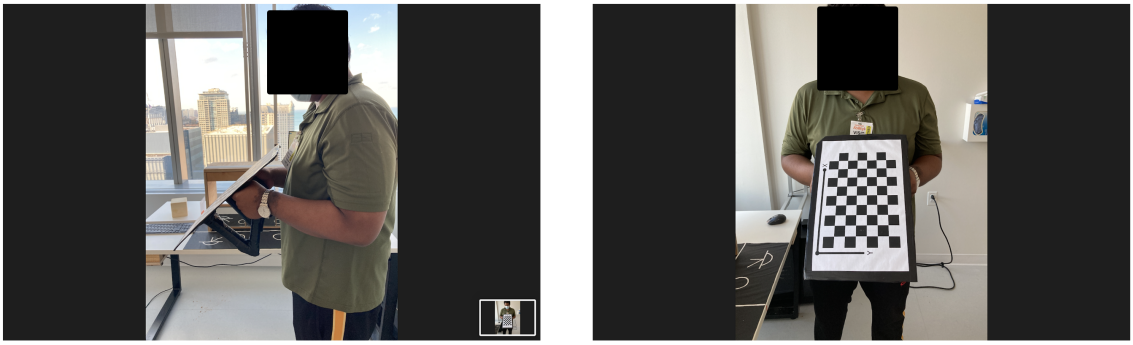

Figure 4.14: Checkerboard calibration object for ARAT camera calibration.

4.4.4 Camera calibration workflow and capture tool integration

To streamline the calibration process workflow, we developed an external synchronized video capture tool on MATLAB, which failed to capture high-quality synchronized videos of checkerboard movement in space. To our advantage, the ARAT capture interface was already capable of capturing high-quality synchronized videos from all the cameras involved in the system. We leveraged this feature of the capture interface to incorporate a camera calibration option on the home screen of the capture tool.

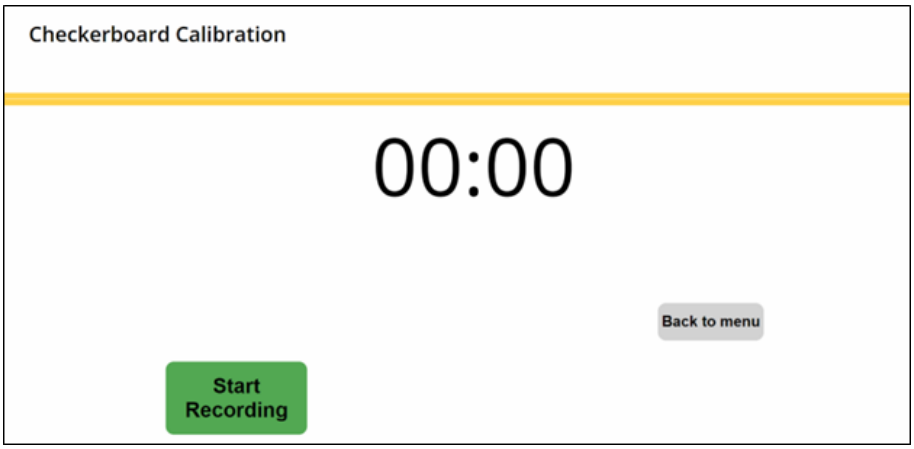

Figure 4.15: Video capture screen for calibration videos with timer.

The calibration screen allows the therapists to start and stop recording when they complete the checkerboard motion, and the videos are automatically downloaded to the local system once the therapist clicks on the "Stop Recording" button. Following the recording process, the therapists are instructed to open a desktop MATLAB application pinned to the taskbar and click on the "calibrate" button provided on the MATLAB interface.

The MATLAB interface we developed consists of a "calibrate" button that performs frame extraction and pair-wise stereo camera calibration on both camera pairs. The advantage of employing the MATLAB calibration application, independent of the capture tool, is to allow the therapists to continue the regular ARAT assessment right after clicking on the

calibrate button on the MATLAB interface. This simultaneously runs the frame extraction and calibration processes in the background, which would otherwise keep the therapists waiting for another 10-15 minutes if implemented into the capture tool. This allowed us to achieve a two-minute calibration process from a therapist's point of view, which proved to be a crucial factor in significantly bringing down the time spent by the therapists on setting up and calibrating the system.

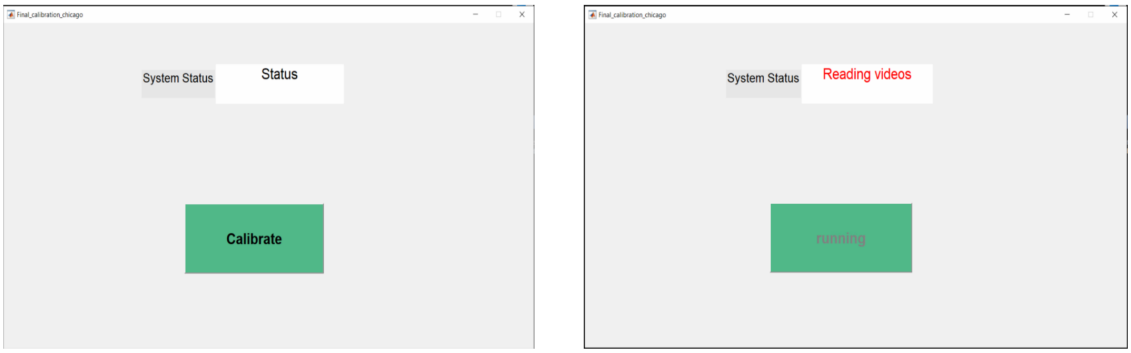

Figure 4.16: MATLAB calibration interface (left - first screen, right - after clicking on "calibrate" button).

The therapists then return to the home page, where they proceed to begin the ARAT assessment. To ensure the integrity of the hardware involved in the rigged system (cameras and the position of the cameras relative to the activity space), the therapists are required to pass through a "Camera Check" screen. The camera checks screen records and stores the coordinates of the Aruco fiducial marker vertices (as introduced in Section 4.1), which allows the development team to keep track of any potential movements undergone by the rigged system. The Camera Check screen also requires the therapists to manually check the boxes on all cameras that are displayed on the screen to ensure that the cameras are fully functional before proceeding with the ARAT assessment process.

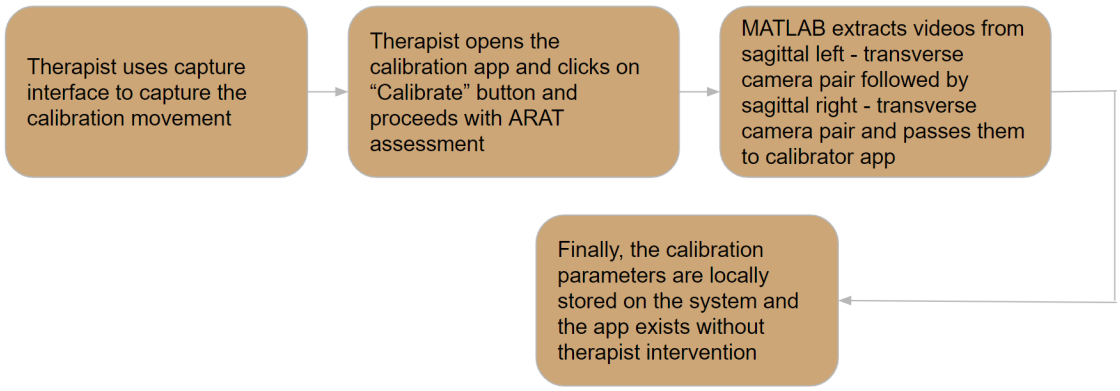

Figure 4.17: Camera calibration workflow.

4.5 Activity Space Reconstruction

Sections 4.1 and 4.2 discussed the role of activity space calibration and multi-camera calibration in capture systems that aim to record and perform automated analysis on upper extremity stroke rehabilitation patient movements. One recurring problem that we encountered in our video-based captures is the apparent loss of data due to the 2D projection of the 3D world scene. This manifests in different problems such as occlusions and loss of high precision data on the hand-finger movement. In this chapter, I thus discuss the efforts made to obtain three-dimensional information from the two-dimensional videos captured through the calibrated capture systems. Specifically, I discuss how we reconstruct the object trajectory, a sparse reconstruction of hand key points obtained from OpenPose [4], and a strategy for the dense hand and object reconstruction along with an additional feature added to the Video Annotation/ Rating tool are discussed. The sections in this chapter build upon the activity space calibration, and camera calibration work discussed previously and justifies the importance of calibration in capture systems.

4.5.1 3D reconstruction of object trajectory

Utilizing the extracted camera parameters resulting from the multi-camera calibration approach, we triangulated the object's position captured from the multi-cameras. The algorithms for triangulation require the stereo calibration parameters generated for a particular pair of cameras and the image frame coordinates of the corresponding points in two video frames. The image frame coordinates of the object at each video frame were initially manually selected and provided to the triangulation function. The future work will involve an automated object detection pipeline which will provide the object coordinate information from each camera. Thus we can fully automate the 3D reconstruction of the objects.

4.5.2 Sparse 3D reconstruction of patient hand

Similar to the reconstruction of object trajectory, we implement the pipeline for a sparse 3D reconstruction of the human hand. To this end, I implemented a fully automated pipeline. We first obtain key points of the hand using the publicly available open-source implementation of Open-pose, which is state-of-the-art software for human pose extraction. By triangulating the 2D key points using the intrinsic and extrinsic parameters of the cameras, we were able to reconstruct a sparse representation of a patient's impaired hand in 3D space.

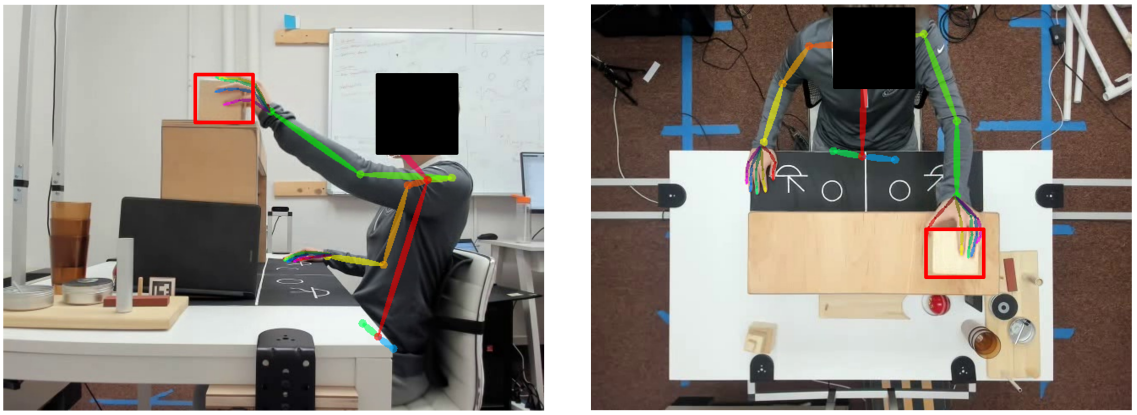

Figure 4.18: Open pose keypoints and object highlighted on sagittal left (left) and transverse cameras (right).

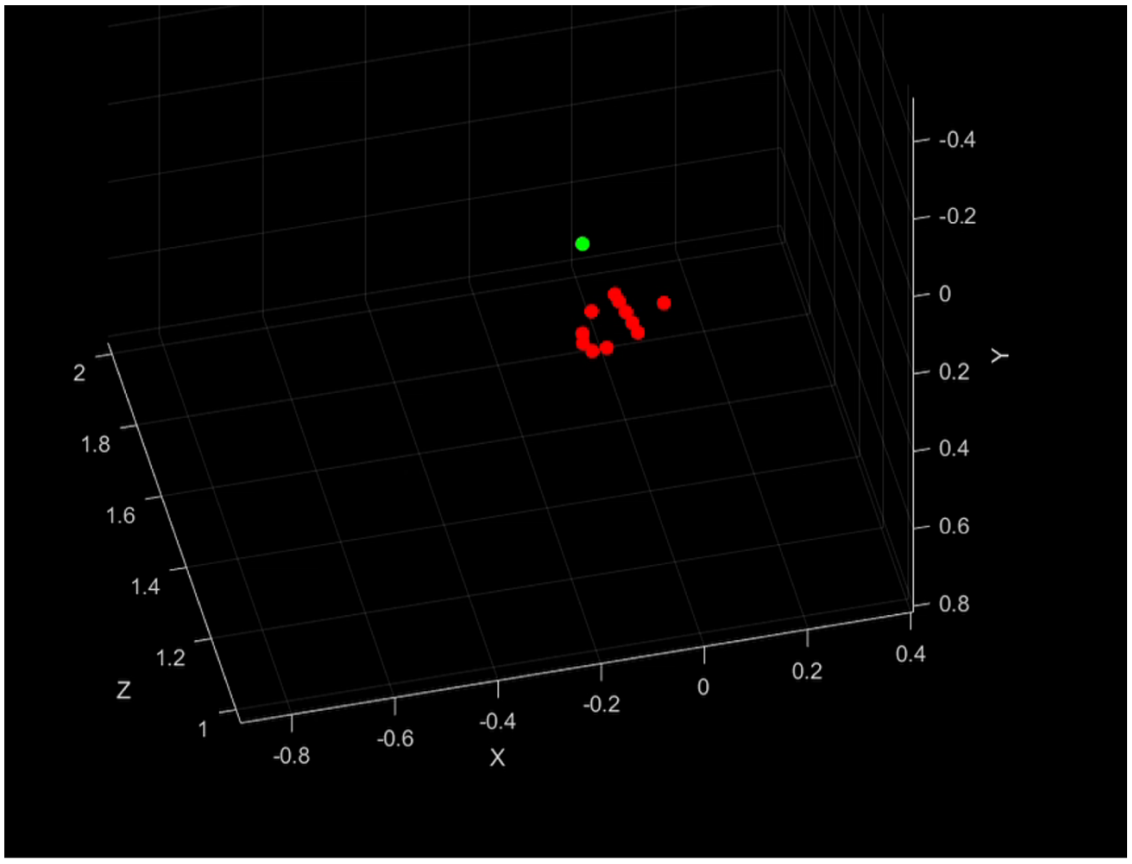

Figure 4.19: Sparse reconstruction of hand key points (red) and object center (green) in three dimensions.

4.5.3 Termination tab on rating tool

To provide the therapist with finer details of the hand object relationship for better ratings, we integrate the 3D reconstructed hand and object into the rating interface. To achieve this integration meaningfully, we created an additional tab called the termination tab on the rating tool, which allows the therapists to choose if they are interested in taking a closer look at the reconstructed hand and object for better clarity on patient performance for respective tasks. To render the 3D perspective itself, we added an additional view tab called "3D view", which allows the therapists to utilize an interactive 3D model of the reconstructed hand and object to take a closer look at both from different viewing angles. The reconstructions generated as the results of future work will be integrated into the rating interface into the termination tab.

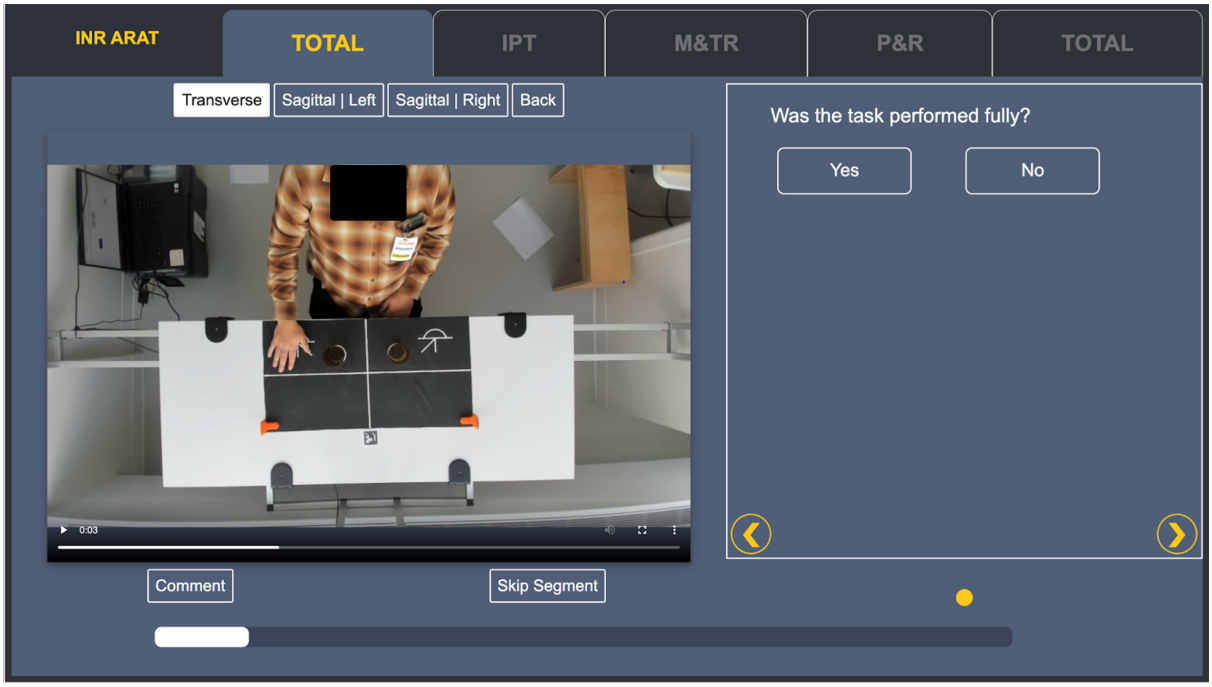

Figure 4.20: Rating tool without termination tab.

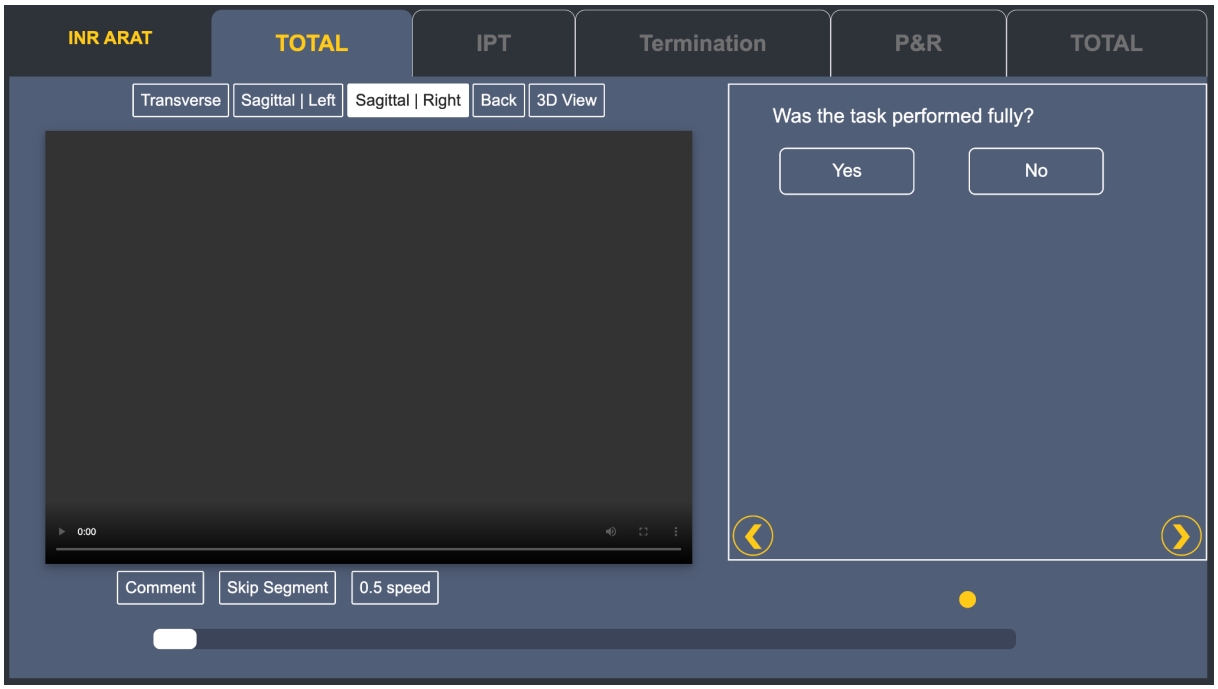

Figure 4.21: Rating tool with termination tab and 3D view.

4.6 Future work

Future work in this area involves a dense reconstruction of the patient's impaired hand and object in three-dimensional space. The INR team is currently working on obtaining a 3D reconstruction of the impaired hands of patients utilizing computer vision-based Machine Learning Models.

The reconstructed 3D hand is integrated into the ARAT video annotation tool to allow the therapists to gain insights into how the patient is interacting and grasping the object. This information is crucial in assessing ARAT exercises and will also be utilized by the Computer vision team to integrate the patient's hand-object relationship into the assessment pipeline.

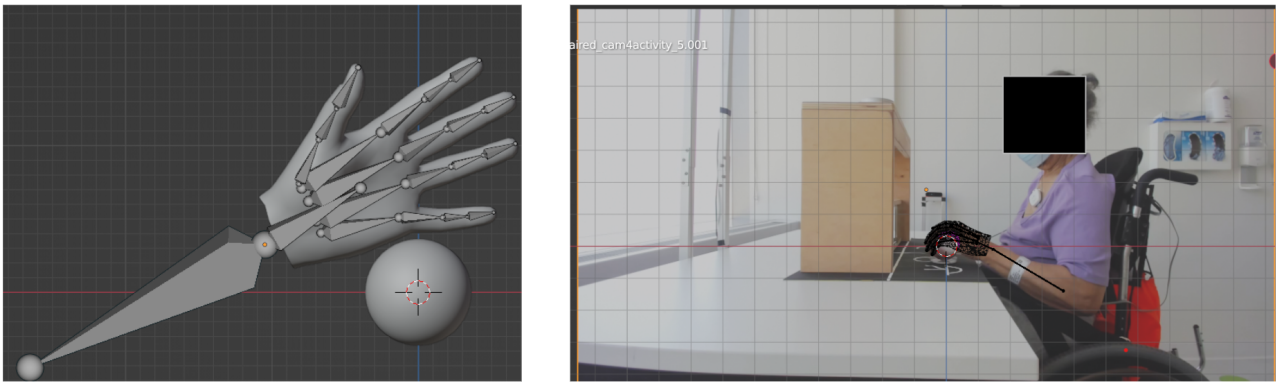

Figure 4.22: Manual annotation of hand and object relationship (left) for training, and reconstructed hand overlaid on top of patient hand (right).

Chapter 5

Results and Evaluation

In this chapter, I discuss the workshops we conducted to test and evaluate the activity space and multi-camera calibration workflows with clinical and occupational therapists. Section 5.1 details the workshop conducted in Roanoke, Virginia (with the SARAH capture system) by the UX members of the INR team, and the demonstrate the qualitative and quantitative results obtained from this workshop. Section 5.2 details two workshops conducted at the Shirley Ryan Ability Lab in Chicago [2] to test the ARAT capture system along with the multi-camera calibration workflow I developed. The results obtained from these workshops provided crucial feedback for future improvements to the calibration workflows of the capture systems.

5.1 Roanoke Study

The activity space calibration approach, integrated into the SARAH capture interface, was tested by six occupational therapists recruited by our team in Roanoke. The primary focus was to test the timing of the calibration process and variance in circle detection in calibrated SARAH setup across various activity space calibrations performed by the participating therapists.

The therapists provided overall positive feedback on performing the setup and calibration process of the SARAH capture system. They raised minor concerns about how it can be te-

dious to pass the camera placement verification step. The process requires precise placement of the tripods relative to the activity mat and table. However, the therapists believed that it was a minor concern that would not deter them from performing the setup and calibration process. "like putting together IKEA furniture, but much simpler", "Absolutely could be in the home of a stroke patient", and "Someone who's not tech-savvy might be able to do it" are some of the comments made by them during the interview sessions conducted by the INR team at the end of the workshop providing positive qualitative feedback. The therapists performed the calibration workflow involving computational verifications in under four and a half minutes on average.

Table 5.1: Time duration of calibration workflow across six therapists (hh:mm:ss) [9]

Therapist 1	Therapist 2	Therapist 3	Therapist 4	Therapist 5	Therapist 6
0:06:46	0:02:59	0:07:14	0:03:19	0:02:42	0:01:38

We further analyzed the collected pixel coordinate data on the coordinates of the detected circles on both camera views. The standard deviations calculated for the detected circles in each row on both camera views are depicted in the tables below.

Table 5.2: Standard deviation of circle pixel coordinates of front camera [9]

Side Camera Calibration (Standard Deviation X, Standard Deviation Y)				
Last Row	(5.09, 8.64)	(14.65, 8.67)	(15.11, 8.77)	(16.29, 8.69)
Middle Row	(14.63, 7.96)	(13.44, 8.22)	(13.76, 13.39)	(15.80, 8.32)
First Row 2	(13.27, 6.95)	(13.40, 7.06)	(14.37, 7.53)	(17.43, 8.03)

Table 5.3: Standard deviation of circle pixel coordinates of side camera [9]

Front Camera Calibration (Standard Deviation X, Standard Deviation Y)			
Last Row	(23.47, 2.6)	(20.19, 1.24)	(21.68, 1.29)
Middle Row - 1	(23.16, 2.98)	(19.51, 1.42)	(21.46, 1.63)
Middle Row - 2	(24.24, 3.5)	(17.91, 1.24)	(21.0, 1.80)
First Row	(23.15, 5.24)	(16.5, 1.73)	(20.10, 2.73)

5.1.1 Machine Learning Improvements

One of the major goals of the activity space calibration workflow is to provide the Computer Vision team with calibrated captures with known prior information on the location of activity space, patient, and the mat on the captured video frames while ensuring that the occupational therapists are able to set up and calibrate the system with ease. We have shown the positive feedback regarding the latter from the Roanoke study. For testing the calibration workflow for improvement in automated assessment by the Machine Learning pipeline, the Computer Vision team performed automated analysis on patient activity captures involving a calibrated SARAH setup. The tables below show the results recorded from this experiment in comparison to the results achieved from the videos captured through an uncalibrated and unstandardized SARAH system.

Table 5.4: Automated segmentation results from uncalibrated SARAH setup [4]

	Ensemble1		Ensemble2		Ensemble3	
	Mean	STD	Mean	STD	Mean	STD
ACC	81.01	1.99	83.76	2.77	85.08	2.14
Precision	77.28	4.38	81.40	3.92	84.34	2.61
Recall	77.88	4.46	81.57	3.66	84.6	2.68

Table 5.5: Automated segmentation results from calibrated SARAH setup [5]

	Train	Test
Experiment 2a	trained model using impaired	unimpaired data
Experiment 2b	75% impaired 25% unimpaired	test set using impaired
Experiment 2c	75% impaired 25% unimpaired	60% impaired 40% unimpaired

	Experiment 2a	Experiment 2b		Experiment 2c	
	Mean	Mean	STD	Mean	STD
ACC	80.52	86.01	1.28	87.85	0.58

We have observed a significant improvement in the standard deviation of automated segmentation of the patient movement analysis. This proved that the calibration workflow, apart from providing standardized patient captures for rating purposes, significantly improved the automated analysis of captured tasks while making sure the setup process was easily performable by the therapists. This is a crucial inclusion for an in-home-based telehealth capture system for stroke rehabilitation, considering the enhancement of automated analysis and the ease of executing the interactive calibration process by the therapists from a Human-Computer Interaction perspective.

5.2 Shirley Ryan Ability Lab Workshops

The INR team conducted two workshops at the Shirley Ryan Ability Lab in Chicago. The workshops were aimed to install the ARAT capture system in real world clinical setting to test and train the clinicians on operating the system. The goal of the INR team was to obtain 100 patient captures at the clinic for automated analysis of patient performance in the ARAT setup.

5.2.1 Shirley Ryan Ability (SRA) Lab - Workshop I

The INR team conducted a preliminary workshop at the Shirley Ryan Ability Lab in Chicago in November 2021. We installed the ARAT captured system for a pilot study involving two stroke patients and five clinical therapists in this workshop. Four sessions of ARAT tests were administered by the clinicians in the presence of the ARAT development team.

Apart from the array of cameras and the rigged system, the ARAT system consists of an

interactive Capture tool and a Video Annotation/ Rating tool developed by the UX team at the INR lab. The capture tool allows the therapists to walk through the exercises administered in the Action Research Arm Test (ARAT). The Video annotation/ Rating tool is utilized by the clinicians to look back and rate the patient's performance according to a rating rubric developed by the INR team [34]. The goal of the annotation tool is to understand the reasoning for the scores assigned for each exercise by the therapists while administering the test.

During the capture sessions at the workshop, Dr. Steve Wolf and the therapists pointed out the difficulty in observing how the patient grasps the ARAT objects during the assessment. One of the limitations of the ARAT capture system is that the clinicians are restricted to staying behind a marking on the floor to avoid obstructing the cameras during the assessment. This affects their ability to observe the patient's interaction with the objects. The three cameras of the rigged system are positioned to capture the patient's upper torso, elbow, and hand movements to perform a holistic capture of the activity space and the patient to facilitate better Machine Learning analysis of the patient's movement. All the captured views are presented in the Rating tool to allow the therapists to review the videos better, leading to a more accurate rating process.

One of the crucial aspects of the ARAT assessment that was discussed during a debrief session at the end of the workshop was the idea of obtaining three-dimensional information from the captured videos. Specifically, a 3D reconstruction of the impaired hand, object, and object's trajectory in a 3D space. The merits of obtaining three-dimensional data include a more informed computer vision analysis of the videos [16]. Moreover, providing a reconstructed view of the patient's impaired hand and object in a 3D space in the rating tool would provide the clinicians with a detailed representation of hand object relation in each exercise and rate the grasping strategy utilized by the patient. The team's consensus regarding the extraction

of 3D information from the 2D videos was that this effort would provide more granular information regarding patients' interaction with the objects, ultimately leading to a better Computer Vision analysis of the activities.

5.2.2 Shirley Ryan Ability Lab (SRA) - Workshop II

We conducted a second workshop at the SRA lab in Chicago in March 2022 with the updated version of the ARAT capture interface, including the camera check screen, camera calibration screen, and the MATLAB interface for performing pair-wise stereo calibration. This workshop marked the beginning of an IRB-approved ARAT data collection initiative undertaken by the INR lab to create the first-ever national database of upper extremity stroke rehabilitation captures.

We installed the ARAT capture system with assistance from the SRA lab staff in a dedicated clinical space assigned for administering ARAT tests. Before transporting it to Chicago, we performed multiple stress tests on the capture interface and the MATLAB calibration interface at the INR lab at Virginia Tech. During the three-day workshop at the SRA lab, we collaborated with expert clinical therapists to administer ARAT tests for four patients who volunteered to participate in the study. We provided a thorough walk-through of the calibration workflow and worked with the clinicians to help them get acquainted with the system. In collaboration with my advisor, Dr. Kelliher, we delivered a detailed written manual on the calibration workflow containing visual aids so that the clinicians could refer to it in case of ambiguity in the absence of the development team. Overall, we received positive feedback on the calibration process, and the therapists were comfortable using it in their daily practice.

The clinical therapists provided valuable feedback and communicated their concerns which

allowed us to make minor improvements and adjustments to the system during the workshop. Some of the suggested improvements were:

1) A summary of scores at the end of every ARAT session so they could inform the patients of their performance - We integrated this feature on the "Thank You" screen to indicate the scores per subscale and the total ARAT assessment score of that session.

2) Provide special instructions for certain exercise screens to remind clinicians to make necessary adjustments to the activity space for the respective exercises. We incorporated the suggested special instructions as per the therapist's requests.

3) Include a button to autofill the comment section when the patient requests to stop the exercise early - We incorporated a "Stopped Early" button that auto-fills the comment section with "Patient requested to stop early". This was one of the most commonly used comments, and removing it allowed the therapists to populate the comment section quickly without manually typing in the comment.

Over the course of the workshop, the therapists were able to perform the camera calibration process involving the checkerboard motion and triggering the MATLAB calibration application in under two minutes as we had aimed for. One of the patient capture results for Sagittal left - Transverse and Sagittal right - Transverse camera pairs are depicted below.

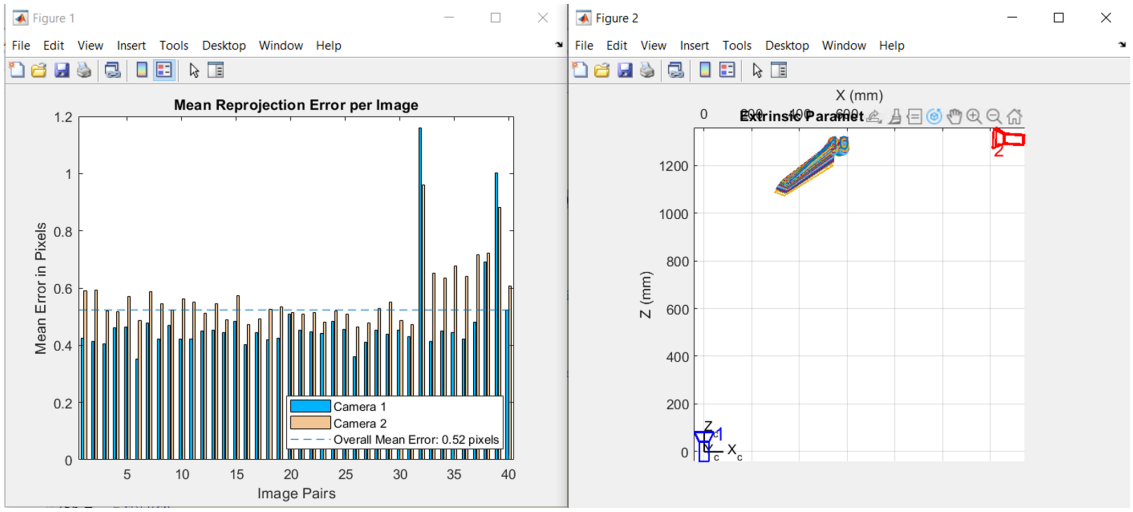

Figure 5.1: Visualization (left) and mean errors (right) in Sagittal left - Transverse camera pair.

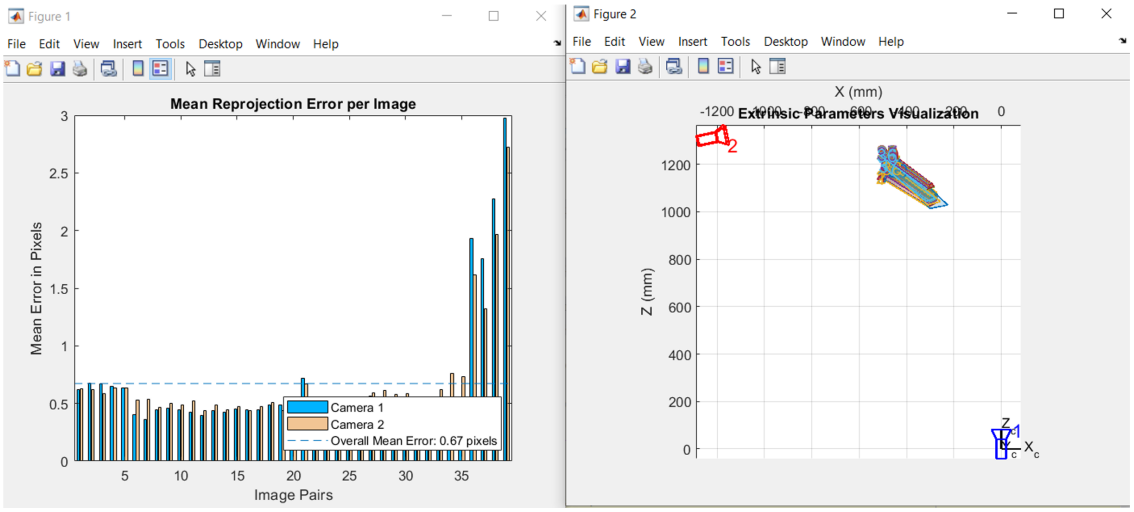

Figure 5.2: Visualization (left) and mean error (right) in Sagittal right - Transverse camera pair.

Chapter 6

Conclusions

Over the course of a 12-month design and development, activity space calibration and multi-camera calibration methods have improved the qualitative and quantitative understanding of captured videos of upper extremity stroke rehabilitation videos. In addition, I successfully developed and deployed the activity space calibration approach to ensure invariance in video captures across multiple stroke rehabilitation sessions and improved automated assessment results generated by the computer vision team in the SARAH system.

The multi-camera calibration approach has successfully generated highly accurate camera calibration results. It is integrated into the ARAT capture system requiring a simple 2-min calibration procedure performed by the clinical therapists. This work has opened up significant research paths toward developing 3D reconstruction approaches to help provide better insights to both the therapist and the Machine Learning models performing automated analysis of upper extremity stroke rehabilitation videos of patients. The discussed methods have shown the idea of designing and calibrating capture systems for stroke rehabilitation purposes to aid the therapists and the Computer vision-based models, resulting in mutual learning between both the therapists and the Machine Learning models.

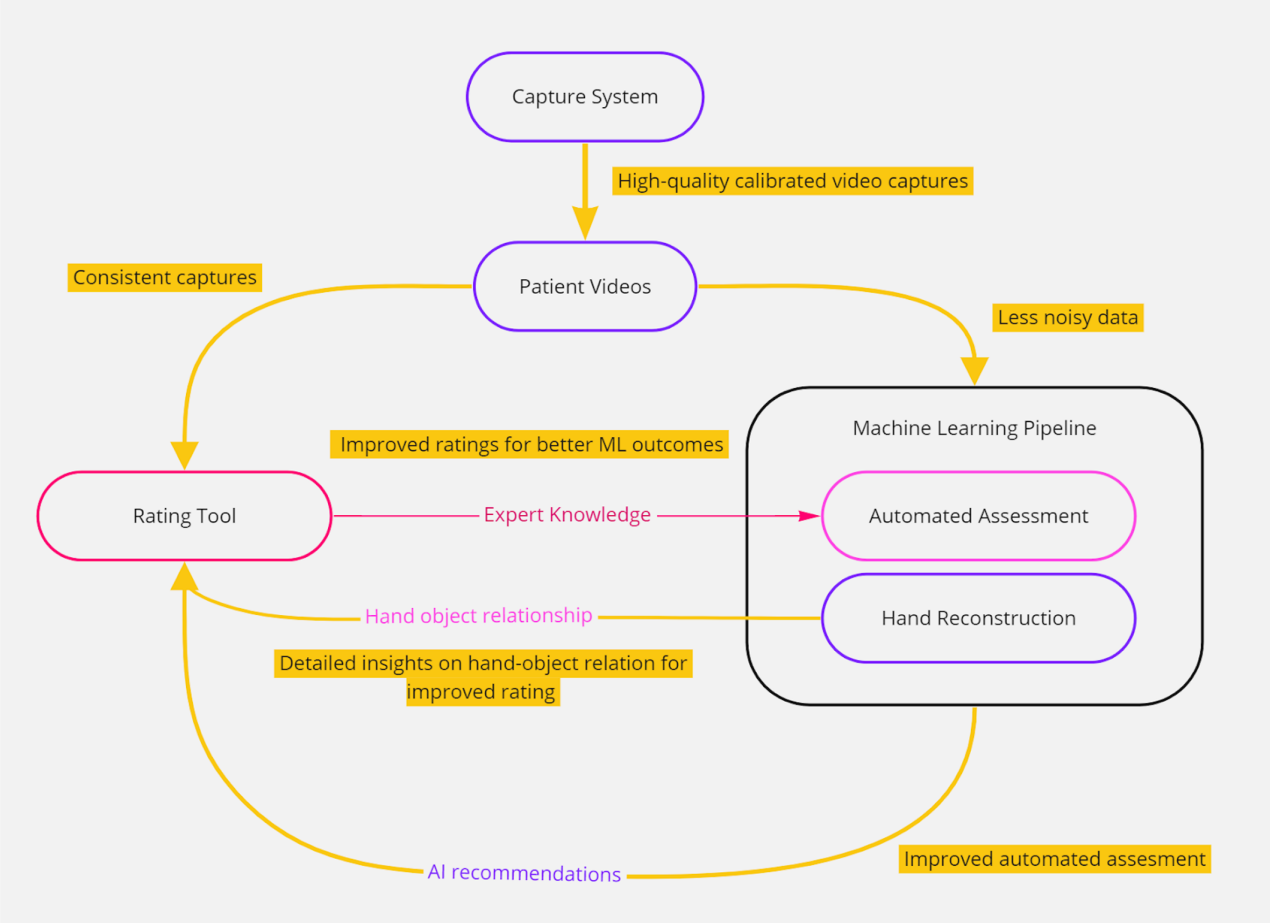

Figure 6.1: Enhanced flow diagram detailing my contributions.

Figure 6.1 depicts an enhanced version of the flow diagram shown in figure 1.1. In this architecture, the primary enhancements are obtained due to the activity space calibration workflow resulting in standardized and invariant video captures, which lead to less noisy data for the machine learning pipeline and the rating tool. This led to improved automated analysis of patient movement and better and more consistent ratings from the therapists. In addition, the multi-camera calibration approach has opened up avenues for facilitating a symbiotic relationship between the machine learning pipeline and the rating tool. Due to the camera parameters being made available to the pipeline, the INR team can develop machine learning models to reconstruct the hand of the patient in three dimensions accurately. This

information provides fine-grained details on the patient's hand object relationship during the exercises and allows the expert therapists to provide better ratings and assessments of the captured videos. The improved expert ratings help machine learning produce better automated assessment results. The activity space and camera calibration workflows are vital components that allowed us to enhance this architecture.

Bibliography

[1] Camera calibration. URL https://www.mathworks.com/help/vision/camera-calibration.html. Last accessed: June 1, 2022.

[2] Abilitylab home. URL https://www.sralab.org/. Last accessed: June 1, 2022.

[3] Stroke facts, Apr 2022. URL https://www.cdc.gov/stroke/facts.htm. Last accessed: June 1, 2022.

[4] Tamim Ahmed, Kowshik Thopalli, Thanassis Rikakis, Pavan Turaga, Aisling Kelliher, Jia-Bin Huang, and Steven L Wolf. Automated movement assessment in stroke rehabilitation. *Frontiers in Neurology*, page 1396, 2021.

[5] Tamim Ahmed, Thanassis Rikakis, Setor Zilevu, Aisling Kelliher, Kowshik Thopalli, Pavan Turaga, and Steven L Wolf. A Hierarchical Bayesian Model for Cyber-Human Assessment of Rehabilitation Movement. *medRxiv*, 2022. doi: 10.1101/2022.05.25.22275480. URL https://www.medrxiv.org/content/early/2022/05/27/2022.05.25.22275480.

[6] Filippo Bergamasco, Andrea Albarelli, Emanuele Rodola, and Andrea Torsello. Rune-tag: A high accuracy fiducial marker with strong occlusion resilience. In *CVPR 2011*, pages 113–120. IEEE, 2011.

[7] Samson LW Sheingold S Taplin C Tarazi W Bosworth A, Ruhter J and Zuckerman R. Beneficiary use of telehealth visits: Early data from the start of covid-19 pandemic. 2020. URL https://ASPEaspe.hhs.gov.

[8] Juliet Clark, Setor Zilevu, Tamim Ahmed, Aisling Kelliher, Sai Krishna Yeshala, Sarah Garrison, Cathleen Garcia, Olivia C Menezes, Minakshi Seth, and Thanassis Rikakis. Hybrid workflow process for home based rehabilitation movement capture. In *ACM International Conference on Interactive Media Experiences*, pages 241–246, 2021.

[9] Juliet Ariana Clark. Designing telehealth rehabilitation systems for diverse stakeholder needs, May 2021. URL https://vtechworks.lib.vt.edu/handle/10919/103526.

[10] Ludovic David, Guillaume Bouyer, and Samir Otmane. Towards an upper limb self-rehabilitation assistance system after stroke. *Proceedings of the Virtual Reality International Conference - Laval Virtual 2017*, 2017.

[11] William W. Gaver. Technology affordances. In Scott P. Robertson, Gary M. Olson, , and Judith S. Olson, editors, *CHI 91 Conference on Human Factors in Computing New Orleans, LA, USA April 27 - May 02, 1991*, pages pp.79–84, New Orleans, Lousiana, United States, 1991. ACM. URL https://hal.archives-ouvertes.fr/hal-00692032.

[12] Donald B Gennery. Stereo-camera calibration. In *Proceedings ARPA IUS Workshop*, pages 101–107, 1979.

[13] Jose-Joel Gonzalez-Barbosa, Teresa Garcia-Ramirez, Joaquin Salas, Juan-Bautista Hurtado-Ramos, and Jose-de-Jesus Rico-Jimenez. Optimal camera placement for total coverage. In *2009 IEEE International Conference on Robotics and Automation*, pages 844–848, 2009. doi: 10.1109/ROBOT.2009.5152761.

[14] Hyowon Ha, Michal Perdoch, Hatem Alismail, In So Kweon, and Yaser Sheikh. Deltille grids for geometric camera calibration. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 5354–5362, 2017. doi: 10.1109/ICCV.2017.571.

[15] Eva Hörster and Rainer Lienhart. On the optimal placement of multiple visual sensors. In *Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks*, pages 111–120, 2006.

[16] Anastasia Ioannidou, Elisavet Chatzilari, Spiros Nikolopoulos, and Ioannis Kompatsiaris. Deep learning advances in computer vision with 3d data: A survey. *ACM Computing Surveys (CSUR)*, 50(2):1–38, 2017.

[17] HUO Ju, LI Yunhui, and Yang Ming. Multi-camera calibration method based on minimizing the difference of reprojection error vectors. *Journal of Systems Engineering and Electronics*, 29(4):844–853, 2018.

[18] Aisling Kelliher, Andrew Gibson, Eric Bottelsen, and Edward Coe. Designing modular rehabilitation objects for interactive therapy in the home. In *Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 251–257, 2019.

[19] J. A. Kleim and T. A. Jones. Principles of experience-dependent neural plasticity: implications for rehabilitation after brain damage. *J Speech Lang Hear Res*, 51(1):S225–39, 2008. ISSN 1092-4388 (Print) 1092-4388 (Linking). doi: 10.1044/1092-4388(2008/018). URL https://www.ncbi.nlm.nih.gov/pubmed/18230848.

[20] Jeff Kleim and Theresa Jones. Kleim ja, jones taprinciples of experience-dependent neural plasticity: implications for rehabilitation after brain damage. j speech lang hear res 51:s225-s239. *Journal of speech, language, and hearing research : JSLHR*, 51:S225–39, 03 2008. doi: 10.1044/1092-4388(2008/018).

[21] Arturo la Escalera and Jose María Armingol. Automatic Chessboard Detection for Intrinsic and Extrinsic Camera Parameter Calibration. *Sensors*, 10(3):2027–2044, 2010.

ISSN 1424-8220. doi: 10.3390/s100302027. URL https://www.mdpi.com/1424-8220/10/3/2027.

[22] Bland-M.-Bailey R. Schaefer S. Birkenmeier R. Lang, C. Assessment of upper extremity impairment, function, and activity after stroke: foundations for clinical decision making. *J Hand Ther*, 26(2):104–14;quiz 115, 2013. ISSN 1545-004X (Electronic) 0894-1130 (Linking). doi: 10.1016/j.jht.2012.06.005. URL https://www.ncbi.nlm.nih.gov/pubmed/22975740.

[23] Albert C Lo, Peter Guarino, Hermano I Krebs, Bruce T Volpe, Christopher T Bever, Pamela W Duncan, Robert J Ringer, Todd H Wagner, Lorie G Richards, Dawn M Bravata, et al. Multicenter randomized trial of robot-assisted rehabilitation for chronic stroke: methods and entry characteristics for va robotics. *Neurorehabilitation and neural repair*, 23(8):775–783, 2009.

[24] Lilika Makabe, Hiroaki Santo, Fumio Okura, and Yasuyuki Matsushita. Shape-coded aruco: Fiducial marker for bridging 2d and 3d modalities. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2655–2664, 2022.

[25] Michelle McDonnell. Action research arm test. *The Australian journal of physiotherapy*, 54:220, 02 2008. doi: 10.1016/S0004-9514(08)70034-5.

[26] Alexandros Pantelopoulos and Nikolaos G. Bourbakis. A survey on wearable sensor-based systems for health monitoring and prognosis. *Trans. Sys. Man Cyber Part C*, 40 (1):1–12, January 2010. ISSN 1094-6977. doi: 10.1109/TSMCC.2009.2032660. URL https://doi.org/10.1109/TSMCC.2009.2032660.

[27] Michael Pfeifer, Mehrsheed Sinaki, Piet Geusens, Steven Boonen, Elisabeth Preisinger, Helmut W Minne, and for the ASBMR Working Group on Musculoskeletal Rehabili-

tation. Musculoskeletal rehabilitation in osteoporosis: A review. *Journal of Bone and Mineral Research*, 19(8):1208–1214, 2004. doi: https://doi.org/10.1359/JBMR.040507. URL https://asbmr.onlinelibrary.wiley.com/doi/abs/10.1359/JBMR.040507.

[28] Cristina Ramírez-Fernández, Alberto L Morán, Eloísa García-Canseco, and Felipe Orihuela-Espina. Design factors of virtual environments for upper limb motor rehabilitation of stroke patients. In *Proceedings of the 5th Mexican Conference on Human-Computer Interaction*, pages 22–25, 2014.

[29] David J Reinkensmeyer and Michael L Boninger. Technologies and combination therapies for enhancing movement training for people with a disability. *Journal of neuroengineering and rehabilitation*, 9(1):1–10, 2012.

[30] David J Reinkensmeyer, Etienne Burdet, Maura Casadio, John W Krakauer, Gert Kwakkel, Catherine E Lang, Stephan P Swinnen, Nick S Ward, and Nicolas Schweighofer. Computational neurorehabilitation: modeling plasticity and learning to predict recovery. *Journal of neuroengineering and rehabilitation*, 13(1):1–25, 2016.

[31] David J Reinkensmeyer, Sarah Blackstone, Cathy Bodine, John Brabyn, David Brienza, Kevin Caves, Frank DeRuyter, Edmund Durfee, Stefania Fatone, Geoff Fernie, et al. How a diverse research ecosystem has generated new rehabilitation technologies: Review of nidilrr's rehabilitation engineering research centers. *Journal of neuroengineering and rehabilitation*, 14(1):1–53, 2017.

[32] Manoj Sivan, Rory J O'Connor, Sophie Makower, Martin Levesley, and Bipinchandra Bhakta. Systematic review of outcome measures used in the evaluation of robot-assisted upper limb exercise in stroke. *J. Rehabil. Med.*, 43(3):181–189, February 2011.

[33] Steven L Wolf, Carolee J Winstein, J Philip Miller, Edward Taub, Gitendra Uswatte, David Morris, Carol Giuliani, Kathye E Light, Deborah Nichols-Larsen, for the EX-

CITE Investigators, et al. Effect of constraint-induced movement therapy on upper extremity function 3 to 9 months after stroke: the excite randomized clinical trial. *Jama*, 296(17):2095–2104, 2006.

[34] Kobla Setor Zilevu. Interactive interfaces for capturing and annotating videos of human movement, Jul 2019. URL http://hdl.handle.net/10919/91424.

[35] Janice L. Zimbelman, Stephen P. Juraschek, Xiaoming Zhang, and Vernon W.-H. Lin. Physical therapy workforce in the united states: Forecasting nationwide shortages. *PM&R*, 2(11):1021–1029, 2010. doi: https://doi.org/10.1016/j.pmrj.2010.06.015. URL https://onlinelibrary.wiley.com/doi/abs/10.1016/j.pmrj.2010.06.015.