Dissertation Thesis

Czech
Technical
University
in Prague

F7

Faculty of Biomedical Engineering
Department of Natural Sciences

Modular Motor Rehabilitation System Using Computer Vision, Augmented and Virtual Reality

Ing. Jindřich Adolf

Supervisor: Assoc. prof. Lenka Lhotská, PhD.
Supervisor–specialist: Ing. Jaromír Doležal, PhD.
June 2024

Acknowledgements

First of all, I would like to express my deep gratitude to all those who have supported me during my Ph.D. studies. My heartfelt thanks go to my supervisors for their invaluable advice and guidance. I would also like to express my appreciation and thanks to FBMI and CIIRC, which provided me with the environment for creative work and the opportunity to establish multidisciplinary and international collaborations. I especially appreciate the opportunity to spend part of my study time in foreign workplaces, which helped me to gain the necessary experience for conducting research. My sincere thanks to all my fellow physiotherapists, students, and volunteers. Without their involvement, commitment, and collaboration we would not have been able to achieve the practical results that this thesis presents.

Declaration

I hereby declare that this dissertation represents my original work and contributions. No portion of this work has been submitted in support of an application for any other degree or qualification at this or any other institution. Where I have drawn upon the works, ideas, figures, or any other material from others, I have appropriately cited or quoted the sources.

All research associated with this dissertation was carried out following the ethical standards and guidelines of Czech Technical University in Prague. Any required permissions or consents, where applicable, have been obtained. Prague, June 3, 2024.

Abstract

The primary goal of this dissertation was to develop a novel methodology for capturing human movement during physiotherapy sessions using a standard RGB camera. This innovation addresses the urgent need for accessible and cost-effective motion capture technologies in clinical settings and home environments. The dissertation provides a review of current human motion capture methods, with a particular focus on the use of conventional cameras and the application of advanced computer vision techniques for motion analysis.

An essential element of this research involved practical experimentation, which included extensive measurements and empirical investigations, as well as comparisons between camera systems and virtual reality motion capture systems.

A critical aspect of this dissertation was to base the research on multidisciplinary collaboration with professional physiotherapists. Their expertise was invaluable in aligning technological development with practical clinical needs and ensuring the relevancy and applicability of the motion capture methodology within therapeutic settings.

The dissertation thus summarizes and systematically describes all the experiments, conference papers and three peer-reviewed journal publications into one whole, which is concluded with a separate chapter devoted to the methodology of camera-based systems for human motion capture systems.

Keywords: Camera-based mocap, telerehabilitation, remote physical therapy, physiotherapy, functional tests, computer vision. machine learning.

Supervisor: Assoc. prof. Lenka Lhotská, PhD.

Abstrakt

Hlavním cílem této disertační práce bylo vyvinout metodiku pro snímání lidského pohybu v rámci fyzioterapeutických cvičení s použitím jedné RGB kamery jako snímacího zařízení. Tento přístup řeší potřebu cenově dostupnějšího a efektivnějšího snímání lidského pohybu v klinickém I domácím prostředí. Tato práce podává přehled současných metod snímání lidského pohybu, s důrazem na použití běžných kamer a aplikaci technik počítačového vidění pro analýzu pohybu.

Významnou součástí tohoto výzkumu byla praktická práce, která zahrnovala rozsáhlá měření a empirická šetření, na základě kterých bylo možné popsat možnosti a limity této metody a porovnat je s metodami snímání pohybu ve virtuální realitě. Práce se zaměřuje především na aplikovatelnost v reálných fyzioterapeutických scénářích.

Jedním z hlavních přístupů tohoto výzkumu je multidisciplinární spolupráce s fyzioterapeuty, která umožnila spojit jak praktické zkušenosti odborníků v praxi, tak technický vývoj v oblasti počítačových věd.

Disertační práce tak shrnuje a systematicky popisuje veškeré provedené experimenty, konferenční příspěvky a tři recenzované časopisecké publikace do jednoho celku, který je ukončený samostatnou kapitolou, která se věnuje metodice snímání pomocí kamerového systému.

Klíčová slova: Snímání pohybu kamerou, telerehabilitace, vzdálená fyzická rehabilitace, fyzioterapie, funkční testy, počítačové vidění, machine learning.

Překlad názvu: Modulární systém pohybové rehabilitace využívající počítačové vidění, rozšířenou a virtuální realitu

Contents

List of Abbreviations 1

1 State-of-the-art 3
1.1 Motivation . . . 4
1.2 Standardized Procedures for Physiotherapy Treatment . . . 5
1.3 Physical Telerehabilitation . . . 5
1.3.1 Current Challenges of Physical Telerehabilitation . . . 6
1.4 Motion Capture Systems . . . 6
1.4.1 Marker-Based Motion Capture systems . . . 7
1.4.2 Wearable Sensor-based Motion Capture systems . . . 10
1.4.3 Virtual Reality Motion Capture Systems . . . 12
1.4.4 Markerless Systems . . . 13
1.5 Computer Vision Based Motion Capture systems . . . 13
1.5.1 Optical Flow . . . 13
1.5.2 RGB-Depth Sensors . . . 14
1.5.3 Pose Estimation . . . 15

2 Aim of the Thesis 19
2.1 Aim of the Thesis . . . 20

3 Preliminary Research and Experiments 23
3.1 Juggling in VR: Advantages of Immersive Virtual Reality in Juggling Learning . . . 24
3.1.1 Description of the Study . . . 24
3.1.2 Context of the Dissertation . . . 25
3.2 Affordable Personalized, Immersive VR Motor Rehabilitation System with Full Body Tracking . . . 25
3.2.1 Description of the Approach . . . 25
3.2.2 Usability Analysis in Dissertation Framework . . . 26
3.3 Automatic Telerehabilitation System in a Home Environment Using Computer Vision . . . 26
3.4 Preliminary Motion Capture Analysis with Vicon . . . 27
3.5 Remote Physical Therapy: Requirements for a Single RGB Camera Motion Sensing . . . 27
3.6 Upper Limb Range of Motion Evaluation by a Camera-Based System . . . 28
3.7 Detection of Signs of Fatigue in Functional Tests in Healthy Population by OpenPose . . . 29
3.8 Software for Quantitative Assessment of Movement Activity . . . 29
3.8.1 Video Record Processing . . . 29
3.8.2 Interpretation of Stored Data 31

4 Single Camera-Based Remote Physical Therapy: Verification on a Large Video Dataset 33
4.1 Introductionary Comments . . . 34
4.2 Abstract . . . 34
4.3 Introduction . . . 35
Related Work . . . 36
4.4 Methods . . . 37
4.4.1 Design . . . 37
4.4.2 Video Analysis . . . 38
4.4.3 Keypoint Confidence Extraction . . . 39
4.4.4 Statistical Analysis . . . 41
4.5 Results . . . 42
4.6 Discussion . . . 46
4.7 Conclusions . . . 48

5 Evaluation of Functional Tests Performance Using a Camera-based and Machine Learning Approach 49
5.1 Introductionary Comments . . . 50
5.2 Abstract . . . 50
Introduction . . . 51
Methods . . . 53
Participants . . . 54
Study Design . . . 54
Functional Test Parameters and Description . . . 55
Motion Capture Setup . . . 55
Expert Visual Analysis . . . 56
Signal Processing and Signal Extraction . . . 57
Classification . . . 59
Results . . . 60
Agreement of Expert Visual Analysis . . . 61
Results of Classification . . . 62

Discussion 63
Conclusion 64

6 OffiStretch: Camera-based Real-time Feedback for Daily Stretching Exercises 67

6.1 Introductionary Comments 68
6.2 Introduction 68
6.3 Related work 69
6.3.1 Body Tracking 70
6.3.2 Visual Feedback for Physical Training 71
6.4 The OffiStretch System 72
6.4.1 Body Tracking and Pose Assessment Features 72
6.4.2 Exercise Instruction Authoring 75
6.4.3 OffiStretch Visual Feedback . 76
6.4.4 Hardware and Software Requirements 77
6.5 Online Survey: Stretching in the (Home) Office 77
6.5.1 Online survey participants . . 78
6.5.2 Online survey results 78
6.6 User Study 80
6.6.1 Study Design 81
6.6.2 Study Procedure 81
6.6.3 Selected Exercises 82
6.6.4 Participants 82
6.6.5 Signal Processing 83
6.6.6 Results 84
6.7 Discussion 88
6.7.1 Limitations and Future Work 90
6.8 Conclusion 91

7 Methodology for a Single Camera-based Human Motion Capture in Physical Telerehabilitation 93

7.1 Goal Settings and Purpose Design 94
7.2 Multiple camera setup 95
7.3 Single Camera-based Approach Limits 95
7.3.1 Feature Extraction 97
7.4 Camera Requirements 98
7.5 Environment Setup 98
7.6 Real-Time vs. Offline Processing: A Comparative Overview 100
7.7 Keypoint Extraction 101
7.7.1 Available Models Used for Extraction 102
7.7.2 Extracted Keypoints vs Biomechanical Keypoints 103
7.8 Selecting HW for Processing . . 103

8 Disscussion 105

8.1 Addressing the Research Questions 106
8.1.1 What are the practical differences, advantages, and disadvantages between using virtual reality and a camera-based system for motion capture? 106
8.1.2 How well does the camera-based motion capture detection work? 106
8.1.3 What are the limits of capturing motion with this camera approach? 107
8.1.4 For which exercises or movements is this approach suitable? 107
8.1.5 How demanding is the camera-based system on computational performance? . . . 107
8.1.6 What could be the clinical applications of this camera-based motion capture? 108
8.2 Achievement of Work Objectives 108
8.2.1 Challanges of Virtual Reality in Rehabilitation 109
8.2.2 Concept of Telerehabilitation Using a Camera-based Approach 110
8.2.3 Verify Detection Functionality Using a Large Video Database . 110
8.2.4 Feasibility of Building Machine Learning Models 110
8.2.5 Multidisciplinary Research Connection 111
8.2.6 Real-time Application 111
8.3 Future Work 111

9 Contribution to the field 115

9.1 Theoretical Contributions 116
9.2 Practical Contributions 116

10 Conclusion **119**

Bibliography **121**

11 List of Publications **137**

11.1 Publications Related to the Dissertation 138

11.1.1 Journal publications 138

11.1.2 Conference Publications . . 138

11.2 Other Publications 139

A Acknowledgements of Contribution and Authorship **141**

Figures

1.1 Confidence maps for keypoint "right shoulder", taken from the original OpenPose publication[1]. . 16
1.2 Affinity filed for keypoints "right shoulder" and "right elbow", taken from the original OpenPose publication[1]. 17

3.1 A person wearing several HTC Vive trackers on moving joints, the rest of the human skeleton is computed using inverse kinematics. 25
3.2 Rendered skeleton into a standard RGB image of the test subjects. . . 26
3.3 A woman captured in mid-jump, adorned with reflective sensors, as recorded by the Vicon system at CIIRC, ČVUT. 27
3.4 In one of the experiments, we evaluated the camera's detection quality across various resolutions to determine the optimal setting. 28
3.5 Software For Quantitative Assessment Of Movement Activity 30

4.1 Study design—diagram explaining the sequence of steps. 37
4.2 Orientation of the camera relative to the subject 39
4.3 OpenPose 25 keypoints model . . 40
4.4 Confidence of detection of selected keypoints front view, upper body. . 43
4.5 Confidence of detection of selected KP (keypoints)—front view, lower body. 44
4.6 Confidence of detection of selected KP (keypoints)—¾ view, upper body. 44

5.1 Flowchart of the study design . . 54
5.2 OpenPose model drawn in the image of a person 56

6.1 Screenshot of OffiStretch application with real-time dynamic feedback drawn onto trainee's own body. The arrows encourage the trainee to extend the stance and the green circle encourages greater flexion at the knee joint. The closer the practitioner is to the desired position, the smaller the circle or thinner the line is. 72
6.2 A model showing 17 key points that we use to calculate features characterizing human posture. We use the same keypoint indexing as the original 25-keypoint OpenPose model [1] from which our model is derived. 73
6.3 Three types of features used in our pose assessment. 75
6.4 Responses to Q1 reflect how often participants perform stretching exercises, Q2 indicates the preferred frequency of stretching with an imaginary coaching system and Q3 reveals participants' willingness to use a coaching system that reminds and instructs them to do stretching. The home-office scenario is presented in blue (left side) and dedicated workplace in red (right side). For more details see Table 6.7. 79
6.5 Reference position for comparison with the trainee. A set of six exercises (performed by each participant twice; once with feedback and once without feedback.) 83
6.6 Values of metrics determining error of trainees' poses with respect to reference poses. Metrics were weighted based on qualitative assessment of professional physiotherapists. Condition with visual feedback is displayed in blue and condition without feedback is shown in red. 85

6.7 Data collected from on-site questionnaire. The condition with our visual feedback is indicated in blue and the condition without feedback is indicated in red. Higher value on y axis means more positive response for a given factor. (A) Understanding of the instructions/visualization. (B) Helpfulness of guidance. (C) Subjective performance. (D) Motivation. (E) Preference. 88

7.1 Workflow of Camera-Based Motion Capture System for rehabilitation. 94
7.2 Comparison of Side and from view. 95
7.3 Illumination adjustment in an RGB image using variable Gamma correction parameters. 99
7.4 The basic concept of Keypoint extraction 102

Tables

4.1 Number of videos in each category. Subcategories are: lying down (LY); on the knees (Kn); sitting (Si); standing (St). 38
4.2 Categorization of videos. 39
4.3 The 12 basic keypoints of the appendicular skeleton. 40
4.4 Median confidence value for each specific group. 42

5.1 List of selected ranges of movement 59
5.2 Agreement of the expert ratings of the three physiotherapists 61
5.3 Summary of the AdaBoost Classification for the Two Exercises SDT and SLST 63

6.1 Questions of our online survey, used to evaluate our two hypotheses. Each question was answered twice (once for home office condition and once for dedicated workplace condition). Answers were listed in the opposite order in the questionnaire and we inverted them for consistency of visualization within the publication. 77
6.2 The results of significance assessment by Wilcoxon signed-rank test. The significance of differences between home office and dedicated workplace conditions was assessed for each question from Table 6.7. 80
6.3 For each exercise, a unique combination of features and feedback elements was experimentally selected. The numerical values correspond to the keypoints in Fig. 6.2. 82
6.4 Statistical significance of differences between conditions with and without visual feedback for each exercise. The results were calculated using Wilcoxon signed rank test. . . 85

6.5 Based on the ex-post monitoring of the recordings with the subjects, professionals in the field of physiotherapy determined a weight for each feedback element. Thus, the table always displays the name of the exercise on a row and a list of feedback elements and their respective weights in a second column. 87
6.6 Results of on-site study questionnaire - statistical significance of differences in responses after the exercises with and without feedback. Statistical significance was assessed using Wilcoxon signed-rank test. . 88
6.7 Questions of our online survey, used to evaluate our two hypotheses. Each question was answered twice (once for home office condition and once for dedicated workplace condition). Answers were listed in the opposite order in the questionnaire and we inverted them for consistency of visualization within the publication. 89

7.1 Comparison between Real-Time and Offline Processing in Motion Capture. 101

List of Abbreviations

AI	Artificial Intelligence
BGU	Ben Gurion University
CNN	Convolutional Neural Networks
COCO	Common Objects in Context
CTU	Czech Technical University in Prague
CUDA	Compute Unified Device Architecture
CV	Computer Vision
ČVUT	České Vysoké Učení Technické v Praze
ER-WCPT	European Region of the World Confederation for Physical Therapy
FPS	Frame per Second
GPU	Graphical processing Unit
ICC	Intraclass correlation coefficients
IMU	Inertial Measurement Unit
MCC	Motion Capture Camera
ML	Machine Learning
MOCAP	Motion Capture
MPII	Max Planck Institute Informatics
NN	Neural Networks
OP	OpenPose
PAF	Part Affinity Fields
PT	Physiotherapist
R-CNN	Region-Based Convolutional Neural Network
RGB	Red, Green, Blue
ROM	Range of Movement
SDT	Step Down Test
SLST	Single Leg Squat
UK FTVS	Charles University, Faculty of Physical Education and Sport
VR	Virtual Reality
WC	World Congress
YOLO	You Only Look Once

Chapter 1

State-of-the-art

1.1 Motivation

In the contemporary world, constantly and rapidly evolving, we witness a significant shift from manual labor to professions predominantly requiring a sedentary lifestyle. This trend is a direct consequence of technological advancements over the past decades, which have led to the replacement of physical labor with automated and robotic systems. Since physical activity is no longer an inherent part of many job roles, it is imperative for individuals to actively make an effort to maintain their physical fitness outside of the work sphere. Otherwise, we risk developing musculoskeletal disorders. These impacts are elaborated in the study titled 'Global Burden of Musculoskeletal Disorders and Attributable Factors' [2] highlights the increase in these problems on a global scale.

The increasing number of people needing physical therapy, combined with higher treatment costs, puts significant pressure on the medical community. This scenario underscores the imperative to develop innovative solutions, as underscored by the studies [3] and [2]. This change requires us to include physical activity in our daily routine, which is vital for keeping us healthy.

However, this "sedentary" [4] era is also witnessing exponential growth in technology. In the last decade, we have witnessed a surge in computing power and the development of technologies such as virtual reality, augmented reality, and advanced computer vision models that recognize the human posture with unprecedented accuracy.

It is therefore logical to ask: can these technologies be used to address health problems associated with increased sedentary behavior? Can telerehabilitation become a tool that bridges the gap between modern lifestyles and the need for physical rehabilitation? Research has shown that one of the reasons why patients do not exercise at home is the lack of feedback and control [5].

For the patient to receive accurate and personalized feedback, the use of a motion capture system that can reliably describe the patient's movement is essential.

If this technology could be used well enough to capture patient movement in the home environment, it would greatly improve the efficiency of the system for treating movement disorders. This is primarily because such sensing is based on a single RGB camera, which is currently available on almost every device at a price so low that anyone can afford it. Thus, the motivation for this research is to describe the current methods, define them, verify their advantages and shortcomings through practical experiments and studies, and thus develop a methodology for modern motion capture using any RGB camera.

My goal is thus to approach the problem from a completely new perspective, to build and investigate a system that would solve the problems described above. I believe that the combination of current computer vision methods and current knowledge in physiotherapy can lead to better outcomes in the field of physical rehabilitation, which is the main motivation of this research.

In this chapter, it is essential to address the current trends in home rehabilitation. This involves discussing which methods are effectively in use and showing practical applicability. Additionally, reviewing today's advanced motion capture systems is necessary. A dedicated subsection will also explore the current state of Computer Vision systems.

1.2 Standardized Procedures for Physiotherapy Treatment

Many countries have national physiotherapy or physical therapy associations that provide guidelines and standardized procedures for their members. In the EU, it is the European Region of the World Confederation for Physical Therapy (ER-WCPT) [1]. In the Czech Republic, it is Union of Physical Therapists of the Czech Republic [2] In the realm of physiotherapy, standardized procedures refer to a set of established protocols and guidelines that ensure consistent, evidence-based care for patients across various clinical settings. These procedures are developed based on rigorous research and clinical trials, aiming to provide optimal outcomes for patients with diverse musculoskeletal, neurological, or cardiopulmonary conditions. A crucial component of these standardized procedures is the emphasis on continuity of care, which often includes guidelines for patients to continue specific exercises at home. This ensures sustained progress and rehabilitation even outside the clinical environment.

1.3 Physical Telerehabilitation

Telerehabilitation can be defined as the service of the delivery of rehabilitation services over telecommunication networks and the Internet[6]. This method gained popularity, especially during the global COVID-19 pandemic [7], which lasted from January 30, 2020, to May 5, 2023. Saaei's study [8] conducted with practical physiotherapists and patients shows that new modern approaches are needed. There's no doubt about the advantages and effectiveness of remote physical rehabilitation. In their systematic review, Seron et al.[9] considered fifty-three reviews. They concluded that telerehabilitation is as effective as in-person rehabilitation or even better in the absence of any rehabilitation for conditions such as osteoarthritis, low-back pain, hip and knee replacements, and multiple sclerosis.

Currently, telerehabilitation is still conducted mostly with the remote presence of a therapist, often through video calls [10] alternatively, they function as systems that are capable of playing personalized videos with a trainer[11].

[1] https://www.erwcpt.eu/
[2] https://www.unify-cr.cz/

1.3.1 Current Challenges of Physical Telerehabilitation

Technology Barriers

The common problem of widespread telerehabilitation is a lack of technical knowledge. In a nationwide survey in 2020 [12], only 58.8% of physiotherapists reported having sufficient knowledge about telerehabilitation. Despite the potential advantages of telerehabilitation, its actual implementation and usage in physical therapy settings remained limited. The primary barriers identified were technical issues, staff skills, and the associated high costs.

Similarly, patients also face challenges when it comes to technological proficiency. Many find it difficult to navigate and use state-of-the-art technological approaches for their rehabilitation. This further exacerbates the problem, as not only do therapists face barriers in implementing telerehabilitation, but patients themselves also encounter hurdles in accessing and effectively using these platforms.

For telerehabilitation to reach its full potential and benefit a broader spectrum of patients, the design of future systems must prioritize user-friendliness. These platforms should be as intuitive and straightforward as possible, minimizing the technological barriers for both therapists and patients alike. This would ensure a smoother transition to digital platforms, enhancing patient engagement and optimizing the benefits of telerehabilitation.

Assessment Limitations

One of the challenges of telerehabilitation is the difficulty in assessment and the inability to observe certain aspects up close or physically interact with the patient [13]. While remote technology cannot replace physical contact, it can, thanks to the objective measurement of certain movements, technically enable at least a partial evaluation of quality.

Patient Engagement and Adherence

Whether it's rehabilitation or telerehabilitation, a common problem is the lack of motivation [14]. The study [15] suggests that feedback and progress monitoring can boost motivation. Unlike traditional rehabilitation, telerehabilitation offers much broader possibilities, primarily due to the integration of technology. Therefore, systems can be designed to track a patient's progress, and provide stronger feedback.

1.4 Motion Capture Systems

To assess the quality of movement execution, we cannot avoid needing a reliable motion capture system. The aim of this section is to describe the current state-of-the-art in motion sensing, both from a general perspective and considering its suitability for remote telerehabilitation. I don't review these systems merely based on their accuracy, but more on their applicability for the intended purpose. For our objective, it's appropriate to categorize these systems into three parts: marker-based systems, marker-less systems, and purely camera-based systems. The entire subsequent section will focus on these, given their relevance to this thesis.

1.4.1 Marker-Based Motion Capture systems

Marker-Based Motion Capture systems are the most accurate and widely used technologies in motion analysis. These systems function by placing reflective or emissive markers on specific anatomical landmarks of an individual or object. As the individual moves, multiple cameras placed strategically around the area capture the motion of these markers in real-time. By analyzing these markers' spatial positions and trajectories, a precise representation of the individual's movement can be reconstructed in a digital environment. Renowned for their precision, these systems are instrumental in fields ranging from biomechanics and sports science to film production and video game design. While they offer unparalleled accuracy, their primary limitation is the need for a controlled environment and the potential for marker occlusion, especially in complex movement scenarios." The famous ones are Vicon. [3], OptiTrack [4] and Qualisys. [5]

By capturing the position of these markers from multiple views, the system triangulates the 3D position of each marker in space.

A simple triangulation can be illustrated with two cameras. If you know the position and orientation of each camera and the position of the marker in each camera's image plane, you can draw a ray from the camera through the detected marker position. The point where these rays from the two cameras intersect is the 3D position of the marker.

Triangulation is the process of determining the 3D point $\mathbf{X}$ using the distorted image points $\mathbf{x}'_i$ from multiple cameras and their corresponding projection matrices P_i [16].

$$\mathbf{x}_1 = P_1\mathbf{X} \tag{1.1}$$

$$\mathbf{x}_2 = P_2\mathbf{X} \tag{1.2}$$

Triangulation, especially in the context of real-world systems, does require accounting for camera distortion [17].

The distorted image coordinates, x' and y', are related to the normalized image coordinates, x and y, by:

$$x' = x(1 + k_1r^2 + k_2r^4 + k_3r^6) + (2p_1xy + p_2(r^2 + 2x^2)) \tag{1.3}$$

$$y' = y(1 + k_1r^2 + k_2r^4 + k_3r^6) + (p_1(r^2 + 2y^2) + 2p_2xy) \tag{1.4}$$

where:

[3] https://www.vicon.com/about-us/
[4] https://www.optitrack.com
[5] https://www.qualisys.com

- (x, y) are the normalized coordinates in the image plane before distortion.
- (x', y') are the distorted coordinates.
- $r^2 = x^2 + y^2$.
- k_1, k_2, k_3 are radial distortion coefficients.
- p_1, p_2 are tangential distortion coefficients.

Once the distortion is accounted for, the corrected 2D points in the image plane can then be related to the 3D world points through the camera projection matrix:

$$\mathbf{x}'_i = P_i\mathbf{X} \tag{1.5}$$

where:

- $\mathbf{x}'_i$ is the distorted point in camera i.
- P_i is the projection matrix for camera i.
- $\mathbf{X}$ is the 3D position in homogeneous coordinates.

Advantages of Marker-Based Motion Capture Systems

1. **High Precision:** Marker-based systems often offer unparalleled accuracy due to the distinct markers that can be tracked individually, ensuring precise movement reconstruction. Eichelberger in his study [18] examined the accuracy of Vicon in biomechanical use, considering factors like the number of cameras, measurement location, and movement dynamics. During lower-body assessments of level walking, trueness for marker distances was between (-0.38,0.38) mm with 10 cameras. Uncertainty varied by region: 0.33 mm for the foot, 0.74 mm for the knee, and 1.25 mm for the hip. The comparison of the most advanced Marker Mocap systems is done by Topley in his study [19]. The result shows, that the accuracy of all those systems for biomechanical use is more than sufficient.

2. **Reliability:** The use of physical markers typically yields consistent data, reducing the chance of system errors or anomalies. This study [20] introduces an anthropometric measurement method using a motion capture camera (MCC) to create a database for young males. After a pilot test to confirm the procedure, they validated the data's accuracy, bias, reliability, and precision. The results confirmed the camera's reliability and successfully established anthropometric data for young male participants.

3. **Mature Technology:** Given their longstanding use in various industries, these systems have undergone extensive development and refinement, benefiting from years of user feedback and technological advancements.Vicon[6] is a thriving company with roots dating back to 1979 when its original product was introduced by Oxford Medical Systems. After a management buyout of Oxford Dynamics in 1984, it embarked on a 35-year journey under the Oxford Metrics banner. The Vicon system, renowned for its versatility, has found applications in various fields. These include motion capture for films and video games, biomechanical analysis for sports and rehabilitation, clinical gait analysis for medical purposes, and development in virtual and augmented reality. Additionally, it's utilized in tracking the movement of robots and drones, studying animal motion, conducting ergonomic workplace studies, researching human-machine interactions, advancing animation techniques, and in the design and testing of prosthetics and orthotics.

4. **Comprehensive Data Analysis:** The data captured allows for a detailed analysis of movement, including aspects like velocity, trajectory, and angles, which are crucial in fields like biomechanics. Vicon SW package Nexus [7] is the market's leading all-inclusive tool for movement analysis, designed for the life sciences community and its use is a real golden standard in the field of biomechanics.

5. **Compatibility:** Many marker-based systems are compatible with various software solutions for post-processing, making it easier for researchers and professionals to work with the acquired data. Vicon, Optitrack, and Qualisys all have an integration into Matlab [8] or Python[9].

Disadvantages of Marker-Based Motion Capture Systems

1. **Setup Time:** Placing markers accurately on an individual or object can be time-consuming, especially when capturing complex movements.

2. **Occlusion:** There's a possibility of markers being obscured from the view of some cameras during movement, leading to potential data loss or inaccuracies.

3. **Environmental Constraints:** These systems often require controlled environments, free from external light interferences, which might limit their application in outdoor or uncontrolled settings.

4. **Cost:** High-quality marker-based motion capture systems can be expensive, encompassing not only the cameras but also the specialized markers, calibration tools, and software.

[6] https://www.vicon.com/about-us/
[7] https://www.vicon.com/software/nexus/
[8] https://www.mathworks.com/
[9] https://www.python.org/

5. **Limitations in Natural Movement:** The presence of markers and, in some cases, the suits worn can slightly alter the natural movements of the subject. This can be a concern when studying subtle or fluid motions.

1.4.2 Wearable Sensor-based Motion Capture systems

Inertial Sensors

These, often found in Inertial Measurement Units (IMUs), combine accelerometers, gyroscopes, and sometimes magnetometers to measure acceleration, angular rate, and magnetic fields. They can determine changes in position and orientation by integrating their measurements over time. An example of this technology integrated in a suit is Rokoko[10], DorsaVi [11] or XSens [12]. This study compares the accuracy between Xsens and dorsaVi. [21]. This study compares Xsens vs Vicon, where we can find an error in angles between 0.7 and 14.5 deg [22].

Inertial motion capture uses Inertial Measurement Units (IMUs) containing a 3-axis accelerometer, a 3-axis gyroscope, and sometimes a 3-axis magnetometer. The fundamental equations related to IMUs are:

Inertial motion capture systems rely on Inertial Measurement Units (IMUs) to track and analyze motion. Each IMU generally comprises:

Acceleration:
Given by the accelerometer, it measures linear acceleration. After subtracting gravity (when stationary, the accelerometer will still detect the acceleration due to gravity), the double integration of acceleration gives the change in position.

$$\Delta p = \int \int (a - g)\, dt^2 \tag{1.6}$$

Where:

- Δp is the change in position.
- a is the acceleration measured by the accelerometer.
- g is the acceleration due to gravity.

Angular Velocity:
Given by the gyroscope, measures the rate of change of the angular position (often in radians/second). Integration of angular velocity provides a change in orientation.

$$\Delta\theta = \int \omega\, dt \tag{1.7}$$

Where:

- $\Delta\theta$ is the change in orientation (angular position).

[10] https://www.rokoko.com/
[11] https://www.dorsavi.com/
[12] https://www.xsens.com/

- ω is the angular velocity measured by the gyroscope.

Magnetic Field:
The magnetometer provides data that can be used to determine heading relative to the Earth's magnetic north. The magnetometer reads the magnetic field in 3D space as:

$$B = [B_x, B_y, B_z] \tag{1.8}$$

Where:

- B is the magnetic field vector.
- B_x B_y B_z are the components in a three-dimensional Cartesian coordinate system.

From the IMU data, one can derive the orientation using methods like the complementary filter, Kalman filter, or quaternion-based algorithms [23] to fuse the data from the accelerometer, gyroscope, and optionally the magnetometer.

For motion capture, multiple IMUs are attached to various parts of the body, and the data from each IMU is fused to get a holistic understanding of body motion.

By attaching multiple IMUs to various parts of the body, it becomes possible to capture detailed motion data. This information is essential for producing an accurate and holistic understanding of body movements.

However, one challenge with IMUs is the drift [24] over time, mainly due to the double integration of acceleration data and the integration of angular velocity. To counteract this drift and enhance system accuracy, various sensor fusion techniques are often employed. These techniques might utilize complementary filters, Kalman filters, or quaternion-based algorithms to merge data from the accelerometer, gyroscope, and magnetometer effectively.

Stretch and Bend Sensors

Those sensors are embedded into clothing or worn as bands, these sensors change their resistance [25] or capacitance [26] based on their deformation, enabling the capture of joint angles and body pose. For these sensors, there are some common challenges, such as a complex setup and expensive monitoring equipment, which prevent widespread adoption and accessibility for all patients.

Resistive Sensors: These sensors work based on the principle that the resistance of a conductor changes when it's stretched or compressed. The change in resistance is proportional to the strain experienced by the material.

$$\Delta R = R_0 \times (1 + k \times \epsilon) \tag{1.9}$$

Where:

- ΔR is the change in resistance.

- R_0 is the initial resistance.
- k is the gauge factor of the sensor material.
- ϵ is the strain, which is defined as the change in length divided by the original length.

Capacitive Sensors: These sensors operate based on the principle that the capacitance of a parallel plate capacitor changes when the distance between the plates is altered (as with bending) or when the area of the plates changes (as with stretching).

$$C = \frac{\varepsilon_0 \varepsilon_r A}{d} \tag{1.10}$$

Where:

- C is the capacitance.
- ε_0 is the permittivity of free space.
- ε_r is the relative permittivity (dielectric constant) of the material between the plates.
- A is the area of one of the plates.
- d is the distance between the plates.

These sensors convert mechanical deformation into electrical signals, which can then be interpreted and used in various applications, from health monitoring wearables to robotic control.

1.4.3 Virtual Reality Motion Capture Systems

Virtual reality motion capture systems employ a combination of inertial sensors, optical markers, and sophisticated algorithms to accurately capture and interpret human motion [27]. Inertial sensors, such as accelerometers and gyroscopes, are often embedded in wearable devices to track movement and orientation without the need for external cameras. Optical systems, on the other hand, use cameras to detect specially designed markers placed on the user's body, providing precise spatial data by triangulating the positions of these markers. The collected data is processed using advanced algorithms that filter noise and compute the kinematics of the human body. This allows the system to deliver real-time feedback and interaction within the virtual environment, making it essential for creating fluid and realistic user experiences in VR. These technical solutions collectively ensure high fidelity in motion capture, crucial for applications demanding accurate and responsive movement replication. One of the most commonly utilized systems today is the HTC Vive[13], which operates with base stations. The HTC Vive's base stations, known as Lighthouse stations, employ a combination of infrared LEDs and

[13] https://www.vive.com/us/

rotating laser emitters for precise room-scale tracking. Each base station issues a synchronization flash from its LEDs, followed by sequential horizontal and vertical laser sweeps. Sensors on the Vive headset and controllers capture the exact timing of these laser hits relative to the LED flash, enabling the system to accurately triangulate their positions within the play area. For motion capture, we utilize a set of Vive Trackers[14] placed on the human body and reconstruct the full body using inverse kinematics [28].

The Voxs study [29] compares a commercial VR tracking sensor system (HTC Vive tracker combined with an inverse kinematic model, Final IK [15]) with a marker-based optical motion capture system Qualisys[16], the gold standard for motion analysis, to evaluate their accuracy in measuring joint angles for ergonomic assessments. Results indicate that while the HTC Vive system has potential for mapping joint angles, it shows significant deviations in accuracy ($\pm 6°$ to $\pm 42°$) compared to Qualisys, highlighting the need for improvements to reduce systematic errors in ergonomic evaluations.

1.4.4 Markerless Systems

In the rapidly evolving field of motion capture technology, markerless systems have gained significant attention. Prominent examples include the Microsoft Kinect[17], Leap Motion[18], and Intel's RealSense[19]. While it is essential to acknowledge these systems as potential alternatives to our approach, their reliance on specialized hardware does not fully align with the specific requirements of our application. This discrepancy underscores the need for a solution that balances technical capability with practical applicability in diverse settings.

1.5 Computer Vision Based Motion Capture systems

1.5.1 Optical Flow

This method computes the motion vector of each pixel in the frame, giving a dense flow map. It's particularly useful for capturing subtle or complex movements [30].

1. **Brightness Constancy Assumption:** The brightness (or intensity) of a particular point in an image remains constant over time [31].

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t) \tag{1.11}$$

[14] https://www.vive.com/eu/accessory/tracker3/
[15] https://assetstore.unity.com/packages/tools/animation/final-ik-14290
[16] https://www.qualisys.com/
[17] https://learn.microsoft.com/en-us/windows/apps/design/devices/kinect-for-windows
[18] https://www.ultraleap.com/
[19] https://www.intelrealsense.com/

2. **Taylor Series Expansion:** The brightness constancy equation can be expanded using the Taylor series.

$$I(x+\delta x, y+\delta y, t+\delta t) \approx I(x,y,t) + \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t \tag{1.12}$$

3. **Optical Flow Equations:** The optical flow equation can be derived as:

$$\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} = 0 \tag{1.13}$$

   Where u and v are the horizontal and vertical components of the optical flow (velocity) respectively.

4. **Lucas-Kanade Method:** The flow is essentially constant in a local neighborhood of the pixel under consideration, and then the least squares criterion to solve for the flow parameters [32].

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \Sigma I_x^2 & \Sigma I_x I_y \\ \Sigma I_x I_y & \Sigma I_y^2 \end{bmatrix}^{-1} \begin{bmatrix} -\Sigma I_x I_t \\ -\Sigma I_y I_t \end{bmatrix} \tag{1.14}$$

   Where I_x, I_y, and I_t are the spatial and temporal image gradients.

5. **Horn and Schunck Method:** This method introduces a global constraint of smoothness to estimate optical flow [33].

$$E(u,v) = \int\int [(I_x u + I_y v + I_t)^2 + \alpha^2(|\nabla u|^2 + |\nabla v|^2)]\, dx\, dy \tag{1.15}$$

   Where α is a regularization parameter.

1.5.2 RGB-Depth Sensors

The Microsoft Kinect, launched in 2010, was a pioneer in using RGB-Depth (RGB-D) technology for motion capture. It combined a standard camera with an infrared (IR) depth sensor to track 3D space and movement without needing physical markers, marking a big step forward for computer vision. Kinect measured depth by shining a pattern of IR light and then analyzing how this pattern changed when it bounced off objects. This method allowed it to figure out how far away things were, helping to turn flat images into 3D models. This technology was important not just for games but also for physical therapy, as it helped track movements in a non-invasive way, crucial for rehab exercises. Research, like the studies reviewed by Hondori[34], showed Kinect's value in healthcare, proving it was useful and impactful in various settings.

While Kinect's implementation of RGB-D technology represented a significant leap forward, its reliance on specialized hardware, including an infrared (IR) depth sensor and structured light projector, poses limitations for scalability and universal application. With the discontinuation of Kinect in 2017 and the desire for more versatile and widely deployable solutions, the focus has shifted towards leveraging standard cameras, which are more ubiquitous and can be integrated into a vast array of devices, from smartphones to conventional computing systems.

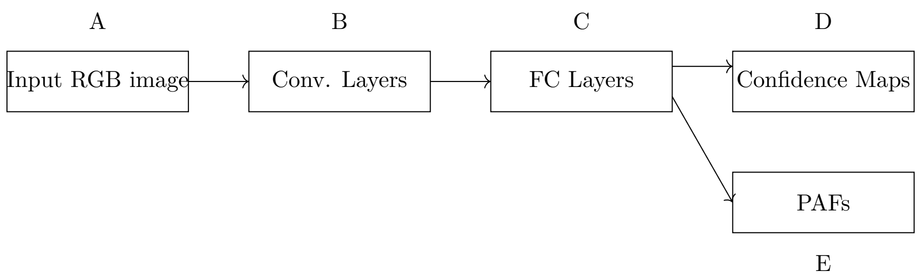
The transition to using standard cameras for motion capture and depth sensing reflects a broader trend in computer vision towards software-based solutions that can interpret depth and motion from conventional RGB video feeds.

1.5.3 Pose Estimation

By identifying and analyzing the structure of an object (like a human body), computer vision techniques can estimate its pose. This involves determining the position of each body part relative to others. This work is primarily based on the concept published in[35], introduced alongside a GitHub repository releasing open-source code for developers. This software utilizes two datasets for training: MPII [36] and COCO [37]. This software is widely used globally for human skeleton detection in various applications such as people counting, human detection from autonomous vehicles, and more. Another widely adopted practical tool is Google's MediaPipe [38], which is also based on the COCO dataset.

Recently, there has been a surge in the development of similar libraries and modified models. Examples include FastNet [39], AlphaPose[40], YOLO-Pose[41] and others. A comparison of these current solutions is presented in the study by Zheng [42]. Various benchmarks, such as PoseTrack[43], are often used for this comparison.

OpenPose [1] uses Convolutional Neural Networks (CNNs) to predict confidence maps S and Part Affinity Fields [44] L. The confidence maps represent the location of key points, while PAFs represent the degree of association between parts.



Where:

A: Represents the input image to the network, this is a color single RGB image in a video stream. The image is preprocessed [45], resized, and normalized.

B: Represents the convolutional layers[46] that learn spatial hierarchies of features. These layers capture low-level features like edges and textures initially and progressively extract higher-level features.

C: Represents the fully connected layers that make predictions based on the features learned[47].

D: Outputs the confidence map of the keypoint in the image for each body part, see fig. 1.1

E: Outputs the Part Affinity Fields (PAFs), see fig. 1.2

Confidence Heatmaps

For generating confidence heatmaps [48] in human pose estimation, each model-predicted 2D heatmap corresponds to a specific keypoint. Within this heatmap, every pixel represents the probability of the associated keypoint being at that specific location in the original image. Due to convolutional operations like pooling or striding, these heatmaps usually possess a lower spatial resolution than the input image. For example, a heatmap of 32x32 pixels would represent an original image of 256x256 pixels, with each pixel in the heatmap accounting for an 8x8 region in the image. The pixel with the highest value in this heatmap pinpoints the most probable position of the respective keypoint.

Figure 1.1: Confidence maps for keypoint "right shoulder", taken from the original OpenPose publication[1].

Part Affinity Fields

It is a novel representation introduced by the authors of OpenPose [1] to effectively detect the orientation and location of limbs (pairs of joints) in an image, even when the image contains multiple people in close proximity or with overlapping body parts. A Part Affinity Field is a 2D vector field for each limb, where each vector points from one joint of the limb to the other. The magnitude of the vector indicates the confidence that a limb exists in that particular position, see image 1.2

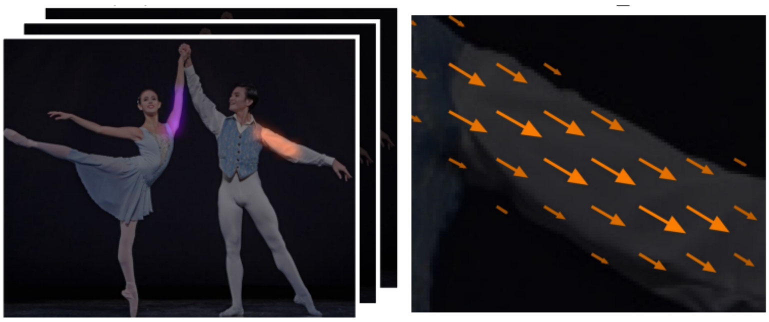


Figure 1.2: Affinity filed for keypoints "right shoulder" and "right elbow", taken from the original OpenPose publication[1].

The final pose estimation is then obtained by parsing the detected keypoints and using the PAFs to associate these keypoints with individual human figures in the image.

Chapter 2

Aim of the Thesis

2.1 Aim of the Thesis

Integrating camera systems into homes for health and fitness tracking has become more popular. These advanced systems, capable of real-time movement and exercise assessment, hold potential in domains such as personalized training, rehabilitation, and preventive measures against injuries. Despite their emerging popularity, critical inquiries persist: How accurately do these systems detect movements? Which exercises align optimally with this technology? What are the inherent limitations in their sensing capabilities? How computationally intensive are they? And, of overarching clinical significance, what are their potential applications within a home environment for health monitoring? What advantage do these systems have over virtual reality systems? This thesis outlines six key research questions to understand the complex nature of these systems.

1. What are the practical differences, advantages, and disadvantages between using virtual reality and a camera-based system for motion capture?

2. How well does the camera-based motion capture detection work?

3. What are the limits of capturing motion with this camera approach?

4. For which exercises or movements is this approach suitable?

5. How demanding is the camera-based system on computational performance?

6. What could be the clinical applications of this camera-based motion capture?

This thesis aims to provide a comprehensive view of the benefits, challenges, and potential of camera systems for home health and exercise monitoring.

The development of new technologies and tools often significantly changes how we solve problems and reach our goals. This is especially true in the fields of health and fitness. Employing camera systems and other technologies to monitor and enhance physical well-being at home can offer significant benefits to many.

However, it's crucial to recognize that it's not just about a technical solution. On the contrary, the successful deployment of these technologies necessitates the integration of expertise and experience from professionals in physiotherapy. They typically possess not only technical know-how but also a deeper understanding of anatomy and the musculoskeletal system, which is essential for the appropriate and efficient use of camera systems in a home environment.

Consequently, the advancement of these technologies necessitates the integration of physiotherapy professionals' expertise. By adopting this approach, the solution is optimized to genuinely enhance patient outcomes, facilitating improved exercise and rehabilitation results.

In my research and publications, I've emphasized collaboration with practicing physiotherapists and educators in the field of physiotherapy. I've structured our collaboration so that I provide all the technical and research resources, allowing them to focus exclusively on their expertise. Their role, therefore, was to define clinical criteria and assess the practical applicability of the measured parameters

Based on these questions, the following objectives were proposed:

1. Evaluate the advantages and challenges of using virtual reality for motion capture compared to camera-based systems. This investigation will explore how virtual reality can enhance motion capture with its immersive and interactive capabilities, offering potentially more precise and dynamic data collection in controlled environments. Conversely, it will also examine the limitations of virtual reality systems, such as potential technical complexities and user discomfort.

2. Describe the functional concept of telerehabilitation using a camera. Due to the ubiquity of cameras, this allows for capturing movement virtually anywhere using any device. This approach provides a versatile platform for rehabilitation and can be adapted to various environments, making it a flexible solution for diverse needs.

3. Verify detection functionality using a large video database. This will determine the optimal perspectives for the camera system, facilitating the establishment of the correct methodology for movement recording.

4. Assess the feasibility of building machine learning models with the gathered data. Expert evaluations by practicing physiotherapists will generate a dataset containing both comprehensive movement records and assessments of these movements. Such a dataset can then be used for more sophisticated data modeling.

5. A principal objective of the study is to integrate interdisciplinary insights from cybernetics, biomedical engineering, and physiotherapy, thereby enhancing the translational applicability of the research outcomes.

6. Develop an automated evaluation software tool, which will assist physiotherapists in facilitating and streamlining the diagnosis of exercise execution.

7. Create a functional application based on this system and define feedback elements for interactive exercises. To validate the entire concept in practice, it's essential to develop a working prototype for real-time exercise with an augmented reality mirror. Experimenting with this software will not only gauge the detection's success but more importantly, assess the user experience.

Chapter 3

Preliminary Research and Experiments

To truly understand the topic and provide a broader perspective, I began my research systematically by examining all comparable systems for capturing the human skeleton. This allowed me to form my own opinion on the issue and personally experience various technologies.

This chapter provides a detailed account of the entire research process, including all the steps and dead ends that led to the final concept of the doctoral thesis. While some parts of the preliminary research weren't directly related to the final outcome, they were all included within the framework of the dissertation. These phases were key in determining the right direction for research. Initially, there were two main areas of focus. The first was the use of virtual reality, reflected in our study presented in a publication titled "Juggling in VR: Advantages of Immersive Virtual Reality in Juggling Learning" 3.1. In this study, we primarily investigated how VR affects people's motivation to learn a new task. In the field of VR, I further developed an application for training in VR, and its concept was introduced in a conference paper [49], detailed further in chapter 3.2.1 After discussions with experts and based on our own experiences, I decided to abandon this direction and turned our attention to the use of RGB cameras.

I started with measurements using the standard Vicon system to gain real experience with the golden standard in this field, this part is described in section 3.4. Subsequently, I dedicated myself to determining camera specifications for person detection, this led to publication [50] and is described in section 3.5. A series of practical tests in various areas followed, such as angle measurements in practical goniometry, see section 3.6 or fatigue detection, see section 3.7 from video recordings. Based on practical experiences from these measurements, I collaborated on software development as part of an international project with Israeli partners. My part of this software solution is described in a section 3.8.

3.1 Juggling in VR: Advantages of Immersive Virtual Reality in Juggling Learning

3.1.1 Description of the Study

In a focused study on the applications of Virtual Reality (VR) in motion learning, this research explored its implications for enhancing motion performance, augmenting motivation for learning, and influencing participants' willingness to persist in motion training. Using a VR application that simulates reduced gravity to decelerate motions, 30 participants were trained in three-ball juggling. The findings underscored the beneficial role of VR in boosting motivation and highlighted its potential utility in the realm of motion learning. Such insights offer promising avenues for the application of VR in physical rehabilitation, illustrating its viability as a tool for patient engagement and training efficacy.

3.1.2 Context of the Dissertation

Incorporating Virtual Reality into motion learning, as demonstrated by the three-ball juggling study, presents clear parallels to its potential use in physical rehabilitation. The primary essence of physical rehabilitation is the relearning, adaptation, and strengthening of physical motions, much like the process of acquiring a new motor skill such as juggling. Details of the study are available in my published article [51].

3.2 Affordable Personalized, Immersive VR Motor Rehabilitation System with Full Body Tracking

3.2.1 Description of the Approach

In the evolving landscape of rehabilitative medicine, leveraging technology has become increasingly pertinent. My contribution to this domain was realized through the design and development of an affordable, personalized, and immersive Virtual Reality (VR) system specifically tailored for motor rehabilitation. Utilizing the robust capabilities of the HTC Vive and its accompanying trackers, I employed inverse kinematics to capture, interpret, and translate full-body movements into a virtual environment. Central to the system is a meticulously designed 3D human model that acts as the user's virtual representation. By integrating this model into Unity, users are empowered to view themselves in a virtual mirror setup, in real-time, aiding in feedback and self-awareness during rehabilitation exercises. One of the system's distinguishing features is its ability to display angle values, providing both therapists and patients quantitative feedback on movement range and accuracy. This synthesis of technology and rehabilitative theory not only showcases the potential of VR in medical applications but also offers an accessible solution for personalized motor rehabilitation. [49]

Figure 3.1: A person wearing several HTC Vive trackers on moving joints, the rest of the human skeleton is computed using inverse kinematics.

3.2.2 Usability Analysis in Dissertation Framework

While the VR-based rehabilitation system showcased innovative potential, several practical issues arose. Firstly, the setup was complex, requiring specific hardware, making everyday use challenging. Feedback from consulting physiotherapists highlighted concerns: the system wasn't as immersive due to occasional technical issues, and compared to RGB camera systems, it was less efficient. Older users, in particular, found the VR environment disorienting, limiting the system's applicability to a narrower age group. Given these challenges, especially when contrasted with the straightforwardness of RGB camera systems, I decided to shift my research focus to the camera based systems. The RGB camera approach offered broader adaptability with a more straightforward setup.

3.3 Automatic Telerehabilitation System in a Home Environment Using Computer Vision

In this conference paper, I summarize the whole concept of rehabilitation, using simple examples and experiments to describe the whole pipeline of use, evaluation, and feedback for patients. This publication functioned as a concept for future practical use. Details regarding this introductory article to the topic of using computer vision methods in physical telerehabilitation can be found in my published paper [52].

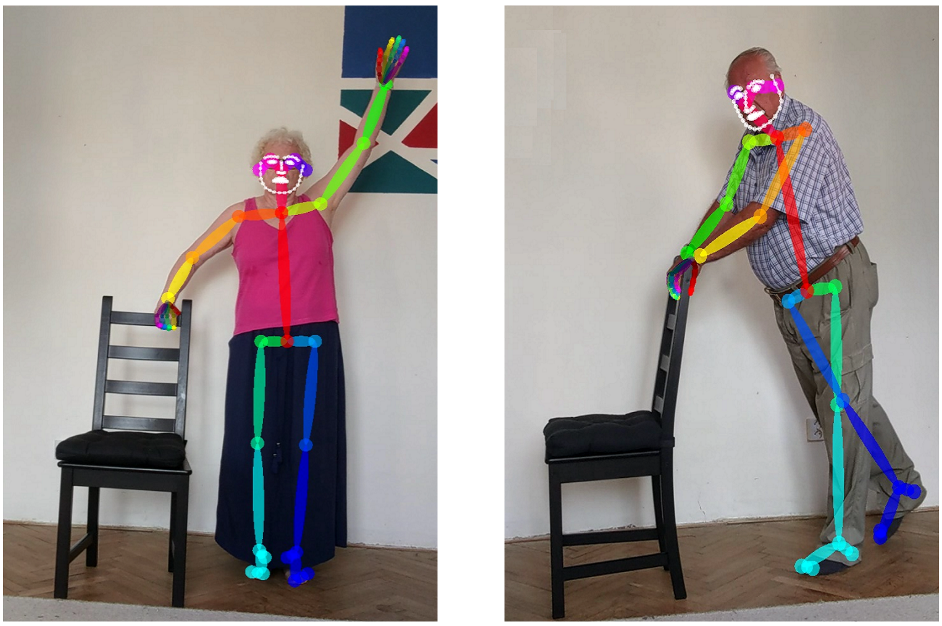


Figure 3.2: Rendered skeleton into a standard RGB image of the test subjects.

3.4 Preliminary Motion Capture Analysis with Vicon

In our quest for hands-on understanding, we undertook a measurement utilizing equipment owned by ČVUT. We quickly found out that although the system is very advanced, setting it up was complicated and took a lot of time. For rehabilitation, where it's important to be efficient and easy to use, this was a big issue. Even though we expected some difficulties, the real challenge of setting up the system became clearer during our tests. Together with worries about how practical it would be in a rehabilitation environment, where patients need quick and simple methods, we decided it might not be the best option for us. However, it's comforting to know that we can still use the ČVUT equipment for future research and comparisons if our goals or focus change.

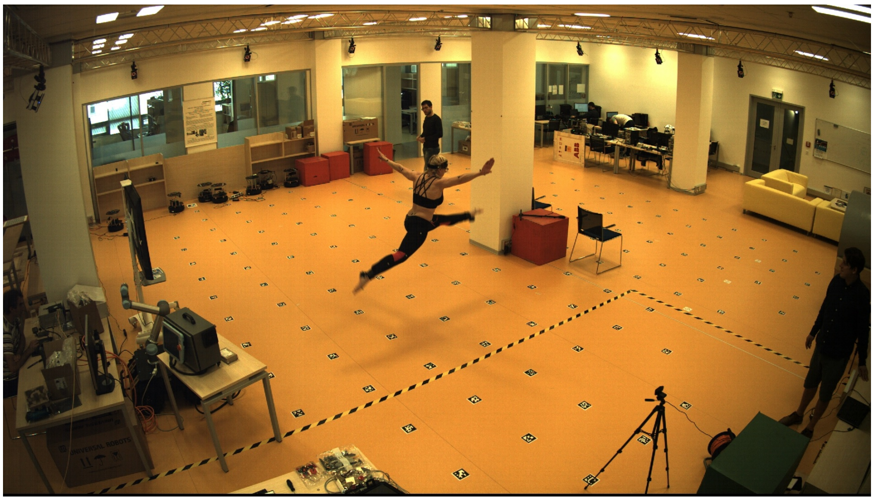


Figure 3.3: A woman captured in mid-jump, adorned with reflective sensors, as recorded by the Vicon system at CIIRC, ČVUT.

3.5 Remote Physical Therapy: Requirements for a Single RGB Camera Motion Sensing

The objective of this study was to determine the basic technical prerequisites for utilizing a standard RGB camera for motion detection in typical household settings. We empirically validated the capturing criteria needed for subsequent motion evaluation. Our research aided in advancing telerehabilitation methods that didn't rely on specialized hardware, making remote rehabilitation more accessible to the wider population. For a comprehensive breakdown and detailed insights into our methodologies, findings, and discussions, readers are directed to consult my conference contribution [50]. This document provides a detailed look at the study, covering all the subtle details and complexities that were crucial to our research findings.

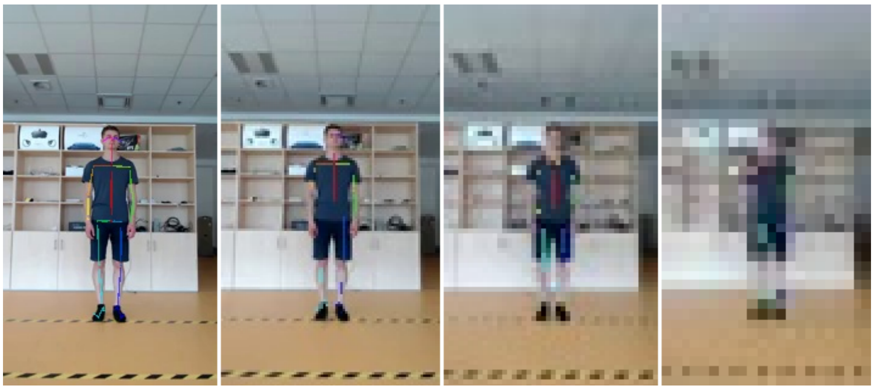


Figure 3.4: In one of the experiments, we evaluated the camera's detection quality across various resolutions to determine the optimal setting.

3.6 Upper Limb Range of Motion Evaluation by a Camera-Based System

The aim of this work is to evaluate the upper limb range of motion using a camera-based system for body pose estimation. For our research, we use the Openpose system based on deep neural networks. The advantage of this system is the possibility to use only a conventional RGB camera for human body motion capturing. OpenPose works best when the subject is standing directly facing the camera and the angles are measured in the frontal plane. In this work, we focus on verifying the applicability for practically used measurements in the supine position in the sagittal plane. To validate the accuracy of the camera-based system, we performed measurements on 48 healthy subjects. Each subject was measured in six different supine positions. A total of 14 shoulder and elbow angles were measured. The reference measurement was performed using a standardized method with a goniometer. The goniometric measurements were conducted by two professionals, and their agreement was taken as a reference. At the same time, the subjects were recorded by the camera. We extracted the key points from the video and calculated the corresponding angles. OpenPose was not able to detect the keypoints in all of the cases, but when the keypoints were detected, the angle estimation error is at the inter-rater error level, meaning that the angles were estimated with accuracy comparable to human experts. We further show the limitations of this approach and discuss the possibilities of using this system for future clinical applications.

I have published our collaborative research findings with UK FTVS at the WC2022 conference [53]. However, the event organizers have not yet released the full proceedings. The findings of this study can be found in the master's thesis of Michaela Sýkorová, titled "Evaluation of the Range of Motion of the Upper Limbs Using the OpenPose Program" [54].

3.7 Detection of Signs of Fatigue in Functional Tests in Healthy Population by OpenPose

This research was a collaborative effort between the Faculty of Physical Education and Sport (FTVS) and myself. Together with students, we established a research study, and details of this endeavor can be found in the diploma thesis of Tereza Skalová, titled 'Detection of Signs of Fatigue in Functional Tests in Healthy Population by OpenPose'. [55]

This research aimed to evaluate whether OpenPose, a markerless motion detection system, could identify changes in motion range and co-movements in other segments due to muscle fatigue. The study focused on two specific movements: repeated shoulder joint abduction with a load and standing on one lower limb. By comparing angle changes and relative distances at the beginning and end of the measurement period, the study sought to determine the system's effectiveness. Participants were healthy individuals aged 18-65, and their movements were recorded during the two tasks. OpenPose processed these recordings to detect human motion. Key point combinations were identified using kinesiology knowledge, and basic statistical analyses were conducted on the motion and posture data, focusing on changes over time. The results showed that changes in shoulder abduction due to fatigue were statistically significant, while changes in standing on one limb were mostly insignificant, with notable differences observed between the right and left sides. The study concluded that OpenPose is capable of detecting certain motion changes caused by muscle fatigue, underscoring its potential practical applications and the need for further research in this area.

3.8 Software for Quantitative Assessment of Movement Activity

During my involvement in the project focused on enhancing robotic physiotherapy treatment using machine learning methods (Identification Number of the Result: LTAIZ19008-V004), I contributed to software development. I was responsible for all aspects related to capturing, exporting motion data, analysis, and calculation of motion parameters. The output of my work was a SW for the ability to generate videos, serving as an easy-to-navigate tool for physiotherapists to review motion data. I was not in charge of the web component.

An example of the output is illustrated in Figure 3.5, where one frame is displayed.

3.8.1 Video Record Processing

This software contains an executable file that automatically scans the content of a specified folder containing videos, and identifies, and processes unprocessed video records. The software operates by analyzing each frame using

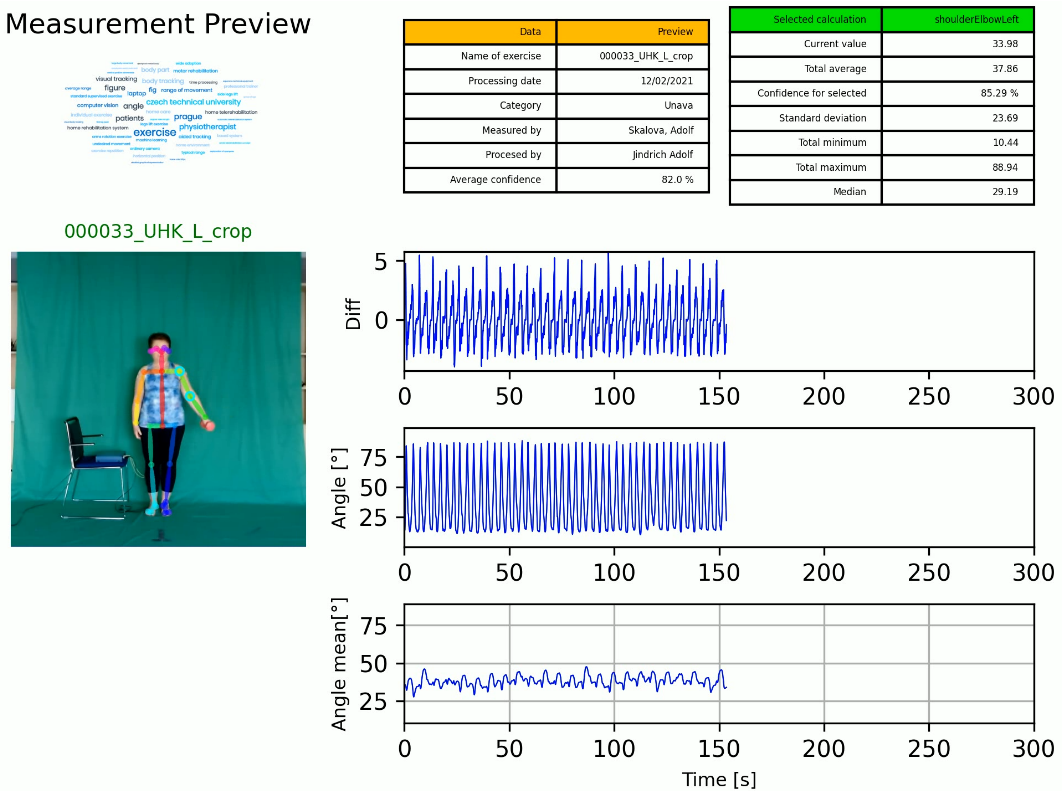


Figure 3.5: The software produces a video identical in length to the input but augmented with extracted data, allowing the physiotherapist to concurrently view the original footage alongside pertinent parameters and curves of chosen features.

OpenPose[1] for pose estimation.

The main features are:

1. Detection of a person in each video frame.
2. Identification of 25 key anatomical points on the human body, including eyes, mouth, ears, and joints.
3. Utilization of the OpenPose tool for detecting these points.
4. Drawing connections between detected points with colored lines.
5. Saving processed frames in JPEG format.
6. Recording the coordinates of detected points and the confidence level of detection.

The software creates a three-dimensional matrix of dimensions (IKB, EKB, DV), where:

- IKB is the number of key points (here 25).

- EKB represents the exported values (coordinates x, y, and confidence), which is 3.
- DV is the length of the processed video, i.e., the total number of frames.

The created matrices are saved in the formats .mat for analysis in the MatLab environment and .npy for further processing in Python using the Numpy library.

3.8.2 Interpretation of Stored Data

In addition to processing video records, the software also offers a tool for interpreting stored data related to the positions of anatomical points. With this tool, users can gain a visual understanding of the position and movement of these points during exercise. The output is a new video where the records are supplemented with a segmented body model. Added to this is a user interface that displays tables and graphs with calculated values of selected parameters.

This tool is invaluable for physiotherapists, who can assess the correctness of a patient's exercise.

For the selected angles, the observed parameters include:

- Current value.
- Mean value and standard deviation throughout the exercise.
- Median, minimum, and maximum value throughout the exercise.
- Range of values throughout the exercise.
- Minimum confidence value for the points.

Chapter 4

Single Camera-Based Remote Physical Therapy: Verification on a Large Video Dataset

4.1 Introductionary Comments

This chapter is primarily derived from the paper "Single Camera-Based Remote Physical Therapy: Verification on a Large Video Dataset," which has been adapted and expanded to fit the context of this dissertation. The focus of this chapter is to address three research questions that are central to understanding the efficacy and limitations of camera-based motion capture in physical therapy. These questions not only guide the structure of this chapter but also align with the broader objectives of this dissertation.

1. How well does the camera-based motion capture detection work?
2. What are the limits of capturing motion with this camera approach?
3. For which exercises or movements is this approach suitable?

The following sections explore each of these questions in detail. Drawing upon the findings and discussions presented in the original paper [56]. The integration of this paper into the dissertation allows for a comprehensive exploration of the camera-based system's capabilities, its constraints, and its applicability to various physical therapy exercises and movements. All co-authors have provided their formal acknowledgments, confirming their contributions to the work and their agreement with the authorship as presented. They collectively acknowledge that the core results and findings primarily originate from my dedicated research and efforts. For detailed information, please refer to Appendix A.

4.2 Abstract

In recent years, several systems have been developed to capture human motion in real-time using common RGB cameras. This approach has great potential to become widespread among the general public as it allows the remote evaluation of exercise at no additional cost. The concept of using these systems in rehabilitation in the home environment has been discussed, but no work has addressed the practical problem of detecting basic body parts under different sensing conditions on a large scale. In this study, we evaluate the ability of the OpenPose pose estimation algorithm to perform keypoint detection of anatomical landmarks under different conditions. We infer the quality of detection based on the keypoint confidence values reported by the OpenPose. We used more than two thousand unique exercises for the evaluation. We focus on the influence of the camera view and the influence of the position of the trainees, which are essential in terms of the use for home exercise. Our results show that the position of the trainee has the greatest effect, in the following increasing order of suitability across all camera views: lying position, position on the knees, sitting position, and standing position. On the other hand, the effect of the camera view was only marginal, showing that the side view is having slightly worse results. The results might also

indicate that the quality of detection of lower body joints is lower across all conditions than the quality of detection of upper body joints. In this practical overview, we present the possibilities and limitations of current camera-based systems in telerehabilitation.

4.3 Introduction

The general concept of remote rehabilitation using motion capture (MoCap) systems has undergone a turbulent change in recent years, as there are several tools for three-dimensional assessments, including sophisticated automation technologies and algorithms, often costing time, expensive equipment and inapplicable inconvenience to the daily practice [57]. Telerehabilitation, or remote physical therapy, is one of the most common types of complex distance medicine that is applied in practice [58]. During recent years, a large number of MoCap systems detecting the pose of a human using a "markerless" approach have emerged [59], these systems work without the necessity of placement of any markers on the human body [60].

This approach reduces the technical and financial requirements and complexity of arrangement [61] and therefore it can be found in the context of distance medicine, not only in specialized clinics but also in the home environment [52].

Considering the application of distance medicine in the home environment, the most promising systems seem to be systems for the evaluation of body movements from commonly used standard video (RGB) records [62].

In this case, we only need a regular camera, which is currently integrated into most common electronic devices, such as smartphones or laptops, or smart TVs.

These systems have great potential for use mainly due to the reduced financial costs of purchasing these systems. These systems have reached such technical levels that they could be used in specific cases as alternatives to costly systems in clinics. However, these systems must also use special camera data processing software [63].

The most commonly used software tools for pose detection are Open-Pose [64], Mask R-CNN [65], Google's Media Pipe [66], Alpha-Pose [40] all available as open source.

The time to the advent of markerless-based systems using neural networks is described in detail in Coyer's review [67]. All the systems mentioned above were operated only under laboratory conditions or used special HW.

All these software tools use artificial intelligence methods, namely neural networks (NN) trained on annotated images [68]. The datasets contain general static images of people in undefined positions, according to which NN learns to recognize anatomical points on the body. Existing benchmarks compare the speed and accuracy of detection using the above algorithms based on NN [68]. A shortcoming that limits wider use of low-cost MoCap systems and the mentioned software is the absence of evaluation of the validity of the data provided by these systems. This raises doubts about the use of telemedicine where it is necessary to know relatively accurately the information about the movement of specific anatomical points that physiotherapists need to monitor and modulate the rehabilitation intervention [69]. Thus, the main aim of our study is to determine whether the systems are sufficient to be used for home rehabilitation and under what specific conditions. We focus on the evaluation of the motion capture of different exercise positions and by different camera views, i.e., camera position relative to the subject. To achieve this aim, it is not possible to rely on existing benchmarks, but it is necessary to evaluate the efficiency of software use on video recordings of people moving while exercising. Thus, our study aims to validate the application of markerless systems using only one, generally positioned camera and thus applicable to home telerehabilitation.

Related Work

Studies [52, 70] show how OpenPose and similar camera-based systems could be used in telerehabilitation, but they are not dealing with the practical telerehabilitation applicability. Studies that evaluate the accuracy of motion detection typically study only one specific type of exercise motion. Hernández [71] concludes that OpenPose is an adequate library for evaluating patient performance in rehabilitation programs that involve the following exercises: left elbow side flexion, shoulder abduction, and shoulder extension based on comparison with Kinect. Ota [63] verifies the reliability and validity of OpenPose-based motion analysis during the bilateral squat based on comparison with Vicon MoCap system. Studies [72, 73] show that OpenPose can be used to capture and analyze both normal gait and abnormal gait. Nakano [74] compares the accuracy of 3D OpenPose with multiple cameras to the Vicon marker system. This study considers common human body movements such as walking, jumping, and ball throwing.

All the aforementioned studies use a system fixed in relation to the moving body and thus represent a one-sided task of interest, i.e., measuring body motion in only one anatomical plane. In general, when it comes to gait or run analysis, the setup is always the same, i.e., the camera is positioned to record only the sagittal plane of the body as accurately as possible. When it comes to measuring range of motion or measuring angles between joints, the setup is such that the person stands in a precisely defined position to the camera and only selected angles in a single anatomical plane are measured. Thus, none of the studies presented here considers the application of a camera based approach in distance rehabilitation and NN based software to recognize anatomical points, where the precise alignment of the camera system perpendicular to the anatomical plane being measured would not be necessary prior to measurement.

4.4 Methods

4.4.1 Design

The advantage of the camera-based approach is the ability to detect motion from any regular RGB video. This allows to use existing recordings and perform a large-scale evaluation. In this study, we choose to use the database created by PhysioTools. PhysioTools is one of the world's most comprehensive exercise libraries [75]. In our study, we consider only the ability of the system to estimate the human pose. We do not analyze the exercise itself. Therefore, we only extract the OpenPose reported confidence of the selected keypoints to infer the quality of detection. The design of our study is shown in Figure 4.1.

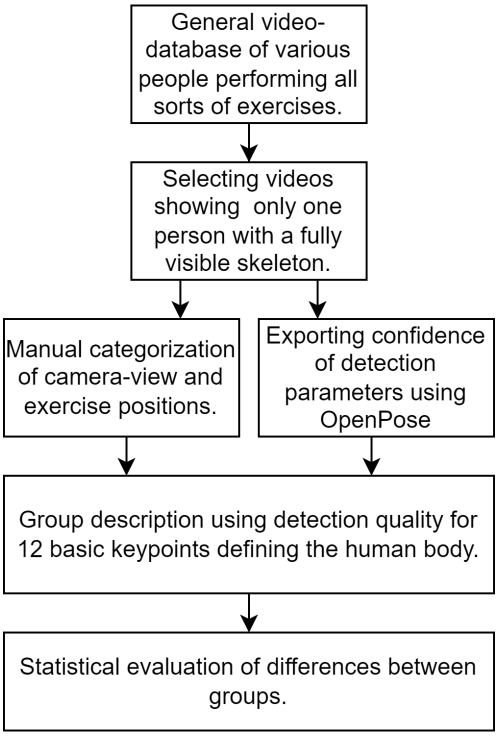


Figure 4.1: Study design—diagram explaining the sequence of steps.

For research purposes, we use a database based on an agreed template of commonly prescribed physical exercises printed from commonly used PhysioTools computer software (PhysioTools©, Product ID RG-PT1ENG, General Exercises Second Edition (English), Tampere, 339 Finland) [76]. PhysioTools is a database of professional trainers and serves as a video aid for exercises in the home environment. The average length of a video is 20 s. Videos have frame rates ranging from 25 fps to 50 fps, resolutions ranging from 0.1–0.6 Mpixel, and bit rate greater than 200 kbs. Our aim is to evaluate practical usability of recordings made in an uncontrolled home environment so we use no additional constraints on video quality.

Unlike typical studies [77] studying only one type of movement, the database we used is composed of hundreds of unique physical exercises, see Table 4.1 for quantity and Table 4.2 for categories.An exercise can be included in the database if it is performed by a single person and shows their whole body. Each video was manually checked to confirm that the entire trainee's body was in view throughout the recording. At the same time, manual categorization into specific groups was done by a single rater. Border cases were excluded from the analysis.

Total (2133)											
Front View (357)				¾ View (1027)				Side view (749)			
Ly	Kn	Si	St	Ly	Kn	Si	St	Ly	Kn	Si	St
75	26	186	70	177	95	165	490	215	92	145	297

Table 4.1: Number of videos in each category. Subcategories are: lying down (LY); on the knees (Kn); sitting (Si); standing (St).

4.4.2 Video Analysis

Since the practical use assumes only one camera, we were interested in the influence of the orientation of the person towards the camera.

By analyzing the database, we determined the following three basic views to be the most common: the frontal view (frontal plane), the side view (sagittal plane), and the ¾ view, which is in between these planes, please see Figure4.2. Although the ¾ plane is not biomechanically defined, it was most frequently used in instructional videos, because it provides an overview of the entire body and a better spatial understanding of records.

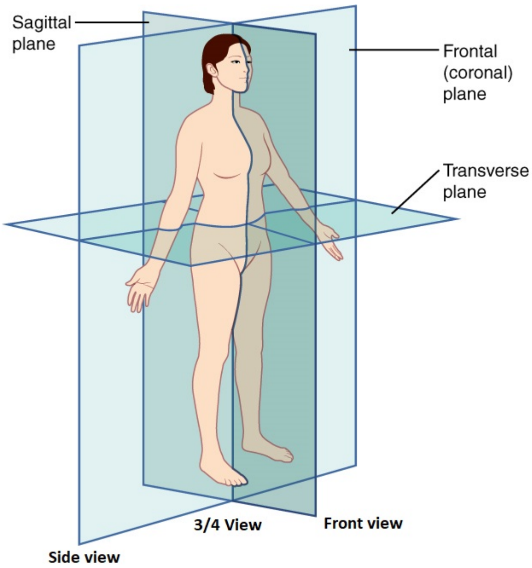


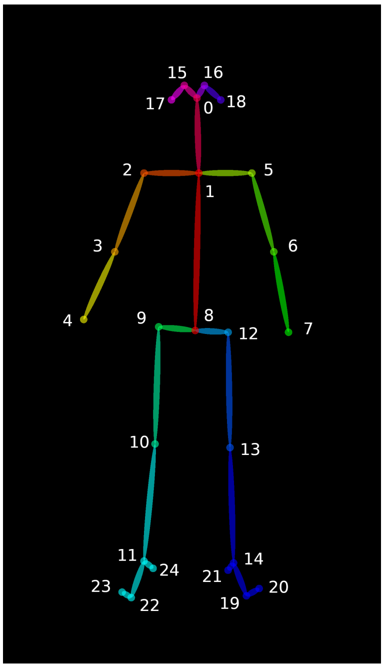
Figure 4.2: Orientation of the camera relative to the subject. Camera view options: front, ¾ and side [78]

Starting Position	Camera View
lying down (Ly)	front view (frontal plane)
on the knees (Kn)	¾ view
sitting (Si)	side view (sagittal plane)
standing (St)	

Table 4.2: Categorization of videos.

4.4.3 Keypoint Confidence Extraction

The videos were analyzed using the OpenPose [44] algorithm. OpenPose uses a model with 25 keypoints. In the context of performing rehabilitation exercises, the body segments that are part of the appendicular skeleton are very often measured [79]. These segments allow for translational movement of the body through cyclic movements such as walking [80]. Thus, we used 12 points that allow us to determine the positions and orientations of segments of the appendicular skeleton for further analysis. These 12 points are described in Table 4.3.



.

Figure 4.3: OpenPose 25 keypoints model [64]

No.	OP Name	The Most Appropriate Name of the Anatomical Point
Upper body		
2	RShoulder	R. acromion, end of the clavicle (collar bone)—top of shoulder
3	RElbow	R. lateral epicondyle of humerus, lateral epicondyle of the humerus. Outside of elbow.
4	RWrist	R. styloid process of the radius; wrist on thumb side.
5	LShoulder	R. acromion, end of the clavicle (collar bone)—top of shoulder
6	LElbow	R. lateral epicondyle of humerus, lateral epicondyle of the humerus. Outside of elbow.
7	LWrist	L. styloid process of the radius; wrist on thumb side.
Lower body		
9	RHip	R. femur greater trochanter
10	RKnee	R. femur lateral epicondyle
11	RAnkle	R. fibula apex of lateral malleolus
12	LHip	L. femur greater trochante
13	LKnee	L. femur lateral epicondyle
14	LAnkle	L. fibula apex of lateral malleolus

Table 4.3: The 12 basic keypoints of the appendicular skeleton.

The outputs of OpenPose for each frame are the x,y coordinate values and the detection confidence for each of the 25 model points, please see Figure 6.2. From this information, we use only the confidence for the twelve points for each frame. This gives us 12 temporal signals for each unique video.

OpenPose processes each frame independently without using the time context. Repeated image analysis returns the same results. OpenPose has almost perfect test–retest reliability within device [63].

In our study, we are not dealing with the absolute position of the detected points. To evaluate the quality of detection, we use directly the confidence returned by OpenPose. Thus, we are not evaluating the accuracy of detection, but the detection capability itself.

Detection accuracy using annotated images is a classical metric for comparing machine learning algorithms; accuracy calculations are performed on large image datasets COCO [37], MPII [36]. In contrast, the evaluation of dynamic events has been studied only for single exercises and sub-joints [77]. Existing annotated 2D datasets deal with either images [81] or deal with a small variety of activities [82, 83].

In contrast to the above-mentioned studies, we deal with many unique rehabilitation exercises.

OpenPose returns the confidence values of the keypoints in the interval of <0, 1>. Points that are not detected have a confidence value of zero. The confidence value is rarely used in single-camera setups because the position of the detected joints is accurate even with an average confidence value. On the other hand, in multi camera setup and 3D reconstruction tasks, which are very sensitive to misclassification, the confidence is used to weight joint positions [84] or to discard joints with a low confidence value, as the misclassification of points appears with values below 0.2 [85].

The result of the processed video is a matrix of 12 keypoint confidences in time.

4.4.4 Statistical Analysis

To compare the records, we calculated the medians of each time signal. This gave us 12 scalar keypoint confidence values defining each video. All records were assigned to exactly one of the subgroups, see Table 4.1. For each subcategory, we calculated the median confidence of the individual keypoints.

To visualize the results we chose box plots where outliers are not shown for clarity. We define outliers as elements more than 1.5 interquartile ranges above the upper quartile (75 percent) or below the lower quartile (25 percent).

The Kruskal–Wallis test was used to determine the statistical significance of group differences. Due to a large amount of data and the significant difference of one of the groups, all results were significant. Therefore, we decided to compare all groups, individually, with each other. To verify the normality of the data we used the Shapiro–Wilks test. The test has shown that values in the groups are not normally distributed. The groups also varied in size, see Table 4.1, therefore we used the Wilcoxon test to determine the statistical significance of the difference between the categories. All statistical calculations were performed using the Matlab (MATLAB and Statistics Toolbox Release 2019b, The MathWorks, Inc., Natick, MA, USA).

4.5 Results

In our study, we analyzed a total of 2133 videos. Each video shows only one trainee performing a unique exercise. Each video belonged to one of the “Camera View” and one of the “Starting position” groups. We present the results of our findings using OpenPose reported confidence values. We can use the confidence returned by the OpenPose algorithm as a measure of detection quality because it correlates with the percentage of correct keypoints metric [86].

The resulting median confidences for each joint and each category are shown in Table 4.4. Keypoints with a confidence value above 0.5 can be considered correctly detected [86]. These confidence values are associated with clearly visible body parts [64].

Medians of Confidence for Selected Group and OP Keypoint													
Camera View		Front view				¾ view				Side view			
Starting possition		Ly	Kn	Si	St	Ly	Kn	Si	St	Ly	Kn	Si	St
No.	KP name												
Upper body													
2	RShoulder	0.67	0.65	0.78	0.80	0.62	0.67	0.77	0.79	0.68	0.68	0.76	0.78
3	RElbow	0.75	0.81	0.82	0.84	0.76	0.81	0.82	0.84	0.78	0.82	0.81	0.83
4	RWrist	0.75	0.79	0.82	0.84	0.75	0.80	0.82	0.83	0.79	0.81	0.81	0.83
5	LShoulder	0.54	0.60	0.79	0.80	0.53	0.63	0.80	0.80	0.52	0.54	0.77	0.78
6	Lelbow	0.39	0.78	0.81	0.83	0.66	0.78	0.79	0.81	0.41	0.72	0.79	0.79
7	LWrist	0.39	0.77	0.79	0.83	0.59	0.80	0.79	0.81	0.30	0.77	0.78	0.80
Lower body													
9	RHip	0.42	0.48	0.64	0.64	0.38	0.51	0.66	0.64	0.40	0.49	0.59	0.64
10	RKnee	0.62	0.75	0.80	0.79	0.61	0.78	0.81	0.78	0.65	0.78	0.78	0.78
11	RAnkle	0.54	0.64	0.73	0.79	0.53	0.68	0.71	0.79	0.56	0.67	0.69	0.78
12	LHip	0.38	0.44	0.65	0.65	0.35	0.47	0.66	0.65	0.35	0.40	0.58	0.63
13	LKnee	0.50	0.72	0.80	0.79	0.41	0.71	0.81	0.79	0.51	0.66	0.79	0.78
14	LAnkle	0.43	0.54	0.71	0.78	0.40	0.65	0.71	0.78	0.47	0.55	0.68	0.78

Table 4.4: Median confidence value for each specific group.

Due to the large amount of data in the groups, the differences in confidence between groups are largely significant, although small in absolute terms. Therefore, for clarity, only differences that were not statistically significant (n.s.) are highlighted in the following boxplots. All the other differences were statistically significant. Box-and-whisker plots of detection confidence for all the categories are show in Figures 4.4, 4.5, 4.6, 4.5, 4.5 and 4.5.

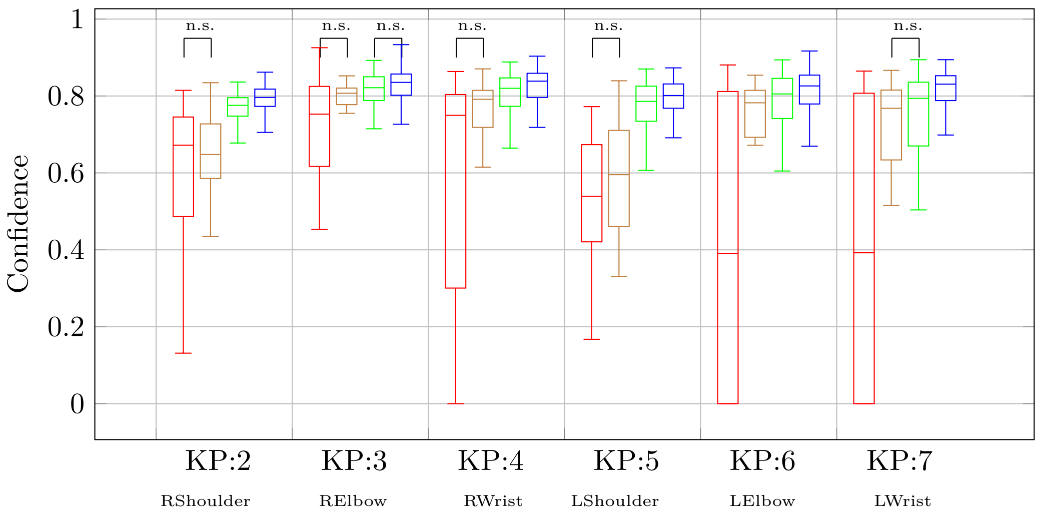


Figure 4.4: Confidence of detection of selected KP (keypoints)—front view, upper body. Camera views are shown in colour as follows: lying down (N = 75); on the knees (N =26); sitting (N =186); and standing (N =70). All the other differences are statistically significant. Only differences that were not statistically significant (n.s.) are highlighted in the boxplots.

When interpreting the plots, it is important to note that the confidence correlates nonlinearly with the detection accuracy. If all keypoints in the image are clearly visible and accurately detected, confidence values in the range of 0.7–0.9 are commonly obtained. From a practical perspective of a single camera-based recording, the differences in accuracy associated with this range of confidences are not relevant. Even points with a confidence value above 0.5 can be considered correctly detected [64]. Lower confidence values are also associated with high-frequency keypoint jitter [87], but this effect can be easily filtered out because the body movements during rehabilitation exercises are slow relative to the sampling rate of the camera. False-positive detections or swapped keypoints can only be expected for low confidence values around 0.2 [85]. Small confidence values (0.1) are associated with guessed and occluded keypoints; the smaller the value,the more false positives detections are likely [64].

To summarize, as long as the value of the lower quartile is above 0.5, we can say that the combination of body position and camera view is practically usable.

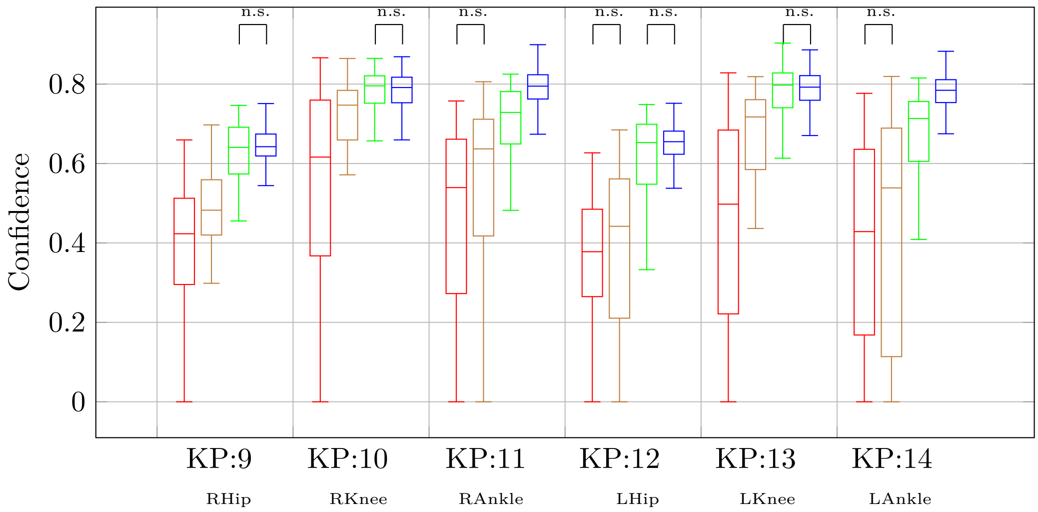


Figure 4.5: Confidence of detection of selected KP (keypoints)—front view, lower body. Camera views are shown in colour as follows: lying down (N = 75), on the knees (N = 26), sitting (N = 186) and standing (N = 70). All the other differences are statistically significant. Only differences that were not statistically significant (n.s.) are highlighted in the boxplots.

Interestingly we can see the differences between upper body points and lower body points. Upper body detection performs better for all camera views.

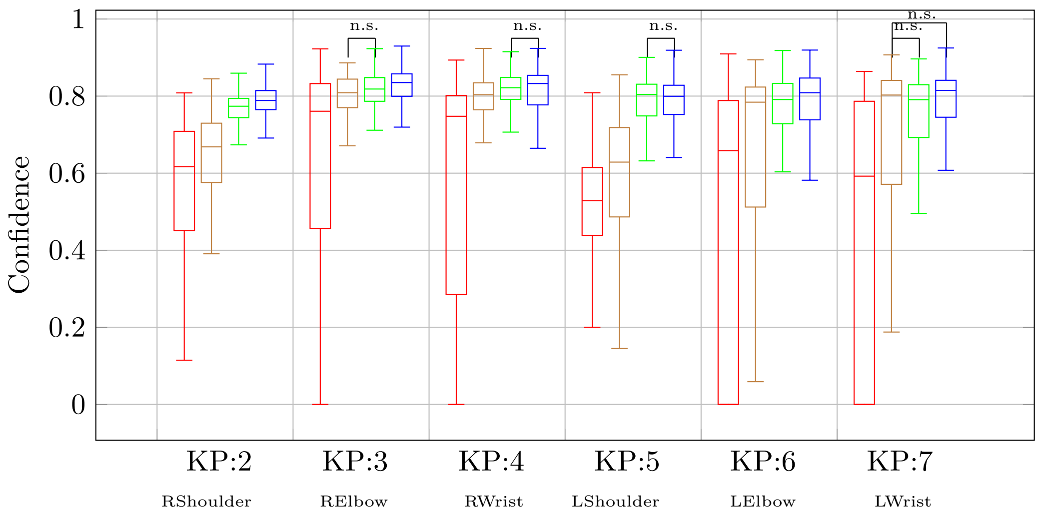


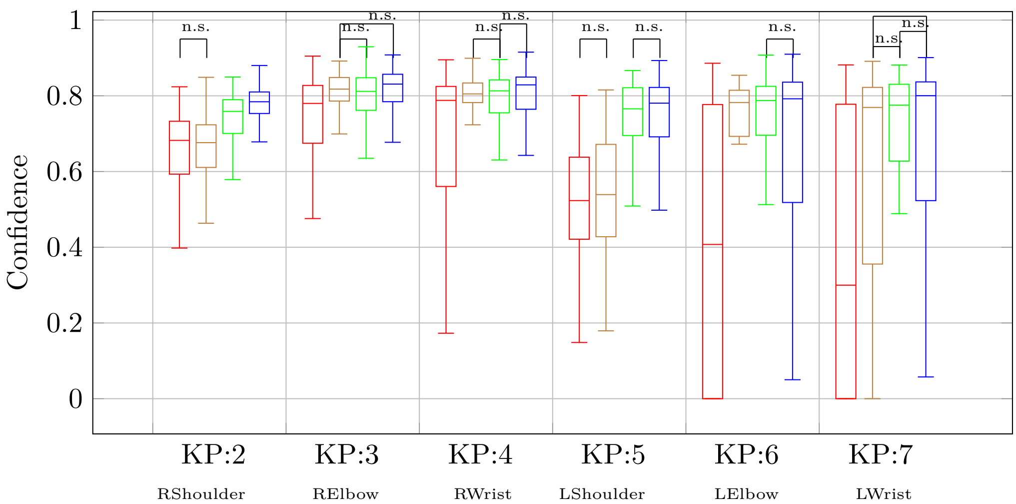
Figure 4.6: Confidence of detection of selected KP (keypoints)—¾ view, upper body. Camera views are shown in colour as follows: lying down (N = 177); on the knees (N =95); sitting (N =165); and standing (N =490). All the other differences are statistically significant. Only differences that were not statistically significant (n.s.) are highlighted in the boxplots.

Non-significant differences in confidence are often found with standing and sitting positions, which is due to the fact that detection works very well.

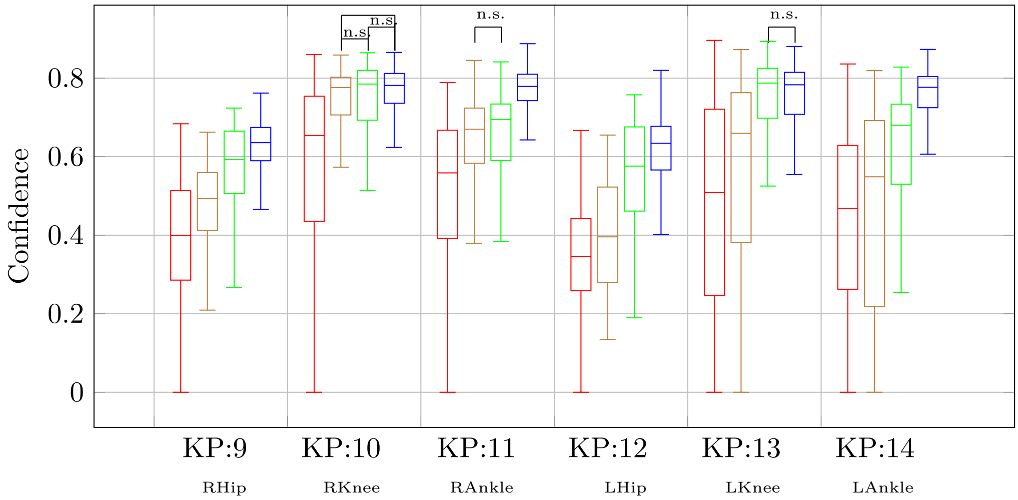
If the value of the lower quartile of confidence is less than 0.5, it is likely that the detection will not work in all cases.

—¾ view, lower body.]Confidence of detection of selected KP (keypoints)—¾ view, lower body. Camera views are shown in colour as follows: lying down (N = 177); on the knees (N =95); sitting (N =165); and standing (N =490). All the other differences are statistically significant. Only differences that were not statistically significant (n.s.) are highlighted in the boxplots.

In all boxplots, we can observe an increasing tendency of confidence values between the groups of starting positions. The worst confidences are achieved by lying down (red), followed by on the knees (brown), and the best results are achieved by exercises performed in sitting (green) and standing (blue).



—side view, upper body.]Confidence of detection of selected KP (keypoints)—side view, upper body. Camera views are shown in colour as follows: lying down (N = 215); on the knees (N =92); sitting (N =145); and standing (N =297). All the other differences are statistically significant. Only differences that were not statistically significant (n.s.) are highlighted in the boxplots.



—side view, lower body.]Confidence of detection of selected KP (keypoints)—side view, lower body. Camera views are shown in colour as follows: lying down (N =215); on the knees (N =92); sitting (N =145); and standing (N =297). All the other differences are statistically significant. Only differences that were not statistically significant (n.s.) are highlighted in the boxplots.

In the boxplots, the black line represents the groups that have no significant differences between them. These are mostly groups of exercises in sitting and standing positions. Thus, we can say that the detection is very reliable in both these groups and there are no significant differences between these groups. The achieved results clearly show that there are significant differences with respect to starting position.

4.6 Discussion

The main objective of our research was to determine the usability of a camera-based system in a home environment. Since human position detection is captured by only one RGB camera, we were mainly interested in the influence of camera view and the position of the trainee, where we expected the greatest impact on the detection of keypoints. The quality of detection was determined by the confidence of detection of each keypoint.

Our main findings are the following: regardless of the camera view, the lying position comes out as the least detectable, followed by the position on the knees. The standing position is the most efficient, but the absolute differences against the sitting position are small. In the case of the camera view, the results were not so convincing. For the lying position and the position on the knees, the differences are not statistically significant in most cases, but no conclusions can be drawn because of the large variances.

For standing and sitting positions, the camera view from the side is a bit worse. From the data we have available, it is not possible to give a clear answer to the question of whether the confidences for the different camera views differ.

We also found out the difference in confidence between upper-body and lower-body keypoints. Confidence of lower-body keypoints is generally lower. This can be explained by the fact that the positions of the upper limbs are more variable than the position of the lower limbs. Generally speaking, joint positions are easier to establish if they are at an angle other than 180 degrees, which is typically the angle of the knee. The hips are not as visible as the shoulders and the ankles are often covered by shoes and trousers, while the position of the wrist can be very easily derived from the palm of the hand.

We can justify generalizing the results about views and postures given the high number of unique exercises, as opposed to works focusing on specific exercises, where only a few different types of exercises are involved.

Before applying the research results in practice, it is important to define several assumptions and limitations. They are closely linked to the application area of telerehabilitation in home settings. The first assumption and limitation at the same time is the use of a single simple camera (smart phone, tablet). The second assumption is the application use by nontechnical users that results in the requirement of simple control and setup of the application. These considerations led us to experiments analyzing the influence of the body position in relation to the camera and evaluation of many different exercises recorded by a single camera.

In the light of these facts, we are well aware of the limitations of the proposed approach, in particular the precise identification of certain motions in the front or side view. For example, abduction of the right arm cannot be well recognized in the side view from the left side, the range of straddling backward or angle of the knee cannot be precisely identified in frontal view.

For practical usability, it is important that there are not too many dropouts, i.e., that the joints are detected, and that they are not mistaken with another part of the body, e.g., the left and right limbs are swapped when viewed from the side, and so on. Another important aspect of the evaluation is that each individual exercise engages different parts of the body, thus only certain points are important for the analysis of the given exercise. The camera view is chosen so that the parts of the body being exercised are clearly visible while at the same time some parts of the body can be obscured. With the side camera view, the other side is often not visible.

Therefore, it can be assumed that points with a lower confidence value do not play a large role in the exercise. Just the fact that low confidences are found for individual joints does not necessarily mean that the exercise cannot be successfully evaluated. This is also the reason why we decided to present the results of individual joints and not evaluate the confidence of the whole exercise.

4.7 Conclusions

Despite the fact that there have been many recent publications describing the possibility of using a camera-based system for home rehabilitation, there has been no work to date that has validated the detection capability on a large dataset consisting specifically of videos of people performing rehabilitation exercises in front of a camera.

We validated the ability of the OpenPose algorithm to detect the keypoints of the human skeleton on more than two thousand videos of people performing rehabilitation exercises.

Based on our findings, we can say that OpenPose, for detection, is a sufficiently robust algorithm that is capable of detecting people during commonly performed exercises in a home environment. Only exercises performed in the lying and on-the-knees positions may not always be correctly detected. In this study, we also analyzed closely the basic landmarks of the human skeleton, see Table 4.3 and gave a summary of which keypoints are more reliably detectable. In that way, we provided an identification of the important points on the skeleton for each exercise, and, thus, offered a practical overview for designers of future camera-based telerehabilitation systems.

Chapter 5

Evaluation of Functional Tests Performance Using a Camera-based and Machine Learning Approach

5.1 Introductionary Comments

The publication 'Evaluation of Functional Tests Performance Using a Camera-Based and Machine Learning Approach' is a vital part of my dissertation, expanding on the work of earlier experiments and publications. It focuses on the fifth research question of my study: 'What are the clinical applications of camera-based motion capture?' The goal is to show how this new method can be used in real situations, connecting expert insights with the use of machine learning algorithms for assessment purposes.

This chapter begins by outlining the methodology employed, emphasizing the integration of camera-based motion capture technology with machine learning algorithms. This approach not only advances our understanding of functional test performance but also highlights the ease of data acquisition compared to other systems, a crucial factor for the effectiveness of machine learning applications.

This research is important because it has the potential to change clinical practices using technology. It merges expert knowledge with sophisticated algorithmic analysis, leading to more precise, faster, and easily accessible clinical evaluations.

In the next sections, I'll explain the specific goals, methods, findings, and implications of this study, always connecting them to the wider objectives and questions of my dissertation. This chapter adds value to the field of telerehabilitation by offering practical solutions and showcases how technological progress can be applied in clinical environments. This method was developed within the framework of the international project TAČR LTAIZ19008. As discussed in the following chapter, the full details of this study can be found in the original paper [88].

I conducted this study in collaboration with colleagues from Charles University, who supported me with an expert design of the study, and together we carried out all measurements. Although we worked as a team, the entire concept and research design stemmed from my initiative, which is confirmed by all co-authors in the document attached as appendix A.

5.2 Abstract

The objective of this study is to evaluate the performance of functional tests using a camera-based system and machine learning techniques. Specifically, we investigate whether OpenPose and any standard camera can be used to assess the quality of the Single Leg Squat Test and Step Down Test functional tests. We recorded these exercises performed by forty-six healthy subjects, extract motion data, and classify them to expert assessments by three independent physiotherapists using 15 binary parameters. We calculated ranges of movement in Keypoint-pair orientations, joint angles, and relative distances of the monitored segments and used machine learning algorithms to predict the physiotherapists' assessments. Our results show that the

AdaBoost classifier achieved a specificity of 0.8, a sensitivity of 0.68, and an accuracy of 0.7. Our findings suggest that a camera-based system combined with machine learning algorithms can be a simple and inexpensive tool to assess the performance quality of functional tests.

Introduction

The number of people that need physical therapy increased during the last decades [89]. The cost of rehabilitation treatment increased consequently. The patients are demanding a better patient experience. There is a need for more doctors and assistants together with better services and individualized approaches. Such a system will be unsustainable shortly. During the last decade, advanced technology allowed us to look at the problem from an entirely new perspective and create systems based on the current knowledge and technology level.

For physiotherapists (PTs) or athletic trainers, visual observation is standard practice. The observational analysis relies on the skill of the evaluators and a clinical evaluation that identifies possible deficiencies in movement expression. Performing functional tests and their clinical, subjective evaluation is a common examination method in the differential diagnostics of physicians and physiotherapists [90]. Such tests typically combine screening of a range of motion, strength, and proprioceptive assessment. Examples of such tests are the well-known Single leg squat (SLST) and Step-down tests (SDT). McGovern [91] describes the implementation and possibility of evaluation by one or more experts observing the patient. Schurr et al. [92] proved that for lower extremity movement, 2D analysis is comparable to the frontal plane of 3D motion analysis commonly regarded as the gold standard. However, this only applies if the camera capturing the person is perpendicular to the frontal plane and is positioned at the center of the patient's body. If an error in the exercise execution can be detected by an expert from a video, we hypothesize that the same error can be detected by machine learning algorithms. Harris-Hayes et al. [93] demonstrated reliability based on visual assessment of lower extremity movement patterns by observing classic 2D RGB recordings.

A PT can use simultaneous motion assessment to assess knee joint dysfunction or pain [94, 95], as well as to assess the hip, pelvic, and trunk deviations, which are also important in people with hip pain [91, 96, 97, 98]. The development of modern motion capture (MoCap) systems makes it possible to automatically evaluate the performance of functional tests. Methods for automatic evaluation of similar tests using 3D MoCap techniques have been proposed in the past. Mostaed [99] compared the visually-assessed quality of the step-down test and corresponding Vicon data at [99], Ageberg [100] compared Vicon and Visual analysis for single-limb mini squat. Barker-Davis [101] measured subjects performing the leg squat exercise using Vicon and compared the data obtained with the subjective assessment of five independent experts who evaluated the correctness of the performance by observing videos taken at the same time.

In their comprehensive analysis, Debnath [102] and colleagues provide an extensive review of computer vision-based systems used in physical rehabilitation over the past two decades. The authors propose an innovative taxonomy, categorizing these systems according to the perspective of rehabilitation and assessment. They classify the mode of rehabilitation into two types: Virtual and Direct. In terms of assessment, the authors suggest three different approaches: Comparison, Categorization, and Scoring. Their exhaustive review not only serves as an excellent overview of all applications and approaches within this field but also suggests a unique perspective to classify and understand them. Coyer [103] has compiled a review outlining the state-of-the-art, markerless systems through 2018. This review includes camera-based systems, as well as more hardware-complex systems. Since then, object detection systems based on deep neural networks have mainly been used in the field of camera systems. Currently, the most used open-source systems based on the above methods are OpenPose [35], AlphaPose [104], [105] and Google Mediapipe [106]. They all work on a similar principle and even use similar databases – usually COCO [107] and MPII [108] – for training to create their models. It is important to notice that OpenPose (OP) requires a powerful GPU to process the video, an issue that might disturb the option to use our solution. In our previous work we have demonstrated where to use a more efficient solution without the need for a GPU to process the videos in real-time [109]. In our next study [50], we proved that OpenPose is robust in terms of quality, video resolution, and lighting changes. Thus, its use is in principle not limited by the environment where the measurement is performed. Ota's study [110] has substantiated the robustness and precision of OpenPose in keypoint extraction, particularly in motion analysis. The investigation utilized OpenPose to study the movements of 20 healthy young participants performing bilateral squats. The joint angles - pertaining to the trunk, hip, knee, and ankle - as calculated by OpenPose were contrasted with those captured by the highly accurate VICON motion analysis device.

Intraclass correlation coefficients (ICCs) indicated an almost flawless consistency between the data generated by OpenPose and VICON, thereby manifesting OpenPose's high reliability. While minor biases were noted for certain joints, they were documented for future corrections. This investigation verifies not only the reliability of OpenPose but also its cost-effectiveness and user-friendliness compared to traditional methodologies.

Further research validating the reliability of systems like OpenPose using these methodologies has been confirmed across several studies. These include comparisons of 2D and 3D accuracy in actions like lifting [111] and squatting [112].

As demonstrated in all of these studies, it is possible to obtain objective data on the quality of testing by using objective measurement methods. For these functional assessments, complex motion capture systems based on markers and complicated setups are required in a laboratory environment in order to obtain reliable results. Our presented method allows evaluating functional tests directly from video acquisition without the need for complex MoCap systems that use active or passive markers placed on the patient's body to determine the validity of the test. With this idea, it would be possible to capture the movements of the subjects during the functional tests using ordinary RGB cameras. A system of this type would considerably increase efficiency and reduce the cost of conventional test measurement and evaluation.

Physical therapy is becoming increasingly necessary for many individuals, resulting in rising rehabilitation costs and a demand for better patient experiences. Traditional observational analysis by physiotherapists (PTs) relies on their skills and clinical evaluations to identify deficiencies in movement expression. This study investigates the feasibility of using a camera-based system and machine learning algorithms to assess the quality of functional tests, specifically the Single Leg Squat Test and Step Down Test. Movement analysis can be tailored based on user rating preference behavior models, and OpenPose keypoint extraction can benefit from image-based feature refinement. The AdaBoost classifier's performance in predicting physiotherapists' assessments demonstrates the potential of camera-based systems and machine learning algorithms in evaluating functional tests.

Our research has made a significant contribution to the field of functional testing. One of the main contributions of our study is the novel use of only one camera and machine learning algorithms to evaluate functional tests. This approach presents a cost-effective and convenient alternative to current evaluation methods, which often require manual evaluation or the use of expensive devices. This methodology stands in contrast to previous studies conducted by Whelan et al. [113], where wearable gyro-accelerometers were utilized to assess single-leg squats, as well as Mitternacht et al. [114], who employed a similar approach to investigate lower-limb motion characteristics. Furthermore, Seifallahi et al. [115] employed Microsoft Kinect to detect Parkinson's disease based on skeletal motion parameters. By introducing our markerless methodology, we address the limitations associated with marker-based systems and wearable devices, offering a streamlined and cost-effective means of evaluating functional tests. We hypothesize that a camera-based system utilizing machine learning algorithms can provide a reliable, cost-effective, and accessible solution to assess functional tests, such as the Single Leg Squat Test and Step Down Test. Specifically, we posit that machine learning algorithms can detect errors in exercise execution that a physiotherapist could identify visually.

Methods

Our goal was to test the hypothesis that we are able to use a camera-based system to detect the correct execution of functional tests. The sequence of steps we performed to confirm this hypothesis can be seen in Fig 5.1. The block diagram presented in Fig 5.1 illustrates a study design combining expert evaluations and video analysis to classify subjects using the AdaBoost algorithm. Channel 1 is used for the training phase, while we use channel two, the patient assessment phase (scoring phase). In Channel 1, experts provide binary assessments for each individual exercise, then we check inter-rater reliability and use them for group size selection. Simultaneously, in Channel 2, the algorithm uses video recordings of subjects; the AI algorithm (OpenPose) extracts anatomical body landmarks (key points), and various features are calculated based on these key points. The outputs from both channels are combined and fed into the AdaBoost Classification algorithm to generate the final system outcome classification. This integrated approach ensures accurate and robust classification, enabling the machine to learn from expert labeling and scores while autonomously performing analysis and measurements of subjects.

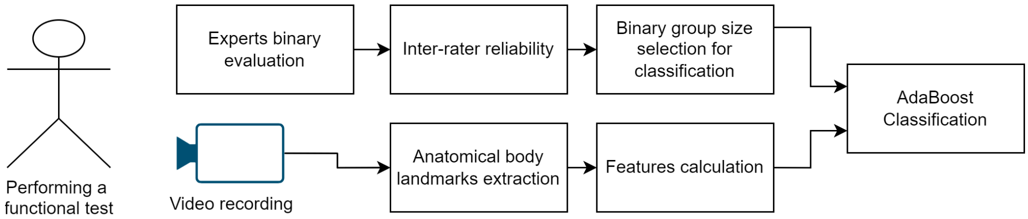


Figure 5.1: Flowchart of the study design, where we can see the camera acquisition, the expert observing the subjects, and the process till the final system outcome classification.

Participants

Forty-six subjects participated in the study: 30 women and 16 men. This was a representative sample of healthy volunteers from a group of students and academic workers from various universities in Prague. The average age was 28.2 ± 7.9 years. The mean height of the subjects was 173.8 ± 9.9 cm, and the most common weight range was 65-70 kg. Half of the subjects identified their right lower limb as dominant, 11 subjects identified their left lower extremity, and 13 subjects had no lateral preference for lower extremity as dominant. Exclusion criteria were any musculoskeletal injury in the past 6 months, acute illness, or pain while performing the test. All participants signed a written informed consent which is institutionally archived. The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the ethics committee of the Faculty of Physical Education and Sport, Charles University under reference number 167/2020.

Study Design

The study design focused on evaluating participants' performance in two functional tests, the Single Leg Squat Test (SLST) and the Step-Down Test (SDT), following a standardized procedure. Participants first signed an informed consent form and were provided with information about the purpose and nature of the experiment. They then received video instructions detailing the proper execution of the functional tests to ensure uniform guidance. During the tests, three physiotherapists simultaneously assessed participants' performance, while a precisely positioned RGB camera recorded the sessions for further analysis. The physiotherapists rated 15 possible execution errors for each test and performed a binary classification for each error (present or absent), providing an objective assessment of the participants' test performance.

Functional Test Parameters and Description

Our study examines the performance of two functional tests, the Single Leg Squat Test and Step Down Test. The performance of the used tests followed the protocol developed by Thonnard [90] and McGovern [91]. Following an explanation and demonstration of the test, the subjects were instructed to repeat the test three times for each lower limb. They always returned to an upright standing position between repetitions. Initially, the overall impression must be assessed (balance, gross arm deviation, ability to perform the test), and evaluated as satisfactory to proceed.

At least one of the three repetitions, evaluated as satisfactory, was sufficient to continue the test. The investigator then marked the following criteria as positive or negative:

1. Trunk flexion (forward lean, lateral rotation, lateral flexion, thoracic rotation)
2. Pelvic posture (tilt, rotation)
3. Hip position (adduction, internal rotation)
4. Knee joint position (valgus knee, tremor)
5. Depth of squat (compared to the other side, orientation with T)

Because the tests were not used as clinical screening for screening purposes, the overall results were not scored positively or negatively, but the parameters of each test were compared between evaluators and OP measurements. Prior to the test, the participants were given the opportunity to try both movement tasks with their right and left lower limbs once. Two images were then displayed on a screen in front of the participant. The first slide showed the initial position of the movement, while the second showed the final position of the movement.

Motion Capture Setup

For video recording, we purposefully chose a regular RGB camera: a Logitech C920 webcam with a resolution of 1280 x 720 and a frame rate of 30 frames per second, and h264 compression. In order to prove that we could make such measurements with an ordinary camera. We set this webcam on a tripod at a height of 130 cm to the top edge, and at a distance of 4m from the subject.

The camera was pointed perpendicular to the subject so that it would capture motion in the frontal plane. The green screen was placed 0.5 m behind the subject, leaving the distance between the camera and the screen at 4.5 m. Thereby we had the maximum chance of picking up body segments. For our study, we use a green screen in the background only for future use of the video dataset with systems operating on a different principle. The models used by OpenPose are trained in real-life environments. Therefore, the use of the system can also be in an environment with any background provided that there are no other people in the image that the system can also detect.

We instructed the subject to perform the aforementioned functional tests as described in the previous section. The Step Down Test was performed on a 20 cm step. Before the test, the subject had the opportunity to try both movements. They were shown two images showing the start and finish positions of both tests for ease of comprehension. Following this, the actual tests were conducted in the order of SLST, followed by SDT. The recorded videos had a length of approximately 15 seconds. A frame of the record can be seen on the right side of the Fig. 5.2.

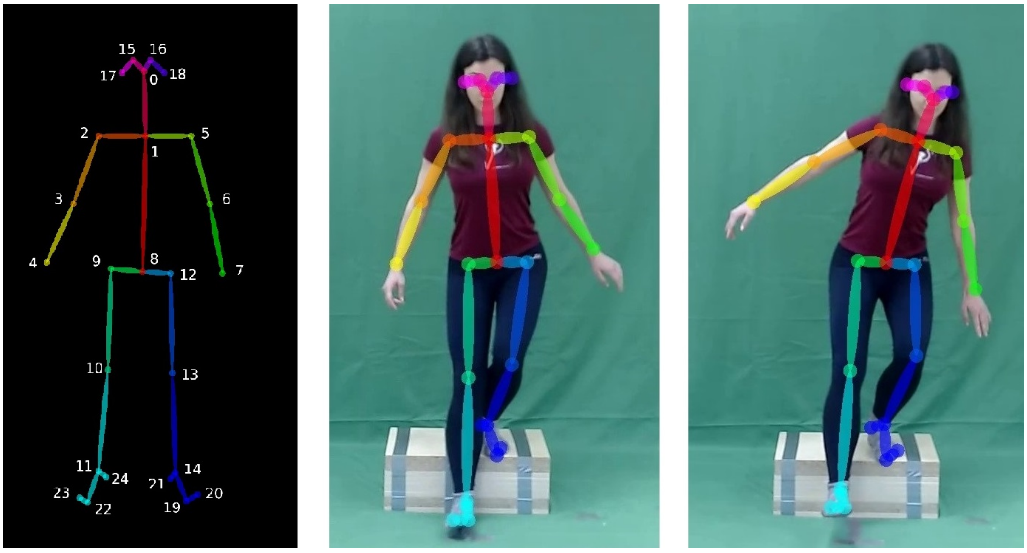


Figure 5.2: In the left part you can see the 25 keypoints (landmarks) model of the OpenPose. In the middle part is an image with the skeleton rendered while performing the functional test without any error. In the right part, the loss of balance error is clearly visible.

Expert Visual Analysis

Our study utilizes the proficient judgment of three physical therapists (PTs), each holding a Master's degree in physiotherapy and a practical experience of five years at the time of the study. These experts carried out an Expert Visual Analysis from a vantage point approximately 4.5 meters behind the recording camera, mirroring the camera's angle to scrutinize the participant's execution of the Forward-Step-Down-Test (SDT).

The selected configuration mirrors an approach supported by an analogous study from the Israeli Physical Therapy Society [116]. This study established that a rater's familiarity with the SDT significantly bolsters the agreement rate, thereby emphasizing the critical role that familiarity plays in ensuring the reliability of the test results. Moreover, the study found that the level of work experience did not influence the agreement rate. These key findings reinforce the reliability and validity of our use of the SDT as an evaluation tool, enhancing the credibility of our methodology and the anticipated outcomes of our study.

Each PT independently evaluated the presence of deviations from correct movement execution using an electronic questionnaire and a binary rating scale. The PTs' scores were evaluated for their inter-rater reliability using Fleiss's Kappa [117].

As a result, we obtained an expert evaluation, which served as a reference. We used this information for categorization in our classification algorithms. Although we may not obtain precise biomechanical motion data, we can still develop an expert semi-objective reference standard based on subjective data acquired in this manner.

However, it is essential to note the potential limitations of this approach. The binary rating scale, while offering simplicity, may not capture the nuanced differences in movement deviations. Moreover, the validity of this scale is predicated on the expertise and subjective judgment of the PTs. Future studies may benefit from using a more detailed rating scale or incorporating additional methods to enhance the validity and reliability of the visual analysis.

Signal Processing and Signal Extraction

Each of one subject's measurements resulted in a single video file containing both tests. In order to minimize distractions, the video image was further cropped to show only the subject and the green screen behind him. With OP, we were able to extract spatiotemporal information from the video files.

OP extracted a human segment model from each frame. The collection of human skeletons is used to illustrate human movement in the video. We grouped the skeletons into a table with a row for each skeleton. The skeleton is anatomically defined by a vector, in which each vector component represents the position of a vertex in the skeleton. Fig5.2 illustrates the two-dimensional Cartesian coordinates of a skeleton. The 25 key points of the model are automatically determined by OP. In our previous work, we show which anatomical points correspond to the points in the 25 landmarks model [56].

Among the detected landmarks, we calculate a signal of the range of motion in three variations. Angles and distances are calculated from the created segmental body model, which is based on identified landmarks. Custom-made software programmed in Python was used to calculate angles and distances.

Joint angles

Joint angles were calculated as the angle of three points in 2D space. The points correspond to the OP landmark model, please see Fig. 5.2. For example, the angle between the R. acromion, end of the clavicle (collar bone) top of the shoulder (A - 2), R. lateral epicondyle of humerus, lateral epicondyle of the humerus, outside of the elbow (B - 3) and R. styloid process of the radius, wrist on the thumb side (C - 4), see equation 5.1.

$$\sphericalangle ABC = \arccos \frac{\overrightarrow{BA} \cdot \overrightarrow{BC}}{|BA||BC|} \tag{5.1}$$

Keypoint-pair orientation

The Keypoint-pair orientation between two points and the horizontal of the image was calculated. The camera was in a horizontal position. Before starting the measurement, we calibrated the camera position using a laser system we developed. [118]. An example of such a calculation was the angle between the center of the pelvis (A - 8), the center of the shoulders (B - 1) determined by OP, and the horizontal plane. This information gave us information about the tilt of the trunk, see equation 5.2.

$$\measuredangle \overrightarrow{AB} = \arctan \frac{\overrightarrow{AB}_y}{\overrightarrow{AB}_x} \tag{5.2}$$

Relative distances

The relative distance was the distance in pixels divided by the distance between two points - (A - 1) and (B - 8). This normalized value can take into account the person's body height and the distance of the person from the camera, i.e. it eliminates intra-population variations in body height and inaccuracies in the distance between the camera and the body of the measured person. Range of Movement (ROM) was calculated as a difference between the minimum and maximum value of the signal, see equation 5.1.

$$|AB|_r = \frac{|AB|}{|K_1 K_8|} \tag{5.3}$$

Feature Definition

Based on the definition of functional tests [90], we selected specific elements to describe the correctness of execution, please see Table 5.1. These key points correspond to the anatomical points, please see study [56].

In this section, we describe signal processing. We used OP for video processing to extract key body points for each time and create a time series. From this time series, we calculated all presented angles and relative distances as signals. We then used these signals to extract feature parameters such as minima, maxima, medians, and means. The resulting data are available in the appendix in a readable format. This approach allowed us to analyze the subject's movements and identify specific features that are suitable for further processing by machine learning algorithms.

ROM/Feature	Category	Selected keypoints
Hips	Keypoint-pair orientation	(9,12)
BothShoulders	Keypoint-pair orientation	(2,5)
Spine	Keypoint-pair orientation	(1,8)
NeckRShoulderRElbow	Joint angle	(1,2,3)
NeckLShoulderLElbow	Joint angle	(1,5,6)
MidHipRHipRKnee	Joint angle	(8,9,10)
MidHipLHipLKnee	Joint angle	(8,12,13)
RHipRKneeRAnkle	Joint angle	(9,10,11)
LHipLKneeLAnkle	Joint angle	(12,13,14)
NoseNeckMidHip	Joint angle	(0,1,8)
NeckMidHip	Relative distances	(1,8)
RWristMidHip	Relative distances	(4,7)
RWristLWrist	Relative distances	(4,8)
MidHipLSmallToe	Relative distances	(8,13)
MidHipLRSmallToe	Relative distances	(8,10)

Table 5.1: List of selected ranges of movements(ROM), from which features were then counted.

We are operating under the assumption that if an expert observes a movement, the relevant information is necessarily captured in the video. Although the expert may not see precise angles or measurements, their experience allows them to determine whether or not errors are present in the movement. Therefore, our focus is not on obtaining a precise biomechanical description, but rather on acquiring data that can distinguish between the presence or absence of evaluated parameters, and investigating whether these derived parameters contain information regarding the correctness of execution. To accomplish this, we utilize machine learning algorithms.

Classification

We have developed a learning system whose function is to answer six research questions by determining whether a specific phenomenon occurred during the execution of the exercise - six binary answers:

1. Loss of balance
2. Gross arm deviation
3. Trunk movement: Forward lean
4. Depth of squat
5. Posture of the hip joint: Drop
6. Overall Impression: Tremor

To train the machine, we utilized a Python library called PyCaret. PyCaret is a high-performance Python library with low code that facilitates the comparison, training, evaluation, tuning, and deployment of machine-learning models. In this library, we are able to evaluate, compare and adjust standard algorithms of different machine learning algorithms in parallel on the basis of a given data set in an efficient and comparative manner.

We choose to utilize AdaBoost, a method of ensemble learning (also called "meta-learning"), to improve our binary classification performance. The AdaBoost algorithm takes an iterative approach to learn from the mistakes of weak classifiers and converted them into stronger ones. AdaBoost is a sequential learning algorithm. Successive models are generated sequentially and their errors are learned by their successors. By giving the mislabeled examples higher weights, this technique exploits the dependency between models. Just as humans learn from their mistakes and do their best to avoid making the same mistakes in the future, the Boosting algorithm attempts to create a stronger learner (predictive model) from the mistakes of several weaker ones. The purpose of boosting is to reduce the bias error that occurs when models are not able to identify relevant trends in the data. AdaBoost (Adaptive Boosting) is a popular boosting technique that aims to combine multiple weak classifiers into a single strong classifier. The definition of a weak classifier is that it performs better than random guessing but is still ineffective at classifying objects. We have implemented the AdaBoost algorithm by using ten poor decision trees, processed sequentially.

For the AdaBoost classifier design in this study, we partitioned the available data into training, validation, and testing sets, allocating 80%, 15%, and 5% respectively. We utilized a Decision Tree Classifier as the base classifier with 50 weak learners employed in the AdaBoost classifier. The learning rate was set to 0.1 and the maximum depth of the decision trees was limited to 3. The feature subset size was determined as the square root of the total number of features and we set the random state to 42 to ensure reproducibility. We initialized the weights of all data points in the training set to be equal, iteratively training a weak learner on the training data, calculating the weighted error of the weak learner on the training set, calculating the weight of the weak learner based on their performance, and updating the weights of the misclassified samples. After the specified number of iterations, the predictions of all weak learners were combined using their weights. Finally, we evaluated the final AdaBoost classifier on the validation and testing data to assess its performance. The data partitioning and parameter details allow other researchers to replicate the experimental setup and enhance the reproducibility of the results.

Results

The results section is divided into two logical units. First, we assess the agreement of individual PTs, which we express using Fleis's kappa. Based on this evaluation, we can proceed to the next section and perform classification on the groups where the agreement, respectively value of Fleis's Kappa was greater than 0.41. This value is referred to in the literature as Moderate agreement [117]. However, most of the groups achieved much higher values, see table 5.2.

Agreement of Expert Visual Analysis

In Section 5.2, we described the 15 categories that the three independent PTs evaluated. Table 5.2 shows Fleis's kappa and the size of each group. Some errors were not observed among the selected participants at all. Such groups could not then be subject to classification. Groups that were suitable for classification are highlighted in bold in the table 5.2.

Functional test parameters	**SDT [kappa (F,NF)]**	**SLST [kappa (F,NF)]**
Overall Impression: Loss of balance	**0,48 (12,34)**	0,56 (8,38)
Overall Impression: Gross arm deviation	**0,58 (8,38)**	0,73 (4,42)
Overall Impression: Disruption of smooth movement	0,37 (7,39)	**0,49 (6,40)**
Overall Impression: Tremor	0,68 (6,40)	**0,59 (10,36)**
Overall Impression: Depth of squat	**0,86 (8,38)**	0,95 (0,46)
Trunk movement: Forward lean	**0,68 (9,37)**	0,95 (2,44)
Trunk movement: Lateroflexion	0,44 (5,41)	0,65 (2,44)
Trunk movement Lateral rotation	0,87 (0,46)	0,93 (3,43)
Posture of the pelvis: Anteversion	1,00 (0,46)	1,00 (0,46)
Posture of the pelvis: Retroversion	1,00 (0,46)	1,00 (0,46)
Posture of the hip joint: Drop	**0,45 (32,14)**	0,84 (5,41)
Posture of the hip joint: Shift	0,24 (9,37)	0,92 (0,46)
Lower limbs: Hip joint, Adduction, internal rotation	0,15(27,19)	0,13 (30,16)
Lower limbs: Knee-joint, valgosity	0,27 (20,26)	0,07(22,24
Lower limbs: Knee-joint, varosity	1,00 (0,46)	1,00 (0,46)

Table 5.2: Agreement of the expert ratings of the three physiotherapists. The agreement is expressed by Fleis's Kappa coefficient. In the literature [117], a kappa greater than 0.41 is considered sufficient. The frequency for each of the binary categories is given in parentheses after this value. The number of subjects who did the given error is labeled as FAULTY and NON-FAULTY respectively. Values subjected to classification are in bold.

By using the Kappa coefficient to assess agreement between raters, we established a reference standard for classification that divides the data into those that can be used - those with high agreement - and those that cannot be used due to insufficient agreement among the experts. We followed standard methods to ensure the validity and reliability of our approach. Specifically, we used well-established procedures for data collection and analysis, and we implemented rigorous quality control measures to ensure that the results were accurate and reproducible. Overall, our approach provides a robust framework for classifying data and enables us to identify reliable and valid results for further analysis.

Results of Classification

Our results show that if two or more raters agree on the presence of error, and there is a sufficient amount of measured data with and without error, our method based on a single 2D camera produces results comparable to the raters themselves. In the event that even three physiotherapists do not agree on the correct result, then it is likely that the gold standard for machine learning methods cannot be established. Due to the fact that our experts were not in agreement on all measurement categories, we were forced to exclude some of the categories from the evaluation. As a consequence, the experts' assessment is still a subjective view and can differ depending on their life experience and the weight they give to individual errors.

For classification, we selected only the datasets fulfilling the conditions of kappa coefficient and minimum sample size, see Table 5.2. These were two separate datasets. Five responses (functional test parameters) to the first dataset (SDT), which contained 15 independent variables, met the conditions. The second data set (SLST) contained 15 independent variables and 2 responses from our panel of expert examiners. We calculated a confusion matrix for each dependent variable.

Our focus was on sensitivity and specificity. It was quite difficult to find subjects who did the movement incorrectly in our sample since only healthy people were included in our sample, according to our experts. In the end, our database is imbalanced because we have included only healthy subjects, the majority of the results showed that the patients were able to perform the exercises correctly. Therefore, we have implemented conventional methods in order to deal with the data imbalance. In this article, we wish to emphasize that since the purpose of the article is to prove the applicability of the method, we have chosen to present a number of cases in which a balance can be illustrated in the analysis. We are therefore very concerned with the sensitivity and specificity of our results in the results section. According to these two metrics, it can be seen that the data we finally used is balanced on the one hand, and on the other hand, they show that our method is able to detect any error by the subject with an average probability of 0.68 (Sensitivity) and that its prediction reliability (i.e. error reporting) is 0.8 (Specificity). The results are displayed in the table 5.3.

Functional test parameters (N=46)	Specif.	Sensit.	Accuracy
Step down Test (SDT)			
Overall Impression: Lost of balance	0.95	0.5	0.86
Overall Impression: Gross arm deviation	0.88	0.75	0.86
Trunk movement: Forward lean	0.55	0.67	0.57
Overall Impression: Depth of squat	0.83	0.4	0.75
Posture of the hip joint: Drop	0.38	0.9	0.75
Single Leg Squat Test (SLST)			
Overall Impression: Lost of balance	0.92	0.5	0.86
Overall Impression: Tremor	0.92	0	0.82
Average values	**0.80**	**0.68**	**0.77**

Table 5.3: Summary of the AdaBoost classification for the two exercises SDT and SLST. In the case of the SDT, it presents five dependent outputs, while for the SLST, two variables are dependent. Dependent variables consist of binary classifications. The quality of classification is determined by the accuracy, specificity, and sensitivity of the classification. The sample is 46, for details, see table 5.2.

We used the AdaBoost Classification algorithm for our data analysis, which is a powerful ensemble learning tool for binary or multiclass classification problems. It integrates many weak classifiers to construct a robust classifier that provides accurate and exhaustive data analysis insights. Our approach trained weak classifiers, such as decision trees, on the weighted training set at each iteration to reduce classification error. We adjusted the instance weights by assigning incorrectly classified cases with greater weights and correctly classified instances with smaller weights, ensuring that subsequent weak classifiers focused on increasingly challenging events. The robust classifier produced predictions by weighing the votes of all weak classifiers, with the final forecast based on the class with the highest weighted votes. The AdaBoost algorithm is effective at delivering precise data analysis insights by leveraging the skills of several weak classifiers and focusing on difficult situations during training, resulting in a generalizable model for new, unobserved data.

Discussion

The aim of our research was to explore the feasibility of using a camera-based approach to automate the standard examination of two functional tests: the Step-down Test (SDT) and the Single-Leg-Stance Test (SLST). A common practice for evaluating these tests is also visual analysis[93], done by experts. According to previous studies, 2D and 3D retroreflective marker analysis[92] is capable of achieving the same results in the frontal plane. In related research, Wouter [119] at. col. uses marker-based 2D analysis to evaluate functional tests of range-of-motion. Remedios et. col [120] compare the absolute differences between the 2D markerless analysis and the 3D marker analysis of the functional load lifting test. In their study, the 2D analysis shows significant bias for the peak values of the ranges of motion. This study records motion in the sagittal plane, which performs worse in detection compared to the frontal plane [50]. In our study, we measure only the frontal plane. In our case, we build a model that does not evaluate the performance using absolute values, but a model that evaluates according to the experts' responses. Thus, it is not essential for us to obtain absolute values of the angles, but the derived values of the 2D model will suffice. This is where our approach is unique and we can afford to compensate for any inaccuracies of the 2D markerless system. This brings an innovative approach that combines modern image processing techniques with expert knowledge. Our interdisciplinary solution enables the transfer of knowledge and experience into automated processes. The result is a comprehensive subject motion analysis that can be used to improve outcomes in clinical applications. Our method brings new possibilities for the diagnosis and treatment of movement disorders and can serve as a supportive tool for physiotherapy practice and other disciplines using movement analysis.

Based on the results of our study, we believe that the design of an assistance system based on our approach is a promising area for future work. Further data for learning and expert evaluation will be necessary to improve the system. This is a significant challenge that requires an interdisciplinary approach and close collaboration with experts from different fields. However, the relatively low cost and scalability of such a system allow the development process to be accelerated and deployed in many places simultaneously.

In the next phase of our research, we plan to strengthen our expert base and prepare our software for mass subjective evaluation. This will improve the whole system and involve more participants and evaluators. We plan to enable remote expert evaluation from video recordings, which will lead to a more robust evaluation of exercise execution.

Overall, our goal is to create a system that can aid in the diagnosis and treatment of movement disorders while serving as a support tool for physiotherapy and other disciplines that use movement analysis. We believe that our interdisciplinary approach and collaboration with experts will lead to significant advances in this field, allowing us to meet the challenges and accelerate the development of an effective assistive system.

Conclusion

Traditionally, functional tests have been evaluated either visually by experts or through expensive automated systems that require human interaction. However, both of these options come with significant costs in terms of time, money, and resources. To address this issue, we propose a novel and cost-effective solution that combines modern computer vision techniques with deep learning algorithms and the expertise of physiotherapists. Our method offers

a more accurate and efficient way to evaluate functional tests, without the need for costly equipment or extensive human involvement. Our approach can detect errors when performing functional tests, allowing for a more comprehensive assessment of performance quality. Further research will allow us to determine the weighting of these parameters to accurately evaluate the overall quality of performance. Our proposed classifier has a high level of accuracy (0.77 on average) due to the heterogeneous data with and without exercise errors and reliable agreement between physiotherapists. Our findings are very encouraging regarding the feasibility of the camera system for use in the home environment. However, we do not believe that these methods can completely replace the work of physiotherapists. Rather, we consider these methods to be useful complementary tools to physiotherapy. We believe that through further research and collaboration with experts in the field, our approach can bring significant advances in the diagnosis and treatment of movement disorders. This study's contribution to physical therapy lies in demonstrating the effectiveness of a simple and cost-effective camera-based system combined with machine learning algorithms to evaluate functional test performance. This approach has the potential to significantly increase efficiency and reduce the cost of conventional test measurement and evaluation, expand the range of assessment tools available to physiotherapists, and potentially improve the accuracy and reliability of their evaluations.

Chapter 6

OffiStretch: Camera-based Real-time Feedback for Daily Stretching Exercises

6.1 Introductionary Comments

One of the defined goals was to create applications with real-time feedback with an augmented mirror. I developed such an application and named it Offistretch. This application utilizes a camera-based system, making it accessible to a wide range of users in a home environment. I submitted an overview of the application and the results of a user study as an article to the Visual Computer journal, published by Springer, under the title "OffiStretch: Camera-Based Real-Time Feedback for Daily Stretching Exercises" [121]. This article has been officially published on May 28, 2024. This paper was submitted with me as the lead author and primary contributor. All co-authors have signed an Acknowledgment of Contribution and Authorship form, thereby confirming that the main results and findings of the paper are derived from my research and work. For detailed information, please refer to Appendix A.

6.2 Introduction

The tendency for home office work has strongly increased due to the pandemic and will likely persist also in the future. This trend impacts peoples' level of physical activity, as they lack movement related to their commute to work, in-person meetings, and social activities with coworkers, but also by using office equipment and furniture that is not ergonomically optimal [122]. It is commonly known that physical inactivity and a sedentary lifestyle can have negative consequences for the general population [123, 124, 125]. As confirmed by our survey results, people are aware of this negative impact on their health, but they do not have enough immediate motivation and personal discipline to exercise.

Motivation can be increased by integrating gamification elements into physical exercise routines, as has been recently studied in connection with video games [126] and exergames. For example, Pacheco's review [127] compares 12 user studies with older participants, concluding that exergames can significantly improve motivation, balance, and mobility. Andrade [128] reviewed studies related to children and adolescents with obesity and reported improvements in self-esteem and self-efficacy through the use of exergames compared to control groups. Soares [129] explored the effect of exergames on the cognitive abilities of older adults compared to conventional exercise. While he found no effect on cognitive function, the use of exergames seems to positively impact motivation. This is supported by Stadiano's study [130] on the development of motivation through exergames.

Fitness trackers represent a further important factor, as they can help to improve motivation [131] by indicating progress towards reaching one's training targets based on measured physical activity and tracked human movements[132]. Fitness trackers usually use GPS, inertial, and physiological sensors for tracking motion and exertion, to provide users with an estimate of their total physical activity during the day. Such systems can provide very good real-time or aggregated values of various biometric properties, like heart rate, step size, or running speed [133]. However, they cannot evaluate whether the user's run was biomechanically correct or not. The same problem appears in the case of other exercises like workouts, where information from an inertial sensor on the user's wrist is not enough to analyze the correctness of the movement for the best outcome and to prevent injury. So, while these devices are great to bolster motivation, they are limited with regard to accuracy for full-body movement measurements.

To address this issue for stationary exercise forms, we propose a vision-based approach using off-the-shelf components for evaluating the correctness of the user's pose based on joint angles and distances between selected body parts. Additionally, we introduce interactive visual feedback that continuously indicates the correctness of the user's pose in a simulated "digital mirror". The digital mirror metaphor is realized using a regular screen that shows the mirrored live capture from a camera. We apply this approach in the context of stretching exercises, where we explore its potential for coaching users to stretch correctly and increasing their motivation for daily activity.

The methods for digital assistance in sports and well-being should always be accompanied by comprehensive studies that investigate whether they affect users in the desired way. Therefore, we first conducted an online survey to investigate users' needs and preferences regarding digital coaching systems for stretching. This was followed by an on-site user study, in which we evaluated users' performance and motivation in performing stretching exercises with and without our visual feedback. Finally, we validated our methods through an expert evaluation with professional physiotherapists.

The main contributions of this paper can be summarized as follows:

- We present a vision-based pose analysis approach using only a single RGB camera.
- We propose a visualization technique for live feedback to indicate pose accuracy.
- We identify user needs and preferences for digital stretching coaches (online survey).
- We report findings on the impact of our feedback on motivation and stretching performance (user study).
- We validate our approach and highlight directions for future work (expert evaluation).

6.3 Related work

Our work builds primarily on research and developments in two major fields. Thus, we will first review body tracking technologies (Section 6.3.1), and then we describe methods for visual feedback in physical training (Section 6.3.2).

6.3.1 Body Tracking

The basis of every interactive method for human motion analysis is a motion capture system. We can divide these systems into three basic categories: (1) user instrumentation with active sensors, (2) marker-based tracking, and (3) markerless camera tracking.

Most active sensors (e.g., wearable or hand-held) are based primarily on inertial sensors [134] that can detect changes in the users' motions. Such sensors, as may be integrated into the smartphone or more recently a smartwatch, have the benefit of being usable in mobile scenarios, without requiring a fixed and calibrated lab installation. However, they are not capable of delivering absolute positions and therefore are subject to drift. Hybrid approaches exist, for example, the hand-held Nintendo game controller known as the Wiimote, for which tracking accuracy could be improved by complementing the inertial measurements with optical tracking through the integrated infrared camera and an extra infrared-emitting sensor bar.

Most marker-based systems usually involve optical tracking with specialized cameras, where the markers may either be passive (i.e., reflecting light) or active (i.e., emitting light). The most accurate marker systems are those used in laboratory conditions such as OptiTrack[1], Vicon[2] and Qualisys[3]. These systems can achieve 6DOF tracking with sub-millimeter accuracy.

The most common sensors in the markerless category are depth sensors such as the Microsoft Kinect[4]. These devices are generally more affordable and simple in their use than marker-based systems, while not requiring any instrumentation of the user.

A functional feedback exercise system using Kinect is the YouMove app created by Anderson and colleagues [135]. The users see themselves in a simulated mirror and they are guided by visual indicators in the image of where to move which limb. If a user reaches the target pose with sufficient accuracy, they are prompted by the system to stay in the position. These systems have the common disadvantage of requiring special hardware. In contrast to that, our approach requires only a standard RGB camera (e.g., webcam or smartphone camera). Surprisingly, even though exergames have been researched extensively, very few RGB camera-based systems can be found in the literature. Losilla and Rosique [136], Kanase et. al. [137] or Hesham et al. [138] follow a similar approach, however, do not contain visual feedback and analysis of the current user pose.

Coyler et al. review the evolution of camera-based motion analysis until 2018 [139]. Since then, body pose detection methods based on deep neural networks have been predominantly used with common examples being OpenPose[1], Alphapose [40] and Media-pipe [**?**]. Badiola-Bengoa and Mendez-Zorrilla discuss the use of such approaches for sports and physical exercise [140].

[1] OptiTrack: `https://www.optitrack.com/`
[2] Vicon Nexus: `https://www.vicon.com/software/nexus/`
[3] Qualisys: `https://www.qualisys.com/`
[4] Kinect: `https://developer.microsoft.com/kinect/`

6.3.2 Visual Feedback for Physical Training

Beyond gaming, domains such as fitness, health, and well-being have actively adopted new technologies, for example for tracking physical exercise and displaying the user's real-time exertion and daily activity on a smartphone or watch. The availability of compact and portable displays has led to a wide variety of visualizations for training progress, from displaying step counters or traveled routes on a map, to ECG-like heart rate visualizations (e.g., on the Fitbit[5]), or "rings" on the Apple Watch[6]. Such visualizations address peoples' craving for a sense of progress and achievement, as well as monitoring their own health and performance. While these visualizations can reflect a user's progress toward their set training goal, they usually provide only aggregated data and do not analyze the poses of individual body parts during the motion to assess their correctness. Failure to do so may lead to less effective workouts and can even risk adverse effects such as physical injury. This aspect may be addressed by live visual feedback of the user's posture and motion, which has been found to positively impact mood [141] and physical well-being [142], and can guide the user to perform movements correctly as is critical for a range of sports like dancing [143, 144], TaiChi [145, 146], or Tennis [147]. Arguably, an increased number of tracking points and accuracy of pose reconstruction can support this better (e.g., approximation of motions in Ring Fit Adventure[7] vs. accurate full body tracking [146, 148],[147]). Related research has explored a variety of different feedback visualizations, with most common designs involving a kind of mirror image [141, 142][149],[150] a third person perspective of oneself [143, 145, 144], or superimposed feedback on the body seen from first person perspective [148, 145].

Similar research to ours is the work of Elsayed et al. [138], who describe the current trends in motion capture systems and their use for home exercise. They compare three different forms of feedback for matching the user's pose with a static posture: a silhouette, a skeleton, and a predefined avatar. An evaluation of this system revealed poor visibility of participants' own bodies through the displayed skeleton, a lack of feedback about which body part was not oriented or positioned correctly, and a lack of audio feedback. Second work named Pose Tutor by Dittakavi [151] et. al. can detect and compare trainee position with predefined position based on The k-nearest neighbors algorithm. This system is a position comparator rather than a complex exercising application. Another similar approach is 3D camera-based system called AIFit and presented by Fieraru et. al. [152]. However, this system uses multiple cameras and the feedback cannot be overlaid directly into the image.

[5]Fitbit ECG: `https://www.fitbit.com/en-ca/technology/ecg`
[6]Apple Watch rings: `https://www.apple.com/watch/close-your-rings/`
[7]Ring Fit Adventure: `https://ringfitadventure.nintendo.com/`

6.4 The OffiStretch System

In this section, we describe our methods for pose analysis and visualization. Additionally, we provide details about the design and development of our application. The name OffiStretch hereby reflects our motivation to encourage and provide interactive guidance during stretching in the (home) office. By comparing the captured stretching pose (from the video stream) to the predefined target pose (static position), we assess the correctness of the user's stretching performance. The result is visualized to the user as a live video stream with visual feedback on an augmented digital mirror.

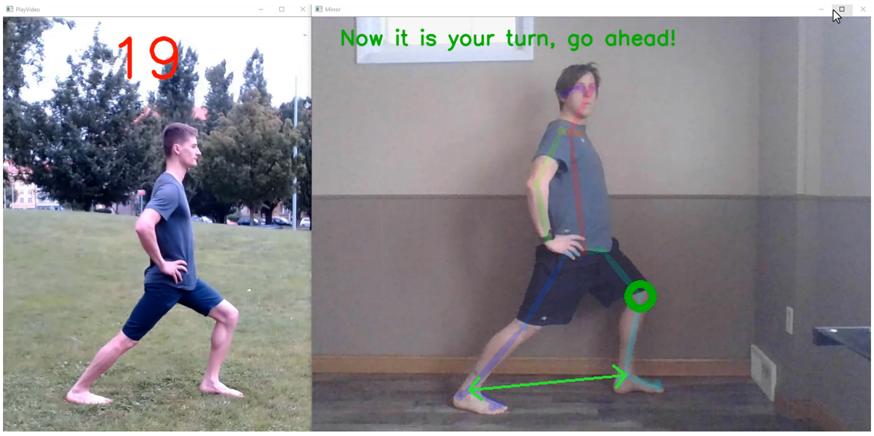


Figure 6.1: Screenshot of OffiStretch application with real-time dynamic feedback drawn onto trainee's own body. The arrows encourage the trainee to extend the stance and the green circle encourages greater flexion at the knee joint. The closer the practitioner is to the desired position, the smaller the circle or thinner the line is.

6.4.1 Body Tracking and Pose Assessment Features

Our application uses the OpenPose [1] system to detect the human skeleton. This approach utilizes image recognition using a deep neural network. To reconstruct the user's pose, the system attempts to match patterns for 25 individual human body parts (keypoints) in the input image. For each of these keypoints, shown in Figure 6.2, the system builds probabilistic heatmaps based on the typical human motion range and then reconstructs the entire human skeleton from these relative keypoint positions. The OpenPose system thereby achieves very high estimation accuracy, with errors in measured angles reported between 0.19° (pelvis joint) to 3.17° (right shoulder) [153].

The advantage of the system is its resistance to light conditions or video quality and requires only minimal setup [1]. The system can be used almost anywhere with any camera. The single condition for successful pose detection is that no other person or image of a person (photograph, poster, drawing) is simultaneously in view.

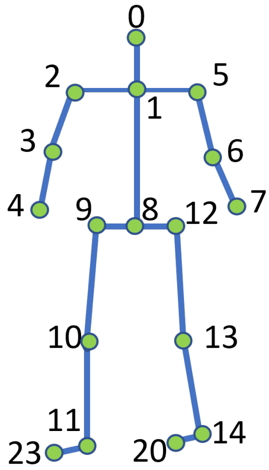


Figure 6.2: A model showing 17 key points that we use to calculate features characterizing human posture. We use the same keypoint indexing as the original 25-keypoint OpenPose model [1] from which our model is derived.

Due to the use of a single static camera, the user's body pose is captured in 2D space. Thus, the trainee must perform the exercises allowing the image sensor a clear (frontal or profile) perspective of the body. We achieve the correct orientation of the trainee to the camera by showing the trainer's video as a guide. The video of the trainer is presented next to the simulated mirror where users can see themselves. Hence, trying to mimic the trainer's posture in the mirror leads users to orient themselves correctly. This method of pose matching is already well-known from previous work [143, 145, 148, 144].

Finally, pose matching is performed as a real-time comparison of the defined target pose pre-recorded by the trainer (reference pose) with the tracked pose of the trainee. The static target pose is described by a number of parameters consisting of the following three measurements: the angle between three keypoints (*joint angles*), the screen-space orientation of the vector between two keypoints (*keypoint-pair orientation*), and the *relative distances* between keypoint-pairs. Details on their computation are provided below and the visual description of these three types of features can be seen in Figure 6.3.

Joint angles

The angle between two vectors, constructed by connecting three keypoints A, B, and C, serves as a basic parameter to describe their mutual constellation. This can be used to reflect tracked joint angles, as seen from the camera perspective. For example, the degree of flexion of the elbow is measured as the angle between the upper arm and forearm, which is described by the keypoints shoulder (A), elbow (B), and wrist (C). This angle is computed in the 2D Cartesian coordinate system by using the dot product as follows.

$$\sphericalangle ABC = \arccos \frac{\overrightarrow{BA} \cdot \overrightarrow{BC}}{|BA||BC|} \tag{6.1}$$

Keypoint-pair orientation

In everyday life, we commonly refer to the horizontal or vertical axis to describe the correct orientation of a body part, which we can formalize based on the relative orientation of keypoint-pairs. For example, the T-pose is commonly understood as a vertical alignment of the spine (e.g., the vector from neck to pelvis: keypoints 1 and 8 in Fig 6.2), straight vertically aligned legs (i.e., vectors between hips and feet: $v(9, 11)$, $v(12, 14)$), as well as horizontal alignment of both arms (i.e., shoulder to wrist: $v(2, 4)$, $v(5, 7)$). Assuming perfect horizontal alignment of the camera, the following equation 6.2 defines the 2D direction of the vector for the keypoint-pair (A, B) in relation to the horizontal (x) axis.

$$\measuredangle \overrightarrow{AB} = \arctan \frac{\overrightarrow{AB}_y}{\overrightarrow{AB}_x} \tag{6.2}$$

Relative Distances

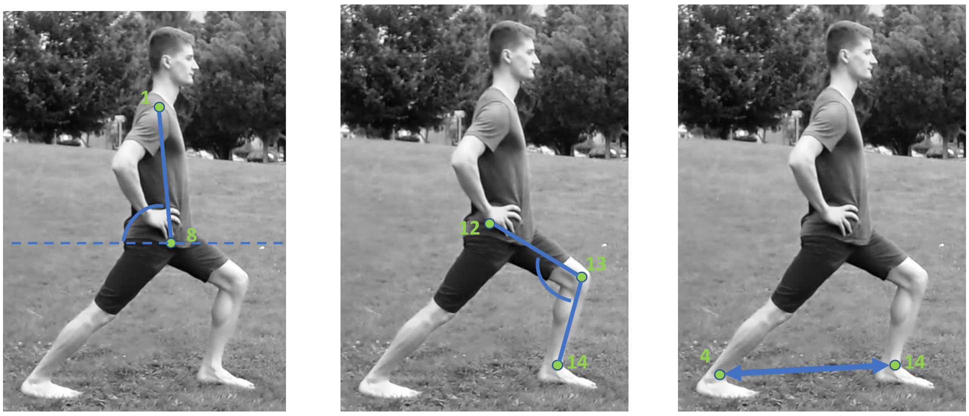
Apart from angles and orientations, distances also play an important role in describing body poses, e.g., placing one's feet hip distance apart. To normalize measured distances between keypoints by a user-specific proportion, we calculate relative distances with respect to the user's spine length: the following formula describes the distance between two keypoints A and B divided by the distance between keypoints K_1 and K_8 (i.e., the keypoints 1 and 8 in Fig. 6.2), measured in pixels. Due to this normalization, we do not need to consider the user's height or distance from the camera when calculating similarity to the reference pose.

$$|AB|_r = \frac{|AB|}{|K_1 K_8|} \tag{6.3}$$

Static Pose Description Features

We can describe each human pose by calculating a number of parameters from the 17 keypoints, based on the measurements described above. Through various keypoint combinations, based on experts' discussion, we defined 109 pose features to represent any body pose: 60 relative distances between keypoint-pairs, 25 joint angles (between three keypoints), and 24 keypoint-pair alignments. The list of all defined features may be found in the supplementary material. Each pose for a given exercise can be stored as a feature vector F:

$$F = (a_1, ..., a_M, b_1, ..., b_N, l_1, ..., l_K) \in \mathbb{R}^{M+N+K} \tag{6.4}$$



(a) : KP-pair orientation **(b) :** Joint angles **(c) :** Relative distances

Figure 6.3: Three types of features used in our pose assessment.

Where:

- M: is the number of joint angles (25)
- N: is the number of keypoint-pair orientations (24)
- K: is the number of relative distances (60)

However, only a subset of these features is used to asses body pose correctness for each exercise. Table 6.3 defines individual selections of features for exercises used in our study. For example for exercise Arm Prayer Stretch (APS) M = 2, N = 1, K = 1. These subsets were defined in consultation with physiotherapists, based on the most relevant and prominent body part configurations required for each exercise.

6.4.2 Exercise Instruction Authoring

The authoring of instructions for a new exercise is achieved simply by including a new video recording of a trainer performing the exercise. Importantly, when recording, attention must be paid to the correct orientation of the trainer in relation to the camera position to ensure good visibility of relevant keypoints for accurate body tracking. The target pose features for the given exercise are then computed from a single manually selected frame in the video, where the trainer is in the static target pose, performing the full stretch. In the system, each exercise is then stored as a video and configuration file. The latter contains details such as the video name, frame, and all 109 descriptive parameters for the target pose. As mentioned before, only a few of these features describe the exercise, while others may not be accurately detectable due to the user's orientation, or can be considered irrelevant for the particular exercise (e.g., elbow angles may be irrelevant for the calf stretch, but critical for the lower arm stretch). This set of most relevant features is manually selected (ideally by professional physiotherapists) and recorded in the configuration file. Pose features vary across exercises, but typically each exercise is described by three to five pose features. These selected parameters are then used to evaluate the error between the trainee's pose and the target pose, which results in the real-time pose assessment that can be visualized using visual feedback explained below.

6.4.3 OffiStretch Visual Feedback

The visual user interface is intended for presentation on a PC monitor or TV screen. The GUI of our application contains two main windows (Figure 6.1): The left window shows the video clip of the trainer with a superimposed countdown and other information about the exercise. On the right side, the users can see themselves in a webcam-simulated digital mirror. To ensure a correct perspective, the camera must be mounted on the respective display.

Each exercise begins with a brief prerecorded verbal explanation of the exercise and a loop of the instruction video showing the trainer performing the stretch. Then, the user is informed that it is their turn to start the exercise (through voice recording and text as shown in Figure 6.1). In this phase, the left window shows a still frame of the trainer in the target pose and a countdown indicating the duration for which the stretching pose should be maintained. Meanwhile, the webcam-simulated digital mirror is augmented with feedback elements, to guide the user to improve her/his pose in real time. Further, every 5 seconds the user receives audio feedback in the form of a voice recording commenting on whether the body pose is correct (within the defined tolerance levels), or needs further adjustment. After the timer has run out, the system starts a new exercise.

The presented elements of visual feedback depend on the chosen set of pose features for which errors are computed in each exercise. We use the following two types of visual feedback to display these errors, as illustrated in Fig. 6.1:

Circles. Any errors in angle (i.e., joint angles and keypoint-pair alignment) are indicated by a circle that is centered on the second keypoint. The size of this circle reflects the magnitude of the difference between the trainee's pose and the target's pose. As the trainee adjusts their pose, the circle gets smaller or larger, conveying whether or not the actual pose is getting closer to the intended pose. The circle disappears when the joint angle or keypoint-pair alignment is correct (within the defined tolerance threshold which was experimentally set to 3 degrees).

Lines with arrows. Error in the relative distance between two keypoints is visualized by a line drawn between them. The magnitude of the error is reflected by line thickness: a thicker line indicates a greater mismatch. Arrow tips at the end of the line indicate in which direction the key points should move to correct the pose. Further, if the distance is smaller than desired, the line is colored red, and green if it is too big. As the trainee adjusts the pose, the lines are updated in real-time, reflecting the progress toward correct stretch execution. As with the circles described above, the lines also vanish once the correct target distance (within the defined tolerance threshold, which was experimentally set to 0.2) is achieved.

6.4.4 Hardware and Software Requirements

The core component of our system is OpenPose [1], with which real-time processing is possible, albeit computationally demanding. Using a laptop with Nvidia GTX 1070 GPU we achieved 16 fps. Application of our approach for a more dynamic exercise or running the system on a low-performance device can reasonably be assumed possible, but it would require optimization of the way we compute the keypoints. Possible options include cloud processing of data or using one of the frameworks designed for lower-performance devices such as Google Tensorflow Lite[8].

Selected Survey Questions

1	2	3	4	5	6	7
Q1: How often do you do stretching exercises? (This may be as part of a longer workout, or alone.)						
O	O	O	O	O	O	O
never	less than once a month	at least once a month	at least once weekly	multiple times per week	once per day	multiple times per day
Q2: Imagine a display that gives you real-time visual feedback about the quality of your stretching. How often can you imagine dedicating a few minutes to stretching exercises with such a coaching system during working hours (e.g. in a break)? You may assume this is approved/encouraged by management.						
O	O	O	O	O	O	O
never	less than once a month	at least once a month	at least once weekly	multiple times per week	once per day	multiple times per day
Q3: I would be willing to try a system that reminds me to stretch and instructs me on particular exercises for relieving body strain from a prolonged working pose (e.g., seated at desk, standing at workbench for many hours).						
O	O	O	O	O	O	O
strongly disagree			undecided			strongly agree

Table 6.1: Questions of our online survey, used to evaluate our two hypotheses. Each question was answered twice (once for home office condition and once for dedicated workplace condition). Answers were listed in the opposite order in the questionnaire and we inverted them for consistency of visualization within the publication.

6.5 Online Survey: Stretching in the (Home) Office

During the design and development of our system, we conducted an online pre-study to investigate the stretching habits of people and their willingness to use an interactive system for stretching guidance. We asked participants

[8]Tensorflow lite: `https://www.tensorflow.org/lite/examples/pose_estimation/overview`

to consider two particular conditions: working (1) in their home office and (2) at their dedicated workplace. The study was designed as an online survey with quantitative and qualitative items We aimed to study the following two hypotheses:

- **H1**: People do more stretching exercises during the day when working in the home office compared to their dedicated workplace.
- **H2**: People would prefer to try using interactive stretching guidance in their home office compared to their dedicated workplace, and could also imagine doing so more frequently at home.

The questionnaire was answered twice by all participants (within-groups design), with fixed order of scenarios: First, the questions were asked about the home office and then about the dedicated workplace. Our H1 was addressed by question Q1, while Q2 and Q3 allowed us to explore H2 (see Table 6.7). Further, demographic information was collected and open questions were asked to investigate exercising habits and awareness of the negative effects of a sedentary lifestyle on participants' health and well-being. It should be noted that survey participants were asked to imagine a system that interactively provides stretching guidance on a display, but we did not specify exactly how this system should work or what it would look like. Hence, the details about the systems they envisioned may differ, e.g., based on their prior experiences with smart mirrors or body tracking games. However, as we merely aimed to assess participants' general willingness to use a guidance system based on display technology, we deem these potential differences irrelevant.

6.5.1 Online survey participants

We collected 90 survey responses from 55 men and 35 women. The age distribution of participants in predefined age groups was the following: 9 people in the group between 18-25 years, 28 people in groups 26-33, 20 in 34-41, 13 in 42-49, 7 in 50-57, 9 in 58-65, and 4 participants in a group over 65 years.

More than 90% of the participants indicated a job in academia with low physical demand and many sitting hours. With regards to nationality, 39 participants came from Czechia, 14 from Slovakia, 12 from Austria, 6 from Denmark, and 22 from other countries. Participants who could not respond to questions in both conditions (16/90), because they had no experience of working both in home office and their dedicated workplace, were excluded from the following quantitative analysis.

6.5.2 Online survey results

Stretching activity and coaching preferences

Statistical analysis by Wilcoxon signed-rank test was performed on the quantitative responses to Q1, Q2, and Q3 (see Table 6.7) to explore our

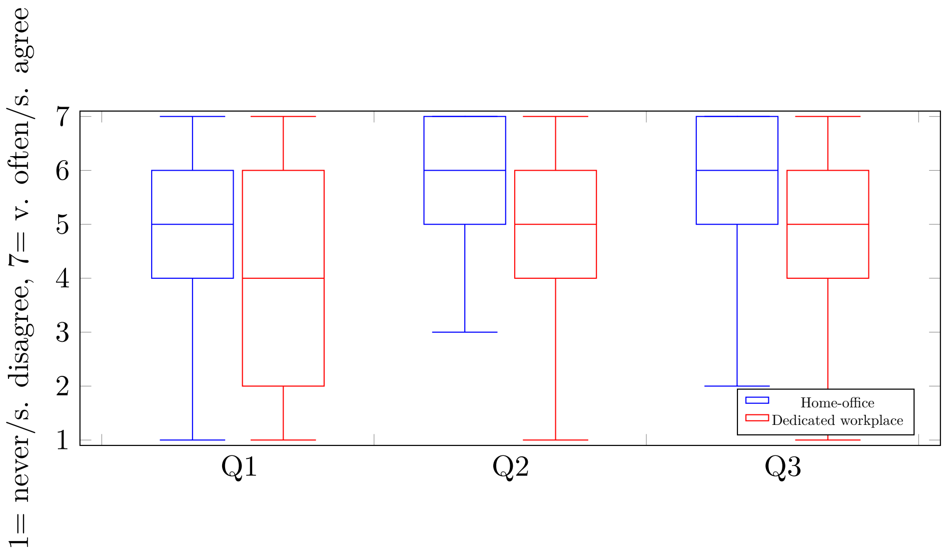


Figure 6.4: Responses to Q1 reflect how often participants perform stretching exercises, Q2 indicates the preferred frequency of stretching with an imaginary coaching system and Q3 reveals participants' willingness to use a coaching system that reminds and instructs them to do stretching. The home-office scenario is presented in blue (left side) and dedicated workplace in red (right side). For more details see Table 6.7.

hypotheses (H1, H2). For all three questions participants' responses, visualized in Figure 6.4, differed significantly between conditions: participants indicated that they performed stretching exercises significantly more often in the home office (median = 5: "multiple times per week"), compared to their dedicated workplace (median = 4: "at least once weekly") (H1). Further, they could also imagine using a digital coaching system more frequently at home (median = 6: "once per day") compared to the workplace (median = 5: "multiple times per week"), and they responded with higher willingness to try such a system that reminds and instructs them to stretch in the home-office scenario (median = 6) compared to the dedicated workplace (median = 5) (H2). While this supports both our hypotheses, it should be noted that responses were very positive for both scenarios, generally indicating healthy stretching habits and high acceptance of using a digital coach. Detailed results are provided in Table 6.2.

Reported health issues, risks awareness and exercising habits

In response to open questions, participants reported about existing health issues, their knowledge of the potential effects of a sedentary lifestyle, and provided details on their exercising habits while at the workplace or home office. We coded and analyzed this data in MaxQDA software. The codes were grouped into 3-7 themes per question[154].

Of the total 90 participants, 19 reported preexisting *diagnosed health conditions*. The most common were pain or mobility issues in the back

Online Question	**Z**	**p**
Q1 - frequency of user's stretching exercise	-4.02	<**0.001**
Q2 - preferred frequency of stretching with a coaching system	-3.83	<**0.001**
Q3 - willingness of trying a coaching system for stretching	-4.60	<**0.001**

Table 6.2: The results of significance assessment by Wilcoxon signed-rank test. The significance of differences between home office and dedicated workplace conditions was assessed for each question from Table 6.7.

(9), shoulders (4), and knees (3). When asked whether they were *aware* of any possible physiological problems caused by a sedentary lifestyle, 61/90 participants gave a positive answer. As examples they listed back pain (24), neck pain (11), wrist issues (11), pain in other joints (6), headache (5), and in lower numbers also heart and blood circulation problems, mental health issues, etc.

In the questions asking about the participants' exercising habits, sources of exercising tips, and obstacles preventing them from exercising, the answers varied depending on the scenario (home office, dedicated workplace). The findings from coding the open questions explain the results from Q1-Q3: People prefer to exercise outside of a dedicated workplace because they do not feel comfortable exercising in front of their *coworkers*, as one of the participants stated: "*I would feel weird doing stretching in the office with my colleagues present.*" This reason for not exercising at their dedicated workplace was listed by 25/90 people - (22.5%). Other obstacles listed for both home office and workplace were related to *personal discipline* (laziness, lack of motivation, non-existing routine, and forgetting to stretch) with 39.6% of received answers for home office and 22.5% for the dedicated workplace. *Workload* or tight schedules were also mentioned for both scenarios (25.2% at home, 29.7% at work). Unsuitable *space* was predictably more often mentioned for the dedicated workplace (14.4%) than at home (3.6%).

From these answers, we conclude a high willingness for stretching with a digital coach. We expect that OffiStretch could help people to exercise especially in their home office setting, where several limitations (coworkers, space) are absent and the coaching system could help with motivational aspects (personal discipline).

6.6 User Study

Upon completion of our OffiStretch prototype, we performed a lab study to evaluate the overall functionality, motivation impact, and potential of our proposed digital coach. In particular, we aimed at exploring the effect of our live visual feedback on users' motivation and performance in stretching.

6.6.1 Study Design

To evaluate our methods for motion assessment and visual feedback we compared two conditions in within-group design (counterbalanced order):

- **NonVF** - video guidance and webcam-simulated mirror *without augmentation*,
- **VF** - video guidance and webcam-simulated mirror augmented by real-time visual feedback about pose correctness.

Both conditions involved the same video recordings showing a trainer performing each stretching exercise, as well as a verbal description (audio recording) of the stretch at the beginning of each. In VF users additionally received real-time audiovisual feedback about the correctness of their actual pose.

We investigated the following hypotheses in the study:

- **H3**: Stretching is performed *more correctly* with visual feedback (VF) than with videos only (NonVF).
- **H4**: Live visual feedback about stretching performance induces *greater motivation* to stretch (and perform stretches regularly) (VF) compared to NonVF.
- **H5**: Users *prefer* stretching with our proposed visual feedback (VF) more than with video guidance only (NonVF).
- **H6**: Our proposed visual feedback for stretching is perceived as effective in terms of (a) understanding/clarity, (b) helpfulness of guidance, and (c) subjective performance.

Data for assessment of performance (i.e., correct stretching, H3) was acquired by direct error measurement in comparison to the reference pose and enriched by qualitative analysis by physiotherapists. The other hypotheses were explored through questionnaires.

6.6.2 Study Procedure

All participants completed a set of six exercises twice, once in the VF condition and once in the NonVF condition. To avoid the effects of order, conditions were counterbalanced resulting in two groups of participants. Upon arrival, all participants were informed about the procedure and data collection, signed their informed consent, and completed an initial questionnaire with personal background information. The first group started with the VF condition and the second with NonVF. After performing a set of 6 exercises with a given condition, they completed a questionnaire reflecting on the activity just performed. The first group continued with the NonVF condition and the second group with VF condition. Afterward, the participants again completed a questionnaire reflecting on the exercise set they had just completed. At the end of the experiment, they completed a questionnaire asking about differences between the exercise sets with different conditions.

6.6.3 Selected Exercises

The six exercises were selected to cover full-body stretching. During the selection of exercises, we also paid an attention to easy detectability with our single-camera body tracking approach. The following exercises were selected for our user study (Figure 6.5):

1. (APS) Arm Prayer Stretch
2. (BER) Bent Elbow Right Side
3. (CSR) Calf Stretch Right
4. (LDM) Latissimus Dorsi Muscle Stretch
5. (SHA) Standing Hamstring
6. (SHS) Standing Hamstring Stretch Right

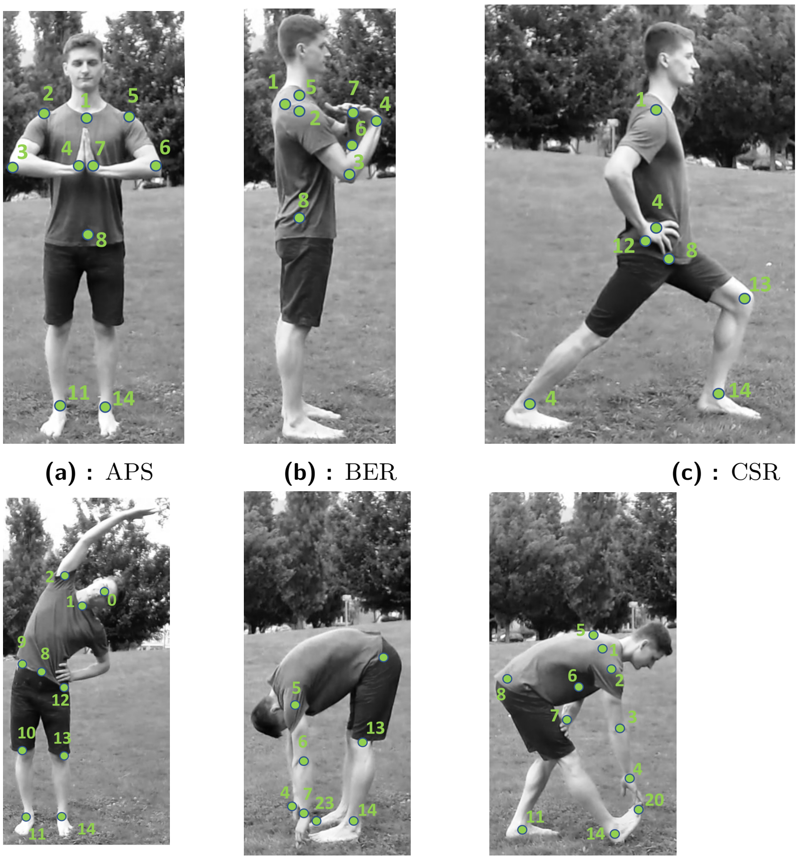
Based on pilot testing, we empirically selected a small number of suitable keypoints as features for each exercise. These are listed in Table 6.3.

Exercise	Joint angles	KP pair orientation	Relative distances
APS	(2,3,4),(5,6,7)	(1,8)	(4,7)
BER	(2,3,4),(9,10,11),(12,13,14)		
CSR	(9,10,11)	(1,8)	(11,14)
SHA	(12,13,14),(5,6,7)	(4,23)	
SHS	(2,3,4),(5,6,7)	(1,8)	(11,14)
LDM	(0,1,5),(8,12,13),(8,9,10)		

Table 6.3: For each exercise, a unique combination of features and feedback elements was experimentally selected. The numerical values correspond to the keypoints in Fig. 6.2.

6.6.4 Participants

The user study was conducted with 14 participants (9 women and 5 men). The distribution in predefined age brackets was as follows: 5 people were between 18-25 years of age, 5 people were 26-33, 2 responded with 34-41, and 2 with 42-49. More than 90% of the study participants were from an academic environment, where physically demanding work is not prevalent. All participants agreed to be video-recorded for signal processing. The questionnaire responses were provided anonymously.



(a) : APS **(b)** : BER **(c)** : CSR

(d) : LDM **(e)** : SHA **(f)** : SHS

Figure 6.5: Reference position for comparison with the trainee. A set of six exercises (performed by each participant twice; once with feedback and once without feedback.)

6.6.5 Signal Processing

From the video recordings of users' exercises, we exported the time series of all feedback element values for both executions (VF and NonVF). These feedback element values corresponded to the differences of each pose to the reference pose for a given exercise. In the next step, we calculated the mean values of these differences across the evaluated time interval. The mean differences were then aggregated across the pose features using weighted average to obtain the final pose correctness metric for each exercise. In a post-hoc step during data analysis, the weights for each individual feedback element were defined by three professional physiotherapists. In summary, the following steps were taken to quantify the correctness of the motion performance with respect to the reference poses:

1. The same time interval was used for all participants and all exercises, which was set by the countdown timer in the application. The participants practiced each exercise for exactly 30s. The 15s time interval between 0:10 to 0:25 was used for the matrics calculations to compare each exercise.

The start was at 0:10 because we already assumed the desired position was reached. The end of the interval was at 0:25 to not consider the movements at the end of the exercise.

2. An average value was determined from each time series, at a selected 15s interval for each pose feature. Thus, if the exercise was defined by 4 features, we obtained 4 average values for the exercise.

3. Physiotherapists determined the importance of each feature for correct use and thus determined the weight of the feature. For example, keeping the spine perpendicular to the ground was more important than keeping the feet together.

4. The exercise performance was determined as a weighted average of all feature distances. The performance metric was compared between two conditions for each exercise (Figure 6.6).

Pose Assessment by Experts

After the study, all videos were presented to professional physiotherapists. The physiotherapists performed two tasks:

a) Determine the weight of each feedback element (pose feature) for each exercise in terms of the correctness of the exercise execution. The individual weights can be seen in Table 6.5.

b) Make an overall assessment of whether participants performed the exercise better with or without feedback.

6.6.6 Results

In this section, we first describe the results of pose matching between the reference motion and the trainee's motion during the exercise using measured data from our system (Section 6.6.6). Second, we present our findings from the qualitative evaluation by physiotherapists (Section 6.6.6). This evaluation was done through a manual visual analysis of all recorded videos. Finally, we provide the results from our questionnaire, which investigated participants' opinions regarding motivation, feedback clarity, correction ability, helpfulness of the coaching system, and user preference (Section 6.6.6).

Pose Matching Performance Metrics

In order to evaluate how well the exercise was performed, we recorded all the movements during the exercise. For comparison, we used the stretching performance metric described in Section 6.6.5 using the selected set of pose features for each exercise.

The statistical results of differences between reference pose and trainees' poses can be seen in Figure 6.6. This figure compares errors of poses between conditions with and without visual feedback. The overall impression of the performance of all 14 participants in the study was aggregated for each

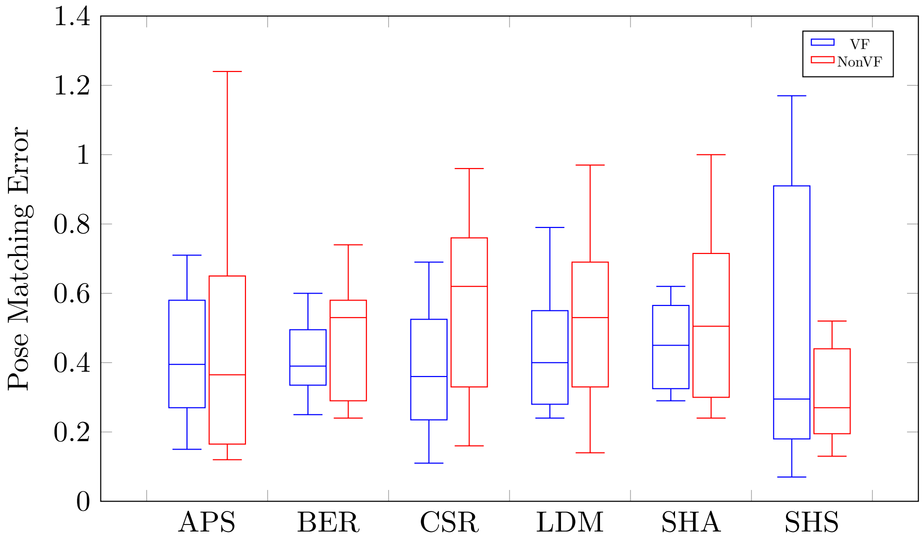


Figure 6.6: Values of metrics determining error of trainees' poses with respect to reference poses. Metrics were weighted based on qualitative assessment of professional physiotherapists. Condition with visual feedback is displayed in blue and condition without feedback is shown in red.

exercise. Despite the fact that we can see trends in the boxplots where the execution with feedback seems to show less error, we did not find a statistically significant difference in the execution of the exercises without feedback and with feedback (Table 6.4).

Exercise	APS	BER	CSR	LDM	SHA	SHS
p-value	0.766	0.644	0.088	0.286	0.460	0.682
z-score	-0.31	0.46	1.71	1.07	0.74	-0.41

Table 6.4: Statistical significance of differences between conditions with and without visual feedback for each exercise. The results were calculated using Wilcoxon signed rank test.

Qualitative Comparison by Professionals

A qualitative assessment was carried out using visual analysis. Three professional physiotherapists watched all videos taken during the study. They evaluated each exercise separately. They watched all videos where subjects performed the exercise with feedback, then watched the videos without feedback. Then, the professionals summarised the common features they found in the exercises with and without feedback. For each exercise, they described how feedback influenced the differences in performance. Based on the observations, the physiotherapists also commented on the appropriateness of the chosen feedback elements. The summary of this assessment is provided as follows:

- **(APS) Arm Prayer Stretch**

There are no significant differences seen in the user performance between VF and NonVF. In both cases, the participants performed the exercises equally well.

- **(BER) Bent Elbow Right Side**
  It is evident that in this exercise people perform the exercise better with feedback than when just watching the video. However, despite the fact that they perform it better, in some cases, they do not perform it quite as well as the trainer.

- **(CSR) Calf Stretch Right**
  For this exercise, the physiotherapists saw slightly better execution with feedback, but observation shows that the correctness of execution varies based on the physical proportions of each subject.

- **(LDM) Latissimus Dorsi Muscle Stretch**
  For this exercise, the physiotherapists did not see any noticeable differences between the performances. they attributed this mainly to poorly chosen feedback elements.

- **(SHA) Standing Hamstring**
  For this exercise, the professionals did not see any major differences in the performance of VF and NonVF. In the case of VF, some users are guided to keep both legs in a vertical position, which is desirable for exercise. Without feedback, these legs are not in a vertical position due to the lack of VF, and buttock displacement occurs.

- **(SHS) Standing Hamstring Stretch Right**
  In this exercise, the professionals observed worse performance in the VF variant. The feedback in this case forces people to get into positions they cannot hold. Here the choice of feedback elements was wrong. In this case, the elements should be chosen in a way that the front leg is extended at the knee. In the VF setting, the leg was bent and therefore the muscles that should be stretched by this exercise were not stretched.

Questionnaire Analysis

The main goal of the questionnaires in our study was to investigate differences between conditions with and without visual feedback. We were interested to study the subjective responses of participants on the understanding of instructions, helpfulness of guidance, subjective performance, motivation, and their preference between two conditions.

The results of the questionnaire analysis can be seen in Figure 6.7 and in Table 6.6. For the majority of measured factors, our visual feedback achieved better subjective ratings than the condition without visual feedback. This was not the case for the subjective performance where the condition without feedback was rated better. As we can see in Table 6.6, we did not find

Exercise	Feedback elements and weights
APS	$\measuredangle(2,3,4)$ W:1.00; $\measuredangle(5,6,7)$ W:1.00; $\measuredangle(1,8)$ W:0.80; RD(4,7) W:1.00; RD(11,14) W:0.20
BER	$\measuredangle(2,3,4)$ W:1.00; $\measuredangle(9,10,11)$ W:0.40; $\measuredangle(12,13,14)$ W:0.40; $\measuredangle(1,8)$ W:0.80; RD(4,7) W:0.20
CSR	$\measuredangle(12,13,14)$ W:1.00; $\measuredangle(1,8)$ W:1.00; RD(11,14) W:1.00; RD(4,9) W:0.20
LDM	$\measuredangle(0,1,5)$ W:0.80; $\measuredangle(8,12,13)$ W:0.80; $\measuredangle(8,9,10)$ W:0.80; RD(11,14) W:0.60; RD(7,12) W:0.60
SHA	$\measuredangle(12,13,14)$ W:1.00; $\measuredangle(5,6,7)$ W:0.60; RD(4,23) W:0.60; RD(7,20) W:0.60
SHS	$\measuredangle(5,6,7)$ W:0.60; $\measuredangle(2,3,4)$ W:0.60; $\measuredangle(1,8)$ W:0.80; RD(11,14) W:0.60; $(4,20)$ W:0.80

Table 6.5: Based on the ex-post monitoring of the recordings with the subjects, professionals in the field of physiotherapy determined a weight for each feedback element. Thus, the table always displays the name of the exercise on a row and a list of feedback elements and their respective weights in a second column.

a statistically significant difference between the conditions for any of the measured factors.

Finally, the preference between stretching with and without visual feedback was measured by a subjective, two-alternative, forced-choice preference approach. Each participant had to select a preferred condition from two experienced conditions, VF and NonVF. Out of 14 participants, 11 stated they preferred our visual feedbackA Chi-square non-parametric test suggests a significant preference for visual feedback ($\chi^2 = 4.571, p = 0.033$).

The main indicated the reason for choosing OffiStretch over just video guidance was the immediate feedback, correcting participants' poses. As noted by one of the participants: "It helped me to put myself in the correct pose and to correct my posture. I find it very helpful since finding the right angle and posture is key for every stretching exercise to bring the desired benefits." The users also offered some ideas for improving the application such as adding more gamifying elements. One of the participants suggested reducing the correction requirements for triggering audio feedback: "It was stressful because many times the voice said to adjust my position. Maybe you could rethink the tolerance of the angles and decrease the number of times it corrects you."

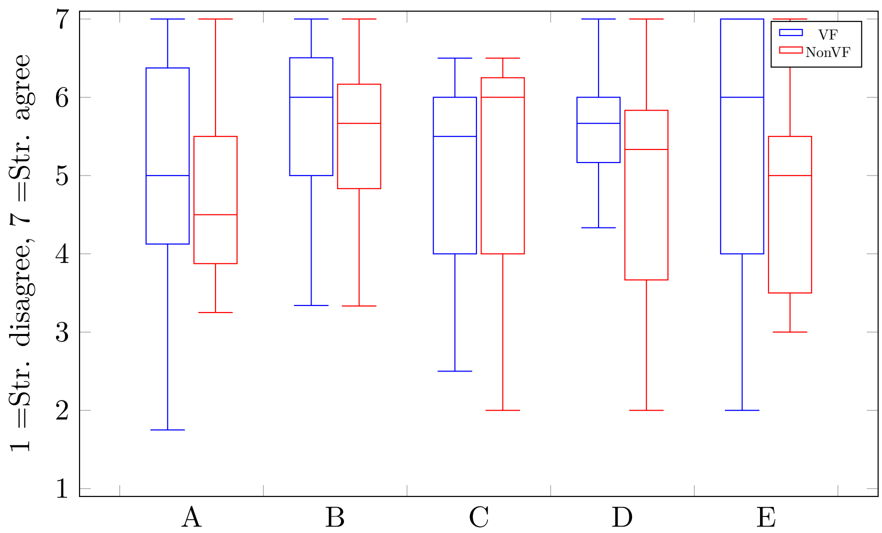


Figure 6.7: Data collected from on-site questionnaire. The condition with our visual feedback is indicated in blue and the condition without feedback is indicated in red. Higher value on y axis means more positive response for a given factor.
(A) Understanding of the instructions/visualization. (B) Helpfulness of guidance. (C) Subjective performance. (D) Motivation. (E) Preference.

Onsite Question	Z-score	p-value
Understanding of the inst./vis.	-0.50	0.61
Helpfulness of guidance	-0.43	0.67
Subjective performance	-0.15	0.88
Motivation	-1.05	0.29
Preference	-1.23	0.22

Table 6.6: Results of on-site study questionnaire - statistical significance of differences in responses after the exercises with and without feedback. Statistical significance was assessed using Wilcoxon signed-rank test.

6.7 Discussion

In our experiment, we used six exercises that cover stretching of different body parts. Based on visual analysis, the importance of visual feedback is not high for simpler exercises (APS). Some exercises take longer to understand (BER, SHA), some are challenging to perform and not all people can do them (SHS), so the feedback that informs trainees that they are not performing the exercise well can be frustrating. Some exercises can be performed worse with feedback than without feedback (SHS).

Our results indicate that feedback is in high demand for people and for some exercises we are able to design feedback that is useful to the trainees. On the other hand, we cannot use all body pose features to simply compare each

Selected Survey Questions

1	2	3	4	5	6	7
Q1: How often do you do stretching exercises? (This may be as part of a longer workout, or alone.)						
O	O	O	O	O	O	O
never	less than once a month	at least once a month	at least once weekly	multiple times per week	once per day	multiple times per day
Q2: Imagine a display that gives you real-time visual feedback about the quality of your stretching. How often can you imagine dedicating a few minutes to stretching exercises with such a coaching system during working hours (e.g. in a break)? You may assume this is approved/encouraged by management.						
O	O	O	O	O	O	O
never	less than once a month	at least once a month	at least once weekly	multiple times per week	once per day	multiple times per day
Q3: I would be willing to try a system that reminds me to stretch and instructs me on particular exercises for relieving body strain from a prolonged working pose (e.g., seated at desk, standing at workbench for many hours).						
O	O	O	O	O	O	O
strongly disagree			undecided			strongly agree

Table 6.7: Questions of our online survey, used to evaluate our two hypotheses. Each question was answered twice (once for home office condition and once for dedicated workplace condition). Answers were listed in the opposite order in the questionnaire and we inverted them for consistency of visualization within the publication.

pose to its reference counterpart. Feedback needs to be looked at in a more complex fashion and each exercise needs to be considered individually (ideally with the advice of physiotherapists). Feedback must only be a supplement. The trainee needs to know that they are being monitored and that their efforts are being recorded and measured. For example, if we detect that a person has stopped exercising at all, we can give them feedback to try to continue.

Our online survey investigated the frequency of users' stretching exercises (H1), preferred frequency of stretching with a coaching system (H2), and willingness of trying a coaching application for stretching (H2) in two conditions: (1) home office and (2) at the dedicated workplace. Our results suggest that the home office scenario is rated significantly higher than a dedicated workplace for all three factors. Therefore, both hypotheses H1 and H2 were supported by the study results.

The online survey explored the overall consciousness of participants about problems with a sedentary lifestyle, willingness to use coaching technology, as well as where and how often such technology could be used. The outcome of the survey indicates the need and user preferences towards research and development of interactive coaching applications for stretching exercises. Complementary to the online survey, our on-site study explored how our proposed system performs in comparison to a video indicating the effectiveness of our methods.

Hypothesis H3 in our on-site study, that people would perform the exercises better with feedback than without it, was not supported by the results of our quantitative error measurements. We observed that for four exercises (BER, CRS, LDM, and SHA) the error with feedback was lower than without feedback, while for two exercises the participants performed better without visual feedback (APS, SHS). These results are displayed in Figure 6.6. None of these differences are significant. Moreover, the performance comparison of VF and NonVF conditions was augmented by the comments of professional physiotherapists. These professional comments reveal additional facts about the body pose assessment, for example, dependency of correctness on physical body proportions (CSR), improper selection of feedback elements (LDM and SHS), and forcing users into improper positions (SHS). These additional comments highlight the importance of the correct selection of feedback elements, individually for each exercise. The comments of professionals on each exercise are detailed in Section 6.6.6.

Hypothesis H4, that our system induces higher motivation to perform stretching at the moment (and also regularly) than videos was only partially supported by our results. The subjectively reported motivation was higher for the condition with visual feedback than for the condition without visual feedback. However, the difference was not statistically significant.

Hypothesis H5, that our visual feedback for stretching is more preferred by the users than video guidance, was supported according to the results of our analysis of the forced-choice preference question.

Finally, hypothesis H6 focused on the differences in subjectively reported understanding of instructions, helpfulness of guidance, and performance (Figure 6.7). This hypothesis was not supported by our results because while the understanding of instructions and helpfulness of guidance were rated higher for the VF condition, the self-reported performance was higher for NonVF condition. Interestingly, this result is in contradiction with the measured quantitative pose errors for some exercises. None of the differences between conditions (in the evaluation of H6) was significant.

6.7.1 Limitations and Future Work

The main limitation of the presented methods is that a single camera and digital mirror limit the possible orientations and postures in which the users can be tracked and see themselves, therefore it can only be used for exercises that permit frontal poses [155] or side-view poses, while for others feedback in a flexible third person perspective may be of advantage [143, 146, 148].

Another limitation is the need for a really thorough selection of features for pose analysis and feedback elements. The results of our user study suggest that some of the elements that were selected and tested before the study were poorly chosen. This was only discovered after visual inspection by physiotherapy professionals. Thus, we emphasize that automated stretching coaching is an interdisciplinary problem and a mere technical solution is just a tool, but the design of similar systems cannot be done without user studies and proper insights from domain experts. In our future work, we want to design a system that facilitates mainly the involvement and evaluation of exercise selection by professionals and to conduct a larger study on more exercises and more subjects.

Our system only works with stretching exercises where the person remains in a static position. The feedback is dynamic and works with video, but the comparison is only with the static position. For the design of dynamic exercises, the system would need to be significantly modified. For some exercises, it would be enough to add more static positions (squat, push up, pull up, and similar) while for others the system would need to be completely redesigned (running, dancing, martial arts).

Another factor with a critical impact on the success of an interactive training application is the intelligibility of the visual feedback. While our indication of joint angles and distances through circles and lines was understood by all study participants, they required prior instructions. This may be improved in the future, for example by presenting the correct pose as an overlay [143, 148] on the user's mirror image.

Based on the feedback from the questionnaires, we will also work on adding more elements of gamification and competition that were suggested by the study participants. Further, the application should include possibilities to adapt exercises for users' individual motion range (e.g., to accommodate physical disabilities), as well as user customization of difficulty level and training goals [141].

6.8 Conclusion

This paper proposes novel methods for pose analysis and visual feedback for personal stretching guidance. Our methods use a single RGB camera and interactive pose estimation to detect and match the body pose of a trainee to a reference pose for stretching exercises. Finally, the detected errors are visualized for the user as an interactive overlay on a webcam-simulated mirror. This allows the user to correct their body pose and thus improve their stretching performance. We present evidence from an online survey suggesting that people prefer to perform stretching exercises more in a home office scenario than at their dedicated workplace and that there is a high willingness to use a system for interactive stretching guidance. Further, we conducted an evaluation of our OffiStretch system in a lab, investigating users' stretching performance when using our system compared to traditional video guidance. For this, we designed six stretching exercises in collaboration with professional physiotherapists. Our study reveals the importance of tailoring feedback elements to each exercise and highlights the relevance of domain knowledge when designing a system for stretching guidance. Finally, our findings suggest that users prefer live visual feedback over plain video guidance, and support the overall feasibility and potential benefit of OffiStretch.

Chapter 7

Methodology for a Single Camera-based Human Motion Capture in Physical Telerehabilitation

This chapter aims to create a guide or methodology for using computer vision techniques in home telerehabilitation. It includes a diagram that outlines the process into separate parts, important for planning this motion capture method, especially for remote rehabilitation.

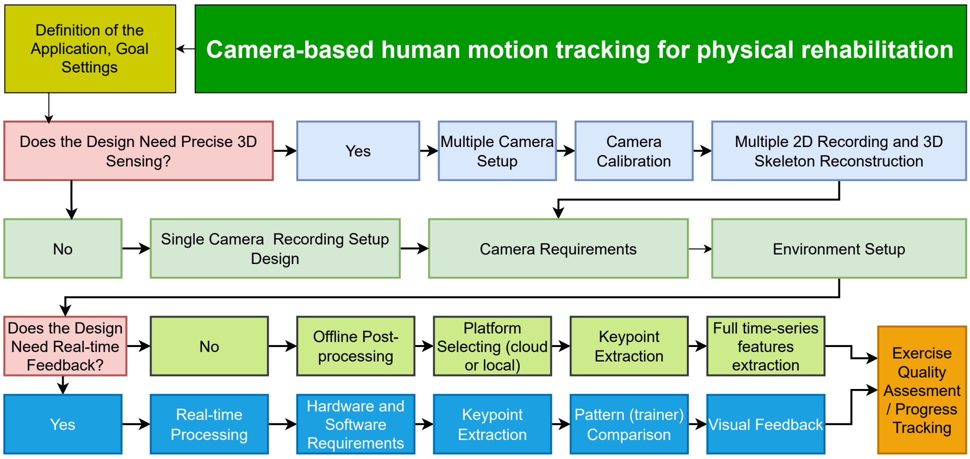


Figure 7.1: Workflow of Camera-Based Motion Capture System for rehabilitation.

7.1 Goal Settings and Purpose Design

Firstly, we need to clearly define the purpose of using camera-based motion capture. This initial goal will then guide further considerations.

- **Accuracy Requirements** What level of precision do I require from the motion capture system? Depending on the purpose of the therapy or study, the accuracy might range from very fine detail (for subtle movements) to a broader perspective (for more general motion tracking).

- **Budget Constrains** What is my budget for this application? Investing in high-end motion capture equipment might provide greater accuracy and more features. In the context of home telerehabilitation, employing such an approach may not be methodologically or economically justifiable.

- **Target group** It's essential to determine whether the aim is to assess gait under controlled laboratory conditions or to create a versatile mobile application for broader use. The choice between these approaches significantly influences the design, functionality, and application of the motion capture system.

When talking about using Motion Capture systems for very accurate results, without worrying about cost or the complicated setup, marker-based systems are the best option. But, for remote rehabilitation at home, the system needs to be easy to use and affordable to set up.

7.2 Multiple camera setup

In certain situations, it may be more appropriate to use multiple cameras, especially in cases where precise tracking in three dimensions is required, or when dealing with movements that are invisible from a single viewpoint. However, in such instances, we must be prepared for a significant increase in complexity of the entire setup, primarily due to the need for camera synchronization and calibration. While our current focus remains on a single-camera approach, it's essential to acknowledge the importance of this alternative for comprehensive understanding and completeness in our field of work.

7.3 Single Camera-based Approach Limits

For those interested in a solution primarily focusing on static and stretching exercises, a single-camera system with adherence to specific capture conditions might suffice. When designing exercises scenarios, the following potential limitations should be taken into account:

- **Lateral Movements and Exercises**: Certain movements when viewed from the side can lead to imprecise detection, or there might even be cases where the detected points get confused or swapped.

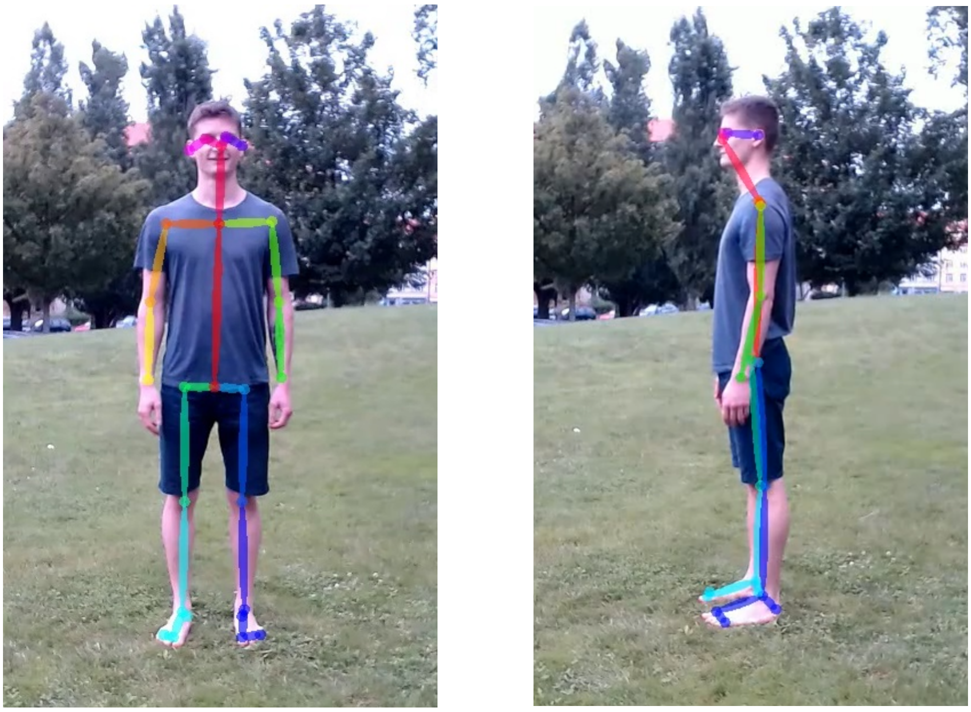


Figure 7.2: This image clearly illustrates the importance of choosing the correct camera angle to ensure that all keypoints are ideally visible.

In this case, it is necessary to instruct the patient in what position to face the camera. An illustrative example can be seen in Figure 7.2. Simplified, one could say that this method shares certain limitations with human observation. For instance, it is evident here that the right arm is obscured behind the body. As a result, neither the camera system nor the human eye can detect it, and can only rely on a statistical model to estimate its position.

- **Movements Where Body Parts Overlap Significantly**: Certain exercises, due to their inherent nature, are practically undetectable using 2D camera-based pose estimation like OpenPose. It is essential to account for these scenarios in the analysis. These exercises often involve positions where body parts overlap or obscure each other significantly, posing a challenge for accurate detection and pose estimation. Recognizing and addressing these limitations is crucial for improving the accuracy and reliability of pose estimation in such cases.

- **Exercises Involving Equipment**: In the case of the use of some aids, it is possible to estimate some parts, but in the case of covering most of the skeleton, there are problems with detection.

- **Complex Multi-plane Movements**:

  - Rolling Movements: Such as somersaults or rolls.
  - Rotational Movements: Hard to track consistently.

  Since single-camera sensing assumes only one plane, rotations in the camera axis cause problems in detection.

- **Supine or Prone Exercises**: Exercises performed in a supine (lying face up) or prone (lying face down) position present unique challenges for 2D camera-based pose estimation systems like OpenPose. In these scenarios, the camera's perspective may limit the visibility of certain body parts, leading to incomplete or inaccurate pose detection. It is important to consider these positional limitations when analyzing exercises in these orientations. More detailed information can be found in our publication [53].

- **Movements with Fast Twisting or Rotation**: Overall, exercises involving rotations across multiple axes pose significant challenges for single-camera systems. In such cases, the amount of reliable data that can be extracted diminishes considerably. The rapid twisting and rotational movements can lead to a loss of tracking accuracy, as the camera may not capture all the necessary angles and positions effectively. This limitation highlights the need for either advanced multi-camera setups or enhanced algorithmic approaches to accurately interpret and analyze movements involving complex, multi-axis rotations.

- **Small Fine-motor Movements**: Capturing small, fine-motor movements presents a distinct challenge for 2D camera-based pose estimation systems like OpenPose. These subtle movements, often involving intricate actions of the fingers, hands, or toes, may not be detected accurately due to their limited scale and the resolution constraints of the camera. The nuances of these fine-motor activities require a high level of detail and precision, which can be difficult to achieve with standard single-camera setups. To overcome these limitations, it's advisable to utilize a specialized model trained specifically for recognizing these detailed movements.

Our experience has shown that the most effective approach is to test all exercises on several individuals before implementing them in our applications. We analyze the signals generated from the time-recorded capture of keypoints. Based on consultations with a practicing physiotherapist, we then incorporate the exercise into our systems [56].

7.3.1 Feature Extraction

From the exported time series data, we can extract various individual features that are crucial for machine learning algorithms, comparative analyses, or descriptive studies. Such features might include maximum ranges during specific time intervals, movement velocities, peak values, as well as average and median values. These parameters offer insights into the dynamics and characteristics of the observed phenomena. For instance, maximum ranges and velocities can be indicative of the agility or flexibility of a subject, while average and median values might reveal typical performance or trends over time. Further, by applying statistical methods or machine learning techniques to these features, we can uncover patterns, make predictions, or even identify anomalies.

In the field of physical rehabilitation, analyzing human movement is essential for assessing progress and designing effective treatment plans. Here are some key equations commonly used for movement analysis:

$$\text{Velocity} = \frac{\Delta\text{Distance}}{\Delta\text{Time}} \tag{7.1}$$

Velocity measures the rate of change in position over time.

$$\text{Acceleration} = \frac{\Delta\text{Velocity}}{\Delta\text{Time}} \tag{7.2}$$

Acceleration quantifies how quickly velocity changes.

$$\text{Keypoint Angle} = \text{Final Angle} - \text{Initial Angle} \tag{7.3}$$

The Keypoint Angle represents the angular position of a joint or any other extracted keypoint.

$$\text{Angular Velocity} = \frac{\Delta\text{Keypoint Angle}}{\Delta\text{Time}} \tag{7.4}$$

Angular Velocity measures the rate of change in joint angle.

$$\text{Linear Displacement} = \text{Final Position} - \text{Initial Position} \tag{7.5}$$

Linear Displacement quantifies the change in position.

In practice, the choice of features depends on the specific goals and requirements of the analysis. Essential features provide a solid foundation for fundamental movement assessment, but the inclusion of standard time series features can enhance the depth of insights and support a more comprehensive understanding of human motion.

7.4 Camera Requirements

- **Frame Rate** At the beginning of the measurement we need to clearly define how fast we want to capture the motion, if we were trying to detect for example fast jumps on a trampoline, we would need a frame above 120 fps, but for normal home stretching exercises, we can make do with 30fps, which offers a standard camera [156].
- **Shutter Speed** Just as with frame rate, shutter speed depends on the movement being captured. In our case, for stretching exercises, it's appropriate to set the shutter speed to twice the frame rate, which would be 1/60 [156].
- **Light Sensitivity** The effect of image brightness was investigated in our conference paper [50], where we concluded that the effect of illumination is practically negligible on the detection of the human skeleton, see Figure 7.3 . We varied the Gamma correction parameter [157] continuously and measured its effect, the detection worked for limits that are practically invisible to the eye from both sides, both overexposed and very dark images.
- **Resolution** In analogy to the considerations surrounding illumination, the spatial resolution has a negligible impact on the accuracy of human skeletal detection algorithms. Utilizing a conventional imaging device with a resolution of 1280x720 pixels — a specification considerably below contemporary standards — and positioning a subject within the full frame at an approximate distance of 3 meters, we observed that the fidelity of detection remains largely invariant even when the operational resolution is decimated to a quarter of the original. For a detailed analysis, refer to our publication [50].

7.5 Environment Setup

- **Camera Distance Based on Subject Height**

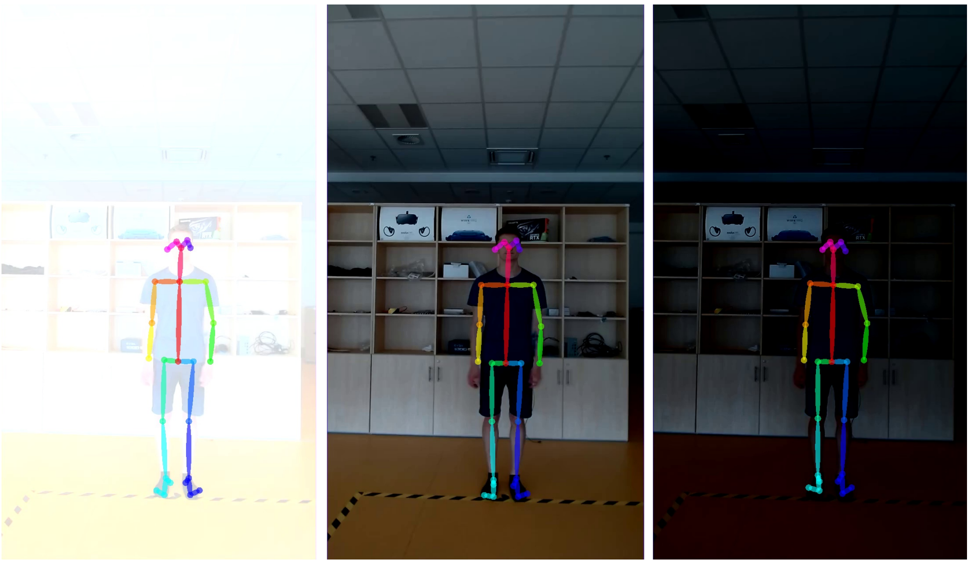


Figure 7.3: Illumination adjustment in an RGB image using variable Gamma correction parameters.

In the realm of pose estimation, determining the correct camera distance to ensure the comprehensive visibility of a subject is very important. When subjects raise their hands or adopt various postures, their effective height in the camera frame can vary significantly. Given the full extended height of the subject (from the feet to the tips of the raised hands) and specific camera parameters, we can derive a formula to calculate the optimal distance for the camera. It is crucial when planning pose estimation recording to make sure the whole body, including raised hands, is fully captured in the frame without being cut off.

To compute the required distance from the camera to the subject, the vertical Field of View (vFOV) is the most important parameter [158]. Given the focal length f and the sensor height s, the vFOV in radians is given by:

$$\mathrm{vFOV} = 2 \times \arctan\left(\frac{s}{2f}\right) \tag{7.6}$$

Given the height of the person h and the derived vFOV, the required distance d to fit the person in the frame is:

$$d = \frac{h}{2 \times \tan\left(\frac{\mathrm{vFOV}}{2}\right)} \tag{7.7}$$

Example: Camera with 1/3" CMOS Sensor and 2.8mm Lens

For a camera commonly used in my experiments (IP Security CAM) equipped with a 1/3" CMOS sensor (with an approximate height s of 4.8mm) and a 2.8mm lens (focal length f), we can plug these values into our formula.

Given:

- Sensor height $s = 4.8\text{mm}$
- Focal length $f = 2.8\text{mm}$

Using the first formula, we can determine the vFOV:

$$\text{vFOV} = 2 \times \arctan\left(\frac{4.8\text{mm}}{2 \times 2.8\text{mm}}\right) \tag{7.8}$$

Then, given the total extended height of the subject h (from the feet to the tips of the raised hands), we can determine the distance d using the second formula. For instance, if the extended height with raised hands is 2.2 meters:

$$d = \frac{2.2\text{m}}{2 \times \tan\left(\frac{\text{vFOV}}{2}\right)} \tag{7.9}$$

Compute the vFOV using the third equation, and then use that result to determine d with the fourth equation.

Given the calculated vFOV of approximately 80.36°, and using the height of the person $h = 2.2$ meters, we can determine the required distance d as:

$$d \approx \frac{2.2\text{m}}{2 \times \tan\left(\frac{80.36^\circ}{2}\right)} \tag{7.10}$$

Evaluating this expression yields:

$$d \approx 1.2\text{m} \tag{7.11}$$

Thus, to capture a subject with an extended height of 2.2 meters (including raised hands) in its entirety using the provided camera parameters, the camera should be placed approximately 1.2 meters away from the subject.

7.6 Real-Time vs. Offline Processing: A Comparative Overview

When designing the application, determining the necessity of real-time versus offline processing is critical. Real-time processing, vital for immediate feedback as explored in our work [121], demands either robust hardware capabilities or a streamlined version of computation. However, the latter may compromise the reliability of detection. For a detailed comparison of the pros and cons, please refer to the table below.

Parameter	Real-Time Processing	Offline Processing
Feedback	Immediate feedback allows for on-the-spot adjustments and interventions.	Delayed feedback, which is suitable for post-processing evaluation.
Computational Intensity	Limited to less intensive analyses due to time constraints.	Allows for comprehensive, detailed, and computationally intensive analyses or a delay in processing.
Flexibility	Requires optimized and streamlined algorithms for rapid analysis.	Provides flexibility for using a range of algorithms and tools without time constraints.
Error Correction	Limited scope for post-processing and error correction.	Allows for detailed post-processing and multiple rounds of error correction.
Cost	For the SOTA models requires specialized equipment(GPU) for real-time processing.	Can be less costly if using post-capture software solutions on standard hardware.

Table 7.1: Comparison between Real-Time and Offline Processing in Motion Capture.

Generally speaking, we can say that real-time processing is a more demanding task both in terms of device performance and algorithms, as well as overall processing complexity. On the other hand, immediate feedback is such a significant advantage for user motivation and engagement that it is necessary to pay tremendous attention to this direction.

7.7 Keypoint Extraction

Regardless of whether we use offline or real-time processing, the fundamental process of this detection method is the extraction of key points from the image. Most of the models I tested operate in a frame-based mode, meaning they process individual RGB frames independently.

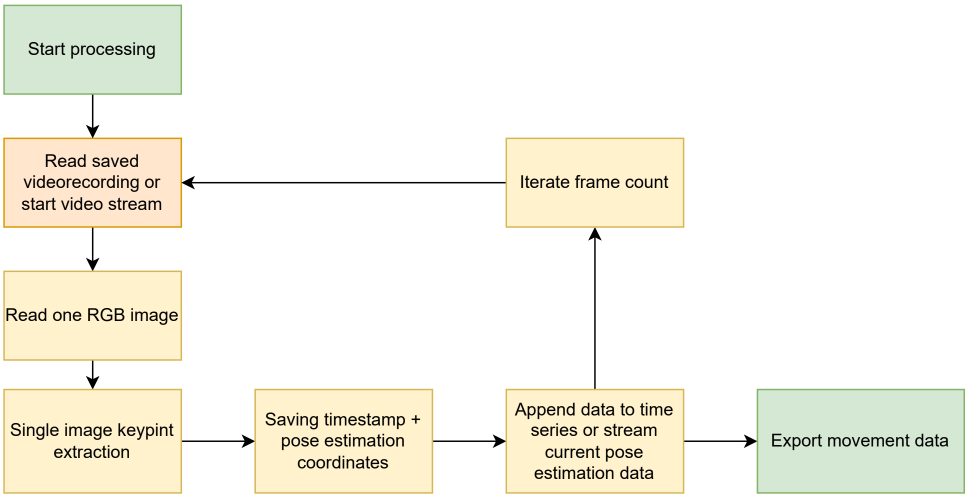
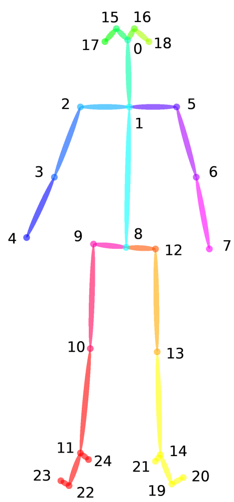


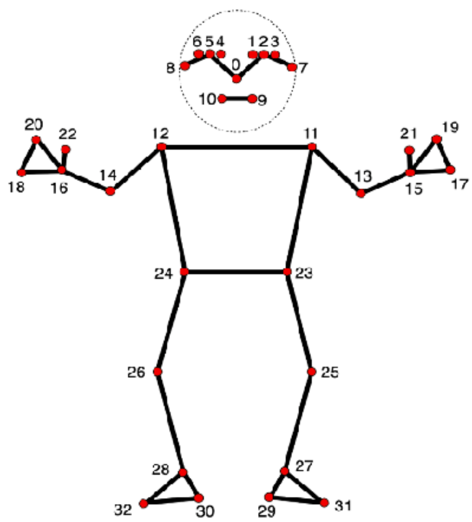
Figure 7.4: The basic concept of Keypoint extraction

7.7.1 Available Models Used for Extraction

In the state-of-the-art section, I have described how this method works in general, what models are usually used, etc. see chapter 1. Different models have distinct ways of recognizing key points. As an illustration, two of the most prevalent models at present are OpenPose (see fig. 7.5a) and MediaPipe (see fig. 7.5b).



(a) : OpenPose 25 key-points model



(b) : Mediapipe 32 keypoints model

7.7.2 Extracted Keypoints vs Biomechanical Keypoints

It's important to recognize that not all points detected from an image may accurately align with anatomical landmarks. This discrepancy often occurs when the individual's orientation relative to the camera deviates from the capturing plane. In such instances, we can resort to computing angles between these detected points. However, it's crucial to note that these computed angles might not always mirror the actual angles formed by the corresponding anatomical landmarks in real space.

Despite these potential inaccuracies, this approach remains suitable for machine learning (ML) purposes. Machine learning algorithms are adept at identifying patterns and trends from large datasets, even if the data contains some level of noise or inaccuracy. In the context of anatomical analysis, ML can learn to correlate the computed angles with specific movements or postures, despite the angles not precisely representing real-world anatomical positions.

7.8 Selecting HW for Processing

Camera-based human pose estimation systems can vary in their computational requirements based on several factors:

- **Model Complexity**:
  - Deep learning models, especially convolutional neural networks (CNNs), can be computationally intensive.
  - Models with more layers and parameters often require more computational power but can achieve higher accuracy.

  The OpenPose system I work with most in my research has the following requirements, at least 1.5GB of GPU memory and CUDA[1] 10 or later.

- **Resolution of the Input Image**:
  - Higher-resolution images provide more details but increase the computational load.
  - Systems may downsample images to speed up processing at the cost of potential accuracy losses.

  For neural networks, especially those processing images (like Convolutional Neural Networks, CNNs), the resolution isn't necessarily a measure of "importance", but it does impact the granularity of features the network can identify.

- **Real-time vs. Offline Processing**:

[1]https://docs.nvidia.com/cuda

- Real-time pose estimation requires rapid processing to provide immediate feedback, demanding more computational resources.
- Offline processing allows for more extensive computations as immediate results are not necessary.

In my experiments, I achieved 16fps on the Nvidia RTX2060 GPU in real-time. However, when considering portable devices, it's necessary to use lower performance due to the limited computational power on mobile devices, battery consumption, etc. This issue can be partially addressed by approximating between frames, where we don't calculate all of them, but only some. We published this approach in a joint paper [109] with our partners from Israeli BGU.

- **Computing Power and Cost Estimation for Video Processing**:

  Systems designed for multiple users require more computational resources than those for individual users. In our offline system, we implemented a queue system that processes users based on the order their videos were received for processing. Therefore, users receive feedback on their exercise with a delay of several minutes. It always depends on the design of the application, and in this case, it wasn't an issue.

  This fact should be taken into consideration, and the computational power can be calculated as follows:

  In the realm of video processing, the computational requirement is often dictated by the number of frames that need to be processed. Given that a standard video is typically shot at 30 frames per second (fps), a 5-minute video contains:

$$\text{Frames} = \text{Video Length (seconds)} \times \text{FPS} = 5 \times 60 \times 30 = 9,000 \text{ frames.} \tag{7.12}$$

  For a single Nvidia RTX2070 GPU, which can process videos at a rate of 16 fps, the processing time required for this video is approximately 9.375 minutes. Consequently, when considering 1000 users uploading such videos daily, the cumulative processing time surges to 156.25 hours. In practical terms, to complete the processing within a day, approximately 7 Nvidia RTX2070 GPUs are essential. The investment cost associated with this setup would be the product of the number of GPUs and the cost of a single GPU. Moreover, the recurring costs encompass the electricity and infrastructure charges, among other overheads. It's necessary to note that this calculation presumes linear scaling, and real-world scenarios might necessitate additional considerations, such as GPU load or software efficiency.

Chapter 8

Disscussion

8.1 Addressing the Research Questions

1. What are the practical differences, advantages, and disadvantages between using virtual reality and a camera-based system for motion capture?
2. How well does the camera-based motion capture detection work?
3. What are the limits of capturing motion with this camera approach?
4. For which exercises or movements is this approach suitable?
5. How demanding is the camera-based system on computational performance?
6. What could be the clinical applications of this camera-based motion capture?

8.1.1 What are the practical differences, advantages, and disadvantages between using virtual reality and a camera-based system for motion capture?

To understand how people approach virtual reality, I first conducted a user study to investigate how well individuals can learn a new motor skill through virtual reality. In my case, this skill was juggling. I compared this with learning to juggle using standard balls. The results were published in the conference paper "Advantages of Immersive Virtual Reality in Juggling Learning [51]".

Next, I developed an application that allowed users to see their entire body in virtual reality and measure the angles between different parts of their body. This work was published in the article "Affordable Personalized, Immersive VR Motor Rehabilitation System with Full Body Tracking [49]". Throughout these experiments, I gathered feedback from participants on how they interacted with the system.

8.1.2 How well does the camera-based motion capture detection work?

The fundamental question of how well the imaging works is challenging to determine for practical imaging using a camera system. This difficulty arises primarily from the inherent nature of how the system works. The system operates based on the principle of classifying specific patterns within an image. The efficacy of such a system is only as good as the data it's trained on. Most models are trained on individuals who are directly facing the camera. There is limited data on scenarios where an individual is partially obscured by other objects or is in unnatural positions, etc. Precisely for this reason, to investigate this effect, we conducted tests on a large dataset. For individual categories and camera viewpoints on the subjects, we determined detection quality. This factor is thoroughly analyzed in my publication, "Single Camera-Based Remote Physical Therapy". Please see chapter 4. The work outlines both the advantages of virtual reality and its drawbacks, particularly the need for specialized hardware and complex setup requirements.

8.1.3 What are the limits of capturing motion with this camera approach?

The limitations of this approach, as well as the recommended methodology for such capture, are detailed in the chapter Methodology 7. In general, we can say that we start from the basic principle that what a user can see with the naked eye can also be seen through a camera. Therefore, it is necessary to plan the design of the recording so that the most important parts are clearly visible in the image, without being obscured by objects or overlapping with other parts of the body.

8.1.4 For which exercises or movements is this approach suitable?

I attempt to answer this question in my publication [56]. Based on analyzing a large database of videos, we categorized exercises into specific categories based on camera angles. This is crucial for the quality of detection. The publication provides detailed statistical evaluations on which views are appropriate. Further results from our experiments are published in the articles [53, 54], where we address one of the most vulnerable positions, the lying-down posture. In this context, we measure 14 different angles in a lying position and perform comparative measurements against the traditional method - the goniometer.

In summary, following an extensive review of my publications that utilize the single-camera technique for skeletal detection in rehabilitation [53, 50, 56], I have incorporated these procedures into Chapter Methodology in section 7.3.

8.1.5 How demanding is the camera-based system on computational performance?

In conclusion, while some camera-based pose estimation systems can be quite demanding in terms of computational performance, careful design, model choices, and hardware considerations can optimize these requirements for various applications. In general, we can say that the more precise the detection we require and the faster we want the data to be processed, the more demanding and costly the processing becomes. For perfect real-time detection, we need top-of-the-line graphics cards, while for post-processing with fewer details, a standard processor can suffice. More details can be found in section 7.8.

8.1.6 What could be the clinical applications of this camera-based motion capture?

To answer this question, I collaborated with the Faculty of Physical Education and Sport (FTVS) at Charles University, focusing on three practical applications.

1. **Angle Measurement Accuracy:** We experimentally verified that when a person faces the camera in the correct plane, the measurements are practically consistent and highly accurate. Our focus was on measuring angles in a lying position. We documented the inaccuracies of this approach at various angles and compared the measured angles with those obtained by several physiotherapists using a goniometer. More details on this research can be found in Section 3.6 and our publication Upper limb range of motion evaluation by a camera-based system [53], as well as in the thesis [54].

2. **Fatigue Detection:** Another experiment involved measuring arm fatigue during weight lifting, while also recording EMG muscle activity. This research is detailed in Section 3.7 of this thesis and in Tereza Skalová's thesis [55].

3. **Functional Tests Evaluation:** In my main publication [88], I discuss the potential use of a camera system for automatically evaluating the correctness of performing functional tests in physiotherapy. This topic is covered in detail in Chapter 5 of this work. One of the primary issues addressed by this approach is the potential for objective evaluation, reducing the influence of the human factor, where different experts may provide varying results when evaluating the tests.

In essence, camera-based motion capture systems provide clinicians with a powerful tool to understand, diagnose, and treat a variety of conditions, enhancing patient outcomes and optimizing therapeutic interventions.

8.2 Achievement of Work Objectives

The upcoming sections will cover each goal in detail. For each one, we'll talk about how well it was met and point out the exact parts of this thesis where you can find the related analysis, experiments, or discussions.

Below, each research question is listed once again, followed by a detailed discussion that aligns with the objectives. This structure ensures a clear and direct correlation between the questions posed at the outset and the subsequent findings and analyses addressed in the respective sections or chapters of the thesis.

1. Evaluate the advantages and challenges of using virtual reality for motion capture compared to camera-based systems. This investigation will explore how virtual reality can enhance motion capture with its

immersive and interactive capabilities, offering potentially more precise and dynamic data collection in controlled environments. Conversely, it will also examine the limitations of virtual reality systems, such as potential technical complexities and user discomfort

2. Describe the functional concept of telerehabilitation using a camera. Due to the ubiquity of cameras, this allows for capturing movement virtually anywhere using any device. This approach provides a versatile platform for rehabilitation and can be adapted to various environments, making it a flexible solution for diverse needs.

3. Verify detection functionality using a large video database. This will determine the optimal perspectives for the camera system, facilitating the establishment of the correct methodology for movement recording.

4. Assess the feasibility of building machine learning models with the gathered data. Expert evaluations by practicing physiotherapists will generate a dataset containing both comprehensive movement records and assessments of these movements. Such a dataset can then be used for more sophisticated data modeling.

5. An overarching aim of the research is to bridge experts from all involved fields. This includes not only a perspective from cybernetics and biomedical engineering but also insights from experts in physiotherapy. This interdisciplinary approach will bolster the practical applicability of the research.

6. Develop an automated evaluation software tool, which will assist physiotherapists in facilitating and streamlining the diagnosis of exercise execution.

7. Create a functional application based on this system and define feedback elements for interactive exercises. To validate the entire concept in practice, it's essential to develop a working prototype for real-time exercise. Testing this software will not only measure the detection's success but, more importantly, evaluate the user experience.

8.2.1 Challanges of Virtual Reality in Rehabilitation

Through my own research and the application I developed [51, 49], I confirmed that virtual reality has the remarkable ability to fully immerse users and enhance their motivation [159]. This immersion opens up endless possibilities for creating rehabilitation scenarios where individuals can interact freely with their environment. The accuracy of motion tracking is also sufficiently high for conducting movement analysis and recording exercise sessions [29].

However, there are two main issues with using VR headsets. The first is the ongoing need for specialized hardware, which comes with the challenge of setting up the entire system [160]. This often involves complex installation, calibration, and significant space requirements. The second issue is the display itself. Users have to wear a headset, which is often heavy and prevents them from seeing their surroundings. This can lead to motion sickness and a sense of insecurity in their movements within the virtual world [161].

Despite the fascinating possibilities virtual reality offers for designing therapeutic scenarios and developing applications, its practical and widespread use is significantly limited by the complexities of setup and hardware requirements. Consequently, after conducting initial experiments, I shifted my research focus from VR to camera-based systems, which are far more practical and applicable in real-world settings.

8.2.2 Concept of Telerehabilitation Using a Camera-based Approach

The concept of telerehabilitation utilizing camera technology is introduced in Section 3.3, with a comprehensive exploration provided in my publication "Automatic telerehabilitation system in a home environment using computer vision"[52]. This concept involves techniques for data acquisition, assessment, server-based data management, user feedback mechanisms, and data analytics, particularly focusing on data trends.

8.2.3 Verify Detection Functionality Using a Large Video Database

I have invested considerable attention in this section and have authored a dedicated chapter for this work (Chapter 4). This chapter was instrumental in determining which exercises are suitable for this approach and provided valuable insights into identifying videos that are easily detectable and those that are not. For a more comprehensive understanding, please refer to the freely accessible publication [56].

8.2.4 Feasibility of Building Machine Learning Models

During my research, I repeatedly encountered the issue of certain parameters being poorly defined. This reflects the fact that projecting 3D motion into a 2D representation results in measured angles that do not correspond to a human's actual biomechanical angles. Even though these angles appear as the main features, we treat them merely as features, and not as actual biomechanical angles. Statistical parameters derived from these angles serve as features that define the execution of an exercise and can be used for model training. To classify the quality of execution, we consulted experts who visually assessed performance across various categories. Their expertise is invaluable as we aim to digitalize their knowledge for more consistent and automated assessments in the future. This approach proved useful and points in the right direction for future research. However, there is a need to define parameters more precisely and to consult with a larger number of physiotherapists. More detailed information can be found in my second impact publication [88] which is a separate chapter 5 of this thesis. However, this remains the primary research direction and a concerted effort to build the most accurate model for highly precise automatic detection. Yet, achieving this goal will be challenging without acquiring a larger dataset and obtaining more expert evaluations from the field.

8.2.5 Multidisciplinary Research Connection

My goal was to apply modern machine learning algorithms and leverage the expertise of physiotherapy professionals. Physiotherapists possess invaluable knowledge of human biomechanics, movement patterns, and rehabilitation nuances, while I, as a computer scientist, focus on algorithms, data processing, and system integration.

It became clear that without a collaborative effort between these disciplines, there was a risk of creating systems that either misinterpreted the nuances of human movement or missed out on the analytical strengths that computer vision offers. During my research, I established a partnership with the Faculty of Physical Education and Sport at Charles University (UK, FTVS) and successfully co-authored several publications together.

8.2.6 Real-time Application

During my experiments and research, I developed an application named Offistretch. This app captures the user's exercise posture in real-time and provides immediate feedback through a graphical interface on the screen. I conducted a user study, the findings of which are detailed in chapter 6. The results of this study indicate that the usability of these systems is good, but only for a limited number of exercises. Additionally, the feedback elements must be carefully and sensitively selected in collaboration with experts in physiotherapy.

8.3 Future Work

Throughout my intensive research, I have had the privilege of exploring a variety of technologies, developing proof-of-concept applications, and conducting user studies. This journey has led to fruitful collaborations with prestigious universities such as Ben Gurion University of the Negev, the Technical University in Vienna, and Charles University in Prague. These partnerships have not only enriched my research but also laid a solid foundation for future joint endeavors.

My work explores the promising area of camera-based systems for telerehabilitation. Building on the findings of my studies, I see several exciting ways for future research:

1. **Advancing Camera-Based Systems in Clinical Applications**: The insights gained from our research highlight the potential of camera-based approaches in transforming clinical practices. While we have successfully

demonstrated their application in automating the evaluation of specific functional tests such as the Step-down Test (SDT) and the Single-Leg-Stance Test (SLST), our future efforts will broaden this scope. We aim to refine these systems to enhance their accuracy and reliability across various clinical assessments and therapeutic exercises. This will involve not only technical improvements in motion detection and analysis but also a deeper integration of expert clinical knowledge. Our goal is to develop versatile tools that can be adapted to a wide range of clinical settings, thereby improving patient assessment, treatment planning, and monitoring outcomes in diverse healthcare environments.

2. **Expanding the Scope of Home Rehabilitation Tools**: The usability of camera-based systems in home environments, as demonstrated by our research with OpenPose, opens up new possibilities. Future efforts will be directed towards addressing the challenges identified, such as the detection issues in certain body positions. We plan to extend our research to include a wider range of exercises and to refine the system for better detection and analysis.

3. **Interdisciplinary Collaboration and System Development**: The next phase of our research will involve strengthening our expert base and preparing our software for broader subjective evaluation. This includes enabling remote expert evaluation from video recordings, which will lead to more robust exercise execution assessments. We also plan to explore the integration of machine learning algorithms to enhance the accuracy and efficiency of these evaluations.

4. **Practical Implementation and Commercial Partnerships**: In partnership with the commercial sector, I have submitted a grant proposal outlining future research trajectories and practical implementations, with a focus on physiotherapy and ergonomics. This proposal is not just the culmination of my current work but also a roadmap for future innovations. We aim to create systems that aid in the diagnosis and treatment of movement disorders and serve as supportive tools for physiotherapy and other disciplines involving movement analysis.

5. **Addressing Limitations and Expanding Capabilities**: Acknowledging the limitations of our current approach, particularly in precise motion identification from various views, future work will also focus on enhancing the system's ability to accurately recognize and evaluate exercises from different camera angles and positions. This will involve a detailed analysis of keypoint detection confidence and the development of algorithms that can adapt to these variances.

In conclusion, the path ahead is rich with opportunities for innovation and advancement. By combining modern computer vision techniques, deep learning algorithms, and the expertise of physiotherapists, we aim to revolutionize the way rehabilitation exercises are evaluated and conducted. Our goal is to create cost-effective, efficient, and accessible tools that complement the work of physiotherapists and enhance patient care in various settings.

Chapter 9

Contribution to the field

This dissertation has explored the potential of camera-based motion capture technology in the field of rehabilitation, aiming to enhance the understanding, diagnosis, and treatment of various conditions through innovative telerehabilitation applications. The research presented herein has made significant contributions to the field, both in terms of theoretical insights and practical applications, which are summarized below.

9.1 Theoretical Contributions

1. **Enhanced Understanding of Camera-Based Motion Capture**: Through extensive experimentation and analysis, this work has contributed to a deeper understanding of how camera-based motion capture can be effectively utilized in rehabilitation. It has clarified the operational limits of this technology, the computational demands of different system designs, and the specific exercises and movements that can be accurately captured and analyzed.
2. **Methodological Advancements**: The development and validation of a comprehensive methodology for capturing and analyzing motion using camera-based systems represent a theoretical contribution. This methodology not only improves the precision of motion capture in rehabilitation settings but also offers a framework for future research to build upon.
3. **Interdisciplinary Insights**: By bridging the fields of computer science, physiotherapy, and rehabilitation science, this dissertation has highlighted the importance of interdisciplinary collaboration in advancing telerehabilitation technologies. The insights gained from this collaborative approach have enriched the understanding of both the technical and clinical aspects of camera-based motion capture.

9.2 Practical Contributions

1. **Proof-of-Concept Applications**: The creation and testing of proof-of-concept applications, such as Offistretch, demonstrate the practical viability of camera-based systems in supporting rehabilitation exercises and providing real-time feedback as an augmented mirror. These applications serve as foundational tools that can be further developed and adapted for wider clinical and home use. A prototype application for motion capture in virtual reality was used to gather insights into the usability and limitations of VR in practice. It demonstrated that the use of cameras is a more suitable direction for research.
2. **Clinical Applications and Evaluations**: Collaboration with the Faculty of Physical Education and Sport (FTVS) at Charles University has led to the identification of specific clinical applications of camera-based motion capture, such as angle measurement accuracy, fatigue

detection, and the evaluation of functional tests. These applications have shown promise in enhancing patient outcomes by offering more precise and objective assessments.

Chapter 10

Conclusion

This dissertation aimed to explore the potential of using a single camera as a sensing device for physical rehabilitation, emphasizing its applicability in telerehabilitation due to the simplicity of capturing exercises with any device with an integrated camera. Initially, the research explored the limits of using virtual reality in telerehabilitation as one of the options. However, it was found that a camera-based system offered greater practicality and accessibility for both therapists and patients. Consequently, the focus shifted to maximizing the effectiveness of single-camera solutions. Central to this investigation were the application of computer vision algorithms and the processing of digital video recordings, addressing questions about the efficacy, limitations, and suitability of camera-based motion capture in physical therapy.

The work is based on the following three studies: the verification of camera-based motion capture using a large video dataset, the evaluation of functional test performance through a camera-based and machine-learning approach, and the development of OffiStretch, a real-time application for daily stretching exercises. These studies collectively addressed key research questions, ranging from the technical performance of camera-based detection to its practical clinical applications.

The first study, "Single Camera-Based Remote Physical Therapy: Verification on a Large Video Dataset," extensively examined the OpenPose algorithm's capability to detect anatomical landmarks under varied conditions, providing foundational insights into the operational parameters and limitations of such systems. This study highlighted how the participant's location and the camera's angle affect detection quality, providing a detailed view of the system's effectiveness.

Furthering the application of camera-based systems, "Evaluation of Functional Tests Performance Using a Camera-based and Machine Learning Approach" ventured into the clinical realm, demonstrating how the integration of machine learning with camera-based systems could perform functional test assessments. This study showed how these systems could be useful in medical settings, providing a method that combines expert opinions with precise algorithms for better medical assessments.

Lastly, the development of OffiStretch showcased the practical application of camera-based systems in promoting physical activity and correcting exercise postures with real-time feedback as an augmented mirror. This initiative highlighted the potential of digital tools to motivate and guide users in their exercise routines, emphasizing the importance of accurate, accessible, and user-friendly technological solutions in addressing the challenges of physical inactivity and sedentary lifestyles.

Collectively, these studies underscore the dissertation's contribution to advancing the field of physical rehabilitation through technological innovation.

In conclusion, this dissertation successfully demonstrated the feasibility, challenges, and clinical relevance of using a single camera for motion capture in rehabilitation contexts. The knowledge and methods developed here provide a valuable roadmap for future system development, ensuring goals are met and setting the stage for ongoing progress in the field.

Bibliography

[1] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh, "Openpose: Realtime multi-person 2d pose estimation using part affinity fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[2] B. Putsa, W. Jalayondeja, K. Mekhora, P. Bhuanantanondh, and C. Jalayondeja, "Factors associated with reduced risk of musculoskeletal disorders among office workers: a cross-sectional study 2017 to 2020," *BMC Public Health*, vol. 22, Aug. 2022.

[3] C. Tersa-Miralles, C. Bravo, F. Bellon, R. Pastells-Peiró, E. R. Arnaldo, and F. Rubí-Carnacea, "Effectiveness of workplace exercise interventions in the treatment of musculoskeletal disorders in office workers: a systematic review," *BMJ Open*, vol. 12, p. e054288, Jan. 2022.

[4] L. J. McGowan, R. Powell, and D. P. French, "Older adults' construal of sedentary behaviour: Implications for reducing sedentary behaviour in older adult populations," *Journal of Health Psychology*, vol. 26, pp. 2186–2199, Mar. 2020.

[5] B. Chen, N. Hu, and J.-H. Tan, "Efficacy of home-based exercise programme on physical function after hip fracture: a systematic review and meta-analysis of randomised controlled trials," *International Wound Journal*, vol. 17, pp. 45–54, Nov. 2019.

[6] A. Peretti, F. Amenta, S. K. Tayebati, G. Nittari, and S. S. Mahdi, "Telerehabilitation: Review of the state-of-the-art and areas of application," *JMIR Rehabilitation and Assistive Technologies*, vol. 4, p. e7, July 2017.

[7] J. P. Bettger and L. J. Resnik, "Telerehabilitation in the age of COVID-19: An opportunity for learning health system research," *Physical Therapy*, vol. 100, pp. 1913–1916, Aug. 2020.

[8] F. Saaei and S. G. Klappa, "Rethinking telerehabilitation: Attitudes of physical therapists and patients," *Journal of Patient Experience*, vol. 8, p. 237437352110343, Jan. 2021.

[9] P. Seron, M.-J. Oliveros, R. Gutierrez-Arias, R. Fuentes-Aspe, R. C. Torres-Castro, C. Merino-Osorio, P. Nahuelhual, J. Inostroza, Y. Jalil, R. Solano, G. N. Marzuca-Nassr, R. Aguilera-Eguía, P. Lavados-Romo, F. J. Soto-Rodríguez, C. Sabelle, G. Villarroel-Silva, P. Gomolán, S. Huaiquilaf, and P. Sanchez, "Effectiveness of telerehabilitation in physical therapy: A rapid overview," *Physical Therapy*, vol. 101, Feb. 2021.

[10] A. J. Buabbas, S. E. Albahrouh, H. N. Alrowayeh, and H. Alshawaf, "Telerehabilitation during the COVID-19 pandemic: Patients and physical therapists' experiences," *Medical Principles and Practice*, vol. 31, no. 2, pp. 156–164, 2022.

[11] C. Wei and J. Finkelstein, "Comparison of alexa voice and audio video interfaces for home-based physical telerehabilitation," in *AMIA Annual Symposium Proceedings*, vol. 2022, p. 496, American Medical Informatics Association, 2022.

[12] S. Aloyuni, R. Alharbi, F. Kashoo, M. Alqahtani, A. Alanazi, M. Alzhrani, and M. Ahmad, "Knowledge, attitude, and barriers to telerehabilitation-based physical therapy practice in saudi arabia," *Healthcare*, vol. 8, p. 460, Nov. 2020.

[13] W. S. Bjorbækmo and A. M. Mengshoel, ""a touch of physiotherapy" — the significance and meaning of touch in the practice of physiotherapy," *Physiotherapy Theory and Practice*, vol. 32, pp. 10–19, Jan. 2016.

[14] V. Hein and A. Koka, "Perceived feedback and motivation in physical education and physical activity," 2007.

[15] R. Essery, A. W. A. Geraghty, S. Kirby, and L. Yardley, "Predictors of adherence to home-based physical therapies: a systematic review," *Disability and Rehabilitation*, vol. 39, pp. 519–534, Apr. 2016.

[16] G. Nagymáté, T. Tuchband, and R. M. Kiss, "A novel validation and calibration method for motion capture systems based on micro-triangulation," *Journal of Biomechanics*, vol. 74, pp. 16–22, June 2018.

[17] Z. Tang, R. Grompone von Gioi, P. Monasse, and J.-M. Morel, "A precision analysis of camera distortion models," *IEEE Transactions on Image Processing*, vol. 26, no. 6, pp. 2694–2704, 2017.

[18] P. Eichelberger, M. Ferraro, U. Minder, T. Denton, A. Blasimann, F. Krause, and H. Baur, "Analysis of accuracy in optical motion capture – a protocol for laboratory setup evaluation," *Journal of Biomechanics*, vol. 49, pp. 2085–2088, July 2016.

[19] M. Topley and J. G. Richards, "A comparison of currently available optoelectronic motion capture systems," *Journal of Biomechanics*, vol. 106, p. 109820, June 2020.

[20] Z. Fikri, R. Hambali, and K. Rahayu, "Validity and reliability of vicon™ motion capture camera over the traditional anthropometric method," *Malaysian Journal of Public Health Medicine*, vol. 18, pp. 142–151, 06 2018.

[21] A. Drapeaux and K. Carlson, "A comparison of inertial motion capture systems: DorsaVi and xsens," *International Journal of Kinesiology and Sports Science*, vol. 8, p. 24, July 2020.

[22] A. Benjaminse, R. Bolt, A. Gokeler, and B. Otten, "A validity study comparing xsens with vicon," *ISBS Proceedings Archive*, vol. 38, no. 1, p. 752, 2020.

[23] X. Yun, E. R. Bachmann, and R. B. McGhee, "A simplified quaternion-based algorithm for orientation estimation from earth gravity and magnetic field measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 3, pp. 638–650, 2008.

[24] J. Zhao, "A review of wearable IMU (inertial-measurement-unit)-based pose estimation and drift reduction technologies," *Journal of Physics: Conference Series*, vol. 1087, p. 042003, Sept. 2018.

[25] J. Xia, L. He, Z. Lu, L. Liu, J. Song, S. Chen, Q. Wang, F. A. Hammad, and Y. Tian, "Stretchable and sensitive strain sensors based on CB/MWCNTs–TPU for human motion capture and health monitoring," *ACS Applied Nano Materials*, vol. 6, pp. 9736–9745, May 2023.

[26] D. Goto, Y. Sakaue, T. Kobayashi, K. Kawamura, S. Okada, and N. Shiozawa, "Bending angle sensor based on double-layer capacitance suitable for human joint," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 4, pp. 129–140, 2023.

[27] V. Omowonuola, Z. Ridgeway, B. Vandeput, Y. Yamashiro, and S. Kher, *Virtual Reality and Motion Capture Training*, p. 792–803. Springer International Publishing, Oct. 2021.

[28] P. Caserman, P. Achenbach, and S. Göbel, "Analysis of inverse kinematics solutions for full-body reconstruction in virtual reality," in *2019 IEEE 7th International Conference on Serious Games and Applications for Health (SeGAH)*, pp. 1–8, 2019.

[29] J. P. Vox, A. Weber, K. I. Wolf, K. Izdebski, T. Schüler, P. König, F. Wallhoff, and D. Friemert, "An evaluation of motion trackers with virtual reality sensor technology in comparison to a marker-based motion capture system based on joint angles for ergonomic risk assessment," *Sensors*, vol. 21, p. 3145, May 2021.

[30] K. F. Sim and K. Sundaraj, "Human motion tracking of athlete using optical flow artificial markers," in *2010 International Conference on Intelligent and Advanced Systems*, pp. 1–4, 2010.

[31] V. L. Guen, C. Rambour, and N. Thome, "Complementing brightness constancy with deep networks for optical flow prediction," in *Lecture Notes in Computer Science*, pp. 121–138, Springer Nature Switzerland, 2022.

[32] B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," in *IJCAI'81: 7th international joint conference on Artificial intelligence*, vol. 2, (Vancouver, Canada), pp. 674–679, Aug. 1981.

[33] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 1, pp. 185–203, 1981.

[34] H. M. Hondori and M. Khademi, "A review on technical and clinical impact of microsoft kinect on physical therapy and rehabilitation," *Journal of Medical Engineering*, vol. 2014, pp. 1–16, Dec. 2014.

[35] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," 2017.

[36] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2d human pose estimation: New benchmark and state of the art analysis," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[37] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, "Microsoft coco: Common objects in context," 2014.

[38] Google, "Mediapipe: Pose landmarker," 2019. Accessed: 24.8.2023.

[39] Y. Luo, Z. Ou, T. Wan, and J.-M. Guo, "FastNet: Fast high-resolution network for human pose estimation," *Image and Vision Computing*, vol. 119, p. 104390, Mar. 2022.

[40] H.-S. Fang, S. Xie, Y.-W. Tai, and C. Lu, "RMPE: Regional multi-person pose estimation," IEEE, Oct. 2017.

[41] D. Maji, S. Nagori, M. Mathew, and D. Poddar, "Yolo-pose: Enhancing yolo for multi person pose estimation using object keypoint similarity loss," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 2637–2646, June 2022.

[42] C. Zheng, W. Wu, C. Chen, T. Yang, S. Zhu, J. Shen, N. Kehtarnavaz, and M. Shah, "Deep learning-based human pose estimation: A survey," *ACM Computing Surveys*, vol. 56, pp. 1–37, Aug. 2023.

[43] M. Andriluka, U. Iqbal, E. Insafutdinov, L. Pishchulin, A. Milan, J. Gall, and B. Schiele, "PoseTrack: A benchmark for human pose estimation and tracking," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, June 2018.

[44] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *CVPR*, 2017.

[45] S. Bhattacharyya, "A brief survey of color image preprocessing and segmentation techniques," *Journal of Pattern Recognition Research*, vol. 1, no. 1, pp. 120–129, 2011.

[46] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, pp. 1–6, 2017.

[47] S. S. Basha, S. R. Dubey, V. Pulabaigari, and S. Mukherjee, "Impact of fully connected layers on performance of convolutional neural networks for image classification," *Neurocomputing*, vol. 378, pp. 112–119, Feb. 2020.

[48] E. Tjoa, H. J. Khok, T. Chouhan, and G. Cuntai, "Improving deep neural network classification confidence using heatmap-based explainable ai," 2022.

[49] J. Adolf, J. Dolezal, and L. Lhotska, "Affordable personalized, immersive vr motor rehabilitation system with full body tracking," *Studies in health technology and informatics*, vol. 261, pp. 75–81, 01 2019.

[50] J. Adolf, J. Dolezal, M. Macas, and L. Lhotska, "Remote physical therapy: Requirements for a single rgb camera motion sensing," in *2021 International Conference on Applied Electronics (AE)*, pp. 1–4, 2021.

[51] J. Adolf, P. Kán, B. Outram, H. Kaufmann, J. Doležal, and L. Lhotská, "Juggling in VR: Advantages of immersive virtual reality in juggling learning," in *25th ACM Symposium on Virtual Reality Software and Technology*, ACM, Nov. 2019.

[52] J. Adolf, J. Dolezal, P. Kutilek, M. Hourova, J. Hejda, I. Milerska, and L. Lhotska, "Automatic telerehabilitation system in a home environment using computer vision," *Studies in Health Technology and Informatics*, vol. 273, no. pHealth 2020, p. 142–148, 2020.

[53] J. A. et al., "Upper limb range of motion evaluation by a camera-based system," *IUPESM World Congres On Medical Physic And Biomedical Enginering*, vol. 11, pp. 13–13, Dec. 2022.

[54] M. Sýkorová, "Evaluation of the range of motion of the upper limbs using the openpose program," Master's thesis, Charles University, Prague, Czech Republic, 2021.

[55] T. Skalová, "Detection of signs of fatigue in functional tests in healthy population by openpose," Master's thesis, Charles University, Prague, Czech Republic, 2021.

[56] J. Adolf, J. Dolezal, P. Kutilek, J. Hejda, and L. Lhotska, "Single camera-based remote physical therapy: Verification on a large video dataset," *Applied Sciences*, vol. 12, no. 2, p. 799, 2022.

[57] A. Demeco, N. Marotta, L. Moggio, I. Pino, C. Marinaro, M. Barletta, A. Petraroli, A. Palumbo, and A. Ammendolia, "Quantitative analysis of movements in facial nerve palsy with surface electromyography and kinematic analysis," *Journal of Electromyography and Kinesiology*, vol. 56, p. 102485, Feb. 2021.

[58] L. Suso-Martí, R. L. Touche, A. Herranz-Gómez, S. Angulo-Díaz-Parreño, A. Paris-Alemany, and F. Cuenca-Martínez, "Effectiveness of telerehabilitation in physical therapist practice: An umbrella and mapping review with meta–meta-analysis," *Physical Therapy*, vol. 101, Feb. 2021.

[59] Y. Desmarais, D. Mottet, P. Slangen, and P. Montesinos, "A review of 3d human pose estimation algorithms for markerless motion capture," *CoRR*, vol. abs/2010.06449, 2020.

[60] R. J. Cotton, "Kinematic tracking of rehabilitation patients with markerless pose estimation fused with wearable inertial sensors," in *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, pp. 508–514, 2020.

[61] A. Peretti, F. Amenta, S. K. Tayebati, G. Nittari, and S. S. Mahdi, "Telerehabilitation: Review of the state-of-the-art and areas of application," *JMIR Rehabilitation and Assistive Technologies*, vol. 4, p. e7, July 2017.

[62] G. A. Farulla, D. Pianu, M. Cempini, M. Cortese, L. Russo, M. Indaco, R. Nerino, A. Chimienti, C. Oddo, and N. Vitiello, "Vision-based pose estimation for robot-mediated hand telerehabilitation," *Sensors*, vol. 16, p. 208, Feb. 2016.

[63] M. Ota, H. Tateuchi, T. Hashiguchi, T. Kato, Y. Ogino, M. Yamagata, and N. Ichihashi, "Verification of reliability and validity of motion analysis systems during bilateral squat using human pose tracking algorithm," *Gait & Posture*, vol. 80, pp. 62–67, July 2020.

[64] G. Hidalgo, Z. Cao, T. Simon, S.-E. Wei, Y. Raaj, H. Joo, and Y. Sheikh, "Openpose library, cmu-perceptual-computing-lab." `https://github.com/CMU-Perceptual-Computing-Lab/openpose`, 2021.

[65] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask r-CNN," IEEE, Oct. 2017.

[66] A. Ahmadyan, T. Hou, J. Wei, L. Zhang, A. Ablavatski, and M. Grundmann, "Instant 3d object tracking with applications in augmented reality," 2020.

[67] S. L. Colyer, M. Evans, D. P. Cosker, and A. I. T. Salo, "A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system," vol. 4, June 2018.

[68] Ł. Kidziński, B. Yang, J. L. Hicks, A. Rajagopal, S. L. Delp, and M. H. Schwartz, "Deep neural networks enable quantitative movement analysis using single-camera videos," *Nature Communications*, vol. 11, Aug. 2020.

[69] N. Marotta, A. Demeco, L. Moggio, and A. Ammendolia, "Why is telerehabilitation necessary? a pre-post COVID-19 comparative study of ICF activity and participation," *Journal of Enabling Technologies*, vol. 15, pp. 117–121, May 2021.

[70] J. Chua, L.-Y. Ong, and M.-C. Leow, "Telehealth using PoseNet-based system for in-home rehabilitation," vol. 13, p. 173, July 2021.

[71] Ó. G. Hernández, V. Morell, J. L. Ramon, and C. A. Jara, "Human pose detection for robotic-assisted and rehabilitation environments," vol. 11, p. 4183, May 2021.

[72] I. Takeda, A. Yamada, and H. Onodera, "Artificial intelligence-assisted motion capture for medical applications: a comparative study between markerless and passive marker motion capture," vol. 24, pp. 864–873, Dec. 2020.

[73] J. Stenum, C. Rossi, and R. T. Roemmich, "Two-dimensional video-based analysis of human gait using pose estimation," vol. 17, p. e1008935, Apr. 2021.

[74] N. Nakano, T. Sakura, K. Ueda, L. Omura, A. Kimura, Y. Iino, S. Fukashiro, and S. Yoshioka, "Evaluation of 3d markerless motion capture accuracy using OpenPose with multiple video cameras," vol. 2, May 2020.

[75] "Physiotools©, product id rg-pt1eng, general exercises second edition (english), tampere, finland."

[76] E. Hay, K. Dziedzic, N. Foster, G. Peat, D. van der Windt, B. Bartlam, M. Blagojevic-Bucknall, J. Edwards, E. Healey, M. Holden, R. Hughes, C. Jinks, K. Jordan, S. Jowett, M. Lewis, C. Mallen, A. Morden, E. Nicholls, B. N. Ong, M. Porcheret, J. Wulff, J. Kigozi, R. Oppong, Z. Paskins, and P. Croft, "Optimal primary care management of clinical osteoarthritis and joint pain in older people: a mixed-methods programme of systematic reviews, observational and qualitative studies, and randomised controlled trials," vol. 6, pp. 1–260, July 2018.

[77] A. Badiola-Bengoa and A. Mendez-Zorrilla, "A systematic review of the application of camera-based human pose estimation in the field of sport and physical exercise," vol. 21, p. 5996, Sept. 2021.

[78] “Anatomical planes.” `https://teachmeanatomy.info/the-basics/anatomical-terminology/planes/`. Accessed: 2021-11-03.

[79] D. Stirling, A. Hesami, C. Ritz, K. Adistambha, and F. Naghdy, “Symbolic modelling of dynamic human motions,” InTech, Feb. 2010.

[80] A. Karatsidis, M. Jung, H. M. Schepers, G. Bellusci, M. de Zee, P. H. Veltink, and M. S. Andersen, “Musculoskeletal model-based inverse dynamic analysis under ambulatory conditions using inertial motion capture,” vol. 65, pp. 68–77, Mar. 2019.

[81] S. Johnson and M. Everingham, “Learning effective human pose estimation from inaccurate annotation,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2011.

[82] W. Zhang, M. Zhu, and K. G. Derpanis, “From actemes to action: A strongly-supervised representation for detailed action understanding,” IEEE, Dec. 2013.

[83] V. Kazemi and J. Sullivan, “Using richer models for articulated pose estimation of footballers,” in *BMVC*, 2012.

[84] M. Garcia-Salguero, J. Gonzalez-Jimenez, and F.-A. Moreno, “Human 3d pose estimation with a tilting camera for social mobile robot interaction,” vol. 19, p. 4943, Nov. 2019.

[85] M. Slembrouck, H. Luong, J. Gerlo, K. Schütte, D. V. Cauwelaert, D. D. Clercq, B. Vanwanseele, P. Veelaert, and W. Philips, “Multiview 3d markerless human pose estimation from OpenPose skeletons,” pp. 166–178, Springer International Publishing, 2020.

[86] K.-D. Ng, S. Mehdizadeh, A. Iaboni, A. Mansfield, A. Flint, and B. Taati, “Measuring gait variables using computer vision to assess mobility and fall risk in older adults with dementia,” *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 8, pp. 1–9, 2020.

[87] L. Needham, M. Evans, D. P. Cosker, and S. L. Colyer, “Can markerless pose estimation algorithms estimate 3d mass centre positions and velocities during linear sprinting activities?,” vol. 21, p. 2889, Apr. 2021.

[88] J. Adolf, Y. Segal, M. Turna, T. Nováková, J. Doležal, P. Kutílek, J. Hejda, O. Hadar, and L. Lhotská, “Evaluation of functional tests performance using a camera-based and machine learning approach,” *PLOS ONE*, vol. 18, p. e0288279, Nov. 2023.

[89] T. Jesus, M. Landry, and H. Hoenig, “Global need for physical rehabilitation: Systematic analysis from the global burden of disease study 2017,” *International Journal of Environmental Research and Public Health*, vol. 16, p. 980, Mar. 2019.

[90] J.-L. Thonnard and M. Penta, "Functional assessment in physiotherapy. a literature review," *Europa medicophysica*, vol. 43, pp. 525–41, 01 2008.

[91] R. P. McGovern, R. L. Martin, J. J. Christoforetti, and B. R. Kivlan, "Evidence-based procedures for performing the single leg squat and step-down tests in evaluation of non-arthritic hip pain: A literature review," *Int J Sports Phys Ther*, vol. 13, no. 3, pp. 526–536, 2018.

[92] S. Schurr, A. Marshall, J. Resch, and S. Saliba, "Two-dimensional video analysis is comparable to 3d motion capture in lower extremity movement assessment," *International Journal of Sports Physical Therapy*, vol. 12, no. 1, pp. 163–172, 2017.

[93] M. Harris-Hayes, K. Steger-May, C. Koh, N. K. Royer, V. Graci, and G. B. Salsich, "Classification of lower extremity movement patterns based on visual assessment: Reliability and correlation with 2-dimensional video analysis," *Journal of Athletic Training*, vol. 49, pp. 304–310, June 2014.

[94] V. Ugalde, C. Brockman, Z. Bailowitz, and C. D. Pollard, "Single leg squat test and its relationship to dynamic knee valgus and injury risk screening," *PM&ampR*, vol. 7, pp. 229–235, Aug. 2014.

[95] P. Kline, D. Johnson, M. Ireland, and B. Noehren, "Clinical predictors of knee mechanics at return to sport following acl reconstruction," *Medicine Science in Sports Exercise*, vol. 48, p. 1, 12 2015.

[96] S.-P. Lee, R. B. Souza, and C. M. Powers, "The influence of hip abductor muscle performance on dynamic postural stability in females with patellofemoral pain," *Gait &amp Posture*, vol. 36, pp. 425–429, July 2012.

[97] P. Levinger, W. Gilleard, and C. Coleman, "Femoral medial deviation angle during a one-leg squat test in individuals with patellofemoral pain syndrome," *Physical Therapy in Sport*, vol. 8, no. 4, pp. 163–168, 2007.

[98] A. L. Hatton, J. L. Kemp, S. G. Brauer, R. A. Clark, and K. M. Crossley, "Impairment of dynamic single-leg balance performance in individuals with hip chondropathy," *Arthritis Care &amp Research*, vol. 66, pp. 709–716, Apr. 2014.

[99] M. F. Mostaed, D. M. Werner, and J. A. Barrios, "2d and 3d kinematics during lateral step-down testing in individuals with anterior cruciate ligament reconstruction," *International Journal of Sports Physical Therapy*, vol. 13, no. 1, pp. 77–85, 2018.

[100] E. Ageberg, K. L. Bennell, M. A. Hunt, M. Simic, E. M. Roos, and M. W. Creaby, "Validity and inter-rater reliability of medio-lateral knee motion observed during a single-limb mini squat," *BMC Musculoskeletal Disorders*, vol. 11, Nov. 2010.

[101] R. M. Barker-Davies, A. Roberts, A. N. Bennett, D. T. Fong, P. Wheeler, and M. P. Lewis, "Single leg squat ratings by clinicians are reliable and predict excessive hip internal rotation moment," *Gait &amp Posture*, vol. 61, pp. 453–458, Mar. 2018.

[102] B. Debnath, M. O'Brien, M. Yamaguchi, and A. Behera, "A review of computer vision-based approaches for physical rehabilitation and assessment," *Multimedia Systems*, vol. 28, pp. 209–239, June 2021.

[103] S. L. Colyer, M. Evans, D. P. Cosker, and A. I. T. Salo, "A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system," vol. 4, June 2018.

[104] H.-S. Fang, S. Xie, Y.-W. Tai, and C. Lu, "Rmpe: Regional multi-person pose estimation," 2016.

[105] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schieke, "Deepercut: A deeper, stronger, and faster multi-person pose estimation model," in *European Conference on Computer Vision (ECCV)*, 2016.

[106] A. Ahmadyan, T. Hou, J. Wei, L. Zhang, A. Ablavatski, and M. Grundmann, "Instant 3d object tracking with applications in augmented reality," 2020.

[107] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, "Microsoft coco: Common objects in context," 2014.

[108] M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2d human pose estimation: New benchmark and state of the art analysis," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3686–3693, 2014.

[109] Y. Segal, Y. Yona, O. Danan, R. Birman, O. Hadar, P. Kutilek, J. Hejda, M. Hourova, P. Kral, L. Lhotska, J. Dolezal, and J. Adolf, "Camera setup and OpenPose software without GPU for calibration and recording in telerehabilitation use," in *2021 International Conference on e-Health and Bioengineering (EHB)*, IEEE, Nov. 2021.

[110] M. Ota, H. Tateuchi, T. Hashiguchi, T. Kato, Y. Ogino, M. Yamagata, and N. Ichihashi, "Verification of reliability and validity of motion analysis systems during bilateral squat using human pose tracking algorithm," *Gait Posture*, vol. 80, 05 2020.

[111] S. M. Remedios and S. L. Fischer, "Towards the use of 2d video-based markerless motion capture to measure and parameterize movement during functional capacity evaluation," *Journal of Occupational Rehabilitation*, vol. 31, pp. 754–767, Sept. 2021.

[112] M. Ota, H. Tateuchi, T. Hashiguchi, T. Kato, Y. Ogino, M. Yamagata, and N. Ichihashi, "Verification of reliability and validity of motion analysis systems during bilateral squat using human pose tracking algorithm," *Gait & Posture*, vol. 80, pp. 62–67, 2020.

[113] T. E. Ward, E. Delahunt, B. Caulfield, D. F. Whelan, and M. A. O'Reilly, "Technology in rehabilitation: Evaluating the single leg squat exercise with wearable inertial measurement units," *Methods of Information in Medicine*, vol. 56, no. 02, pp. 88–94, 2017.

[114] J. Mitternacht, A. Hermann, and P. Carqueville, "Acquisition of lower-limb motion characteristics with a single inertial measurement unit—validation for use in physiotherapy," *Diagnostics*, vol. 12, p. 1640, July 2022.

[115] M. Seifallahi, H. Soltanizadeh, A. H. Mehraban, and F. Khamseh, "Alzheimer's disease detection using skeleton data recorded with kinect camera," *Cluster Computing*, vol. 23, pp. 1469–1481, Nov. 2019.

[116] G. Herman, O. Nakdimon, P. Levinger, and S. Springer, "Agreement of an evaluation of the forward-step-down test by a broad cohort of clinicians with that of an expert panel," *Journal of Sport Rehabilitation*, vol. 25, pp. 227–232, Aug. 2016.

[117] J. L. Fleiss, "Measuring nominal scale agreement among many raters.," *Psychological Bulletin*, vol. 76, no. 5, pp. 378–382, 1971.

[118] P. Kutilek, J. Hejda, L. Lhotska, J. Adolf, J. Dolezal, M. Hourova, P. Kral, Y. Segal, R. Birman, and O. Hadar, "Camera system for efficient non-contact measurement in distance medicine," in *2020 19th International Conference on Mechatronics - Mechatronika (ME)*, IEEE, Dec. 2020.

[119] W. Welling, A. Benjaminse, R. Seil, K. Lemmink, and A. Gokeler, "Altered movement during single leg hop test after ACL reconstruction: implications to incorporate 2-d video movement analysis for hop tests," *Knee Surgery, Sports Traumatology, Arthroscopy*, vol. 26, pp. 3012–3019, Mar. 2018.

[120] S. M. Remedios and S. L. Fischer, "Towards the use of 2d video-based markerless motion capture to measure and parameterize movement during functional capacity evaluation," *Journal of Occupational Rehabilitation*, vol. 31, pp. 754–767, Sept. 2021.

[121] J. Adolf, P. Kán, T. Feuchtner, B. Adolfová, J. Doležal, and L. Lhotská, "Offistretch: camera-based real-time feedback for daily stretching exercises," *The Visual Computer*, May 2024.

[122] S. E. Kotowski, K. G. Davis, and T. Gerding, "Almost a year in: Virtual offices remained an ergonomic trouble spot," *Work*, vol. 71, pp. 319–326, Feb. 2022.

[123] I.-M. Lee, E. J. Shiroma, F. Lobelo, P. Puska, S. N. Blair, and P. T. Katzmarzyk, "Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy," *The Lancet*, vol. 380, pp. 219–229, July 2012.

[124] J. H. Park, J. H. Moon, H. J. Kim, M. H. Kong, and Y. H. Oh, "Sedentary lifestyle: Overview of updated evidence of potential health risks," *Korean Journal of Family Medicine*, vol. 41, pp. 365–373, Nov. 2020.

[125] H. Daneshmandi, A. Choobineh, H. Ghaem, and M. Karimi, "Adverse effects of prolonged sitting behavior on the general health of office workers," *Journal of Lifestyle Medicine*, vol. 7, pp. 69–75, July 2017.

[126] C. González-González, N. G. Río, and V. Navarro-Adelantado, "Exploring the benefits of using gamification and videogames for physical exercise: a review of state of art," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 2, p. 46, 2018.

[127] T. B. F. Pacheco, C. S. P. de Medeiros, V. H. B. de Oliveira, E. R. Vieira, and F. A. C. de Cavalcanti, "Effectiveness of exergames for improving mobility and balance in older adults: a systematic review and meta-analysis," *Systematic Reviews*, vol. 9, July 2020.

[128] A. Andrade, C. K. Correia, and D. R. Coimbra, "The psychological effects of exergames for children and adolescents with obesity: A systematic review and meta-analysis," *Cyberpsychology, Behavior, and Social Networking*, vol. 22, pp. 724–735, Nov. 2019.

[129] V. N. Soares, H. M. Yoshida, T. S. Magna, R. A. C. Sampaio, and P. T. Fernandes, "Comparison of exergames versus conventional exercises on the cognitive skills of older adults: a systematic review with meta-analysis," *Archives of Gerontology and Geriatrics*, vol. 97, p. 104485, Nov. 2021.

[130] A. E. Staiano, M. A. Adams, and G. J. Norman, "Motivation for exergame play inventory: Construct validity and relationship to game play," *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, vol. 13, July 2019.

[131] M. H. Jarrahi, N. Gafinowitz, and G. Shin, "Activity trackers, prior motivation, and perceived informational and motivational affordances," *Personal and Ubiquitous Computing*, vol. 22, no. 2, pp. 433–448, 2018.

[132] G. Hayes, K. P. Dowd, C. MacDonncha, and A. E. Donnelly, "Tracking of physical activity and sedentary behavior from adolescence to young adulthood: A systematic literature review," *Journal of Adolescent Health*, vol. 65, pp. 446–454, Oct. 2019.

[133] L. Laranjo, D. Ding, B. Heleno, B. Kocaballi, J. C. Quiroz, H. L. Tong, B. Chahwan, A. L. Neves, E. Gabarron, K. P. Dao, D. Rodrigues, G. C. Neves, M. L. Antunes, E. Coiera, and D. W. Bates, "Do smartphone applications and activity trackers increase physical activity in adults? systematic review, meta-analysis and metaregression," *British Journal of Sports Medicine*, vol. 55, pp. 422–432, Dec. 2020.

[134] A. Filippeschi, N. Schmitz, M. Miezal, G. Bleser, E. Ruffaldi, and D. Stricker, "Survey of motion tracking methods based on inertial sensors: A focus on upper limb human motion," *Sensors*, vol. 17, p. 1257, June 2017.

[135] F. Anderson, T. Grossman, J. Matejka, and G. Fitzmaurice, "YouMove," in *Proceedings of the 26th annual ACM symposium on User interface software and technology*, ACM, Oct. 2013.

[136] F. Losilla and F. Rosique, "An augmented reality mirror exergame using 2d pose estimation," in *Proceedings of the 14th International Conference on Software Technologies*, SCITEPRESS - Science and Technology Publications, 2019.

[137] R. R. Kanase, A. N. Kumavat, R. D. Sinalkar, and S. Somani, "Pose estimation and correcting exercise posture," *ITM Web of Conferences*, vol. 40, p. 03031, 2021.

[138] H. Elsayed, P. Hoffmann, S. Günther, M. Schmitz, M. Weigel, M. Mühlhäuser, and F. Müller, "Cameraready: Assessing the influence of display types and visualizations on posture guidance," DIS '21, (New York, NY, USA), p. 1046–1055, Association for Computing Machinery, 2021.

[139] S. L. Colyer, M. Evans, D. P. Cosker, and A. I. T. Salo, "A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system," *Sports Medicine - Open*, vol. 4, June 2018.

[140] A. Badiola-Bengoa and A. Mendez-Zorrilla, "A systematic review of the application of camera-based human pose estimation in the field of sport and physical exercise," *Sensors*, vol. 21, p. 5996, Sept. 2021.

[141] K. Gerling, I. Livingston, L. Nacke, and R. Mandryk, "Full-body motion-based game interaction for older adults," in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1873–1882, 2012.

[142] M. Rice, M. Wan, M.-H. Foo, J. Ng, Z. Wai, J. Kwok, S. Lee, and L. Teo, "Evaluating gesture-based games with older adults on a large screen display," in *Proceedings of the 2011 ACM SIGGRAPH Symposium on Video Games*, Sandbox '11, (New York, NY, USA), p. 17–24, Association for Computing Machinery, 2011.

[143] M. Kyan, G. Sun, H. Li, L. Zhong, P. Muneesawang, N. Dong, B. Elder, and L. Guan, "An approach to ballet dance training through ms kinect and visualization in a cave virtual reality environment," *ACM Trans. Intell. Syst. Technol.*, vol. 6, mar 2015.

[144] Y. Yang, H. Leung, L. Yue, and L. Deng, "Generating a two-phase lesson for guiding beginners to learn basic dance movements," *Computers & Education*, vol. 61, pp. 1–20, 2013.

[145] P. T. Chua, R. Crivella, B. Daly, N. Hu, R. Schaaf, D. Ventura, T. Camill, J. Hodgins, and R. Pausch, "Training for physical tasks in virtual environments: Tai chi," in *IEEE Virtual Reality, 2003. Proceedings.*, pp. 87–94, IEEE, 2003.

[146] J. Bailenson, K. Patel, A. Nielsen, R. Bajscy, S.-H. Jung, and G. Kurillo, "The effect of interactivity on learning physical actions in virtual reality," *Media Psychology*, vol. 11, no. 3, pp. 354–376, 2008.

[147] P. Le Noury, T. Buszard, M. Reid, and D. Farrow, "Examining the representativeness of a virtual reality environment for simulation of tennis performance," *Journal of sports sciences*, vol. 39, pp. 1–9, 09 2020.

[148] T. N. Hoang, M. Reinoso, F. Vetere, and E. Tanin, "Onebody: Remote posture guidance system using first person view in virtual environment," in *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, NordiCHI '16, (New York, NY, USA), Association for Computing Machinery, 2016.

[149] H. S. Park, G. A. Lee, B.-K. Seo, and M. Billinghurst, "User experience design for a smart-mirror-based personalized training system," *Multimedia Tools and Applications*, vol. 80, pp. 31159–31181, Nov. 2020.

[150] Q. Zhou, A. Irlitti, D. Yu, J. Goncalves, and E. Velloso, "Movement guidance using a mixed reality mirror," in *Designing Interactive Systems Conference*, ACM, June 2022.

[151] B. Dittakavi, D. Bavikadi, S. V. Desai, S. Chakraborty, N. Reddy, V. N. Balasubramanian, B. Callepalli, and A. Sharma, "Pose tutor: An explainable system for pose correction in the wild," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, June 2022.

[152] M. Fieraru, M. Zanfir, S. C. Pirlea, V. Olaru, and C. Sminchisescu, "AIFit: Automatic 3d human-interpretable feedback models for fitness training," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, June 2021.

[153] O. Stamm and A. Heimann-Steinert, "Accuracy of monocular two-dimensional pose estimation compared with a reference standard for

kinematic multiview analysis: Validation study," *JMIR mHealth and uHealth*, vol. 8, p. e19608, Dec. 2020.

[154] J. Saldaña, "The coding manual for qualitative researchers," *The coding manual for qualitative researchers*, pp. 1–440, 2021.

[155] T. Helten, M. Muller, H.-P. Seidel, and C. Theobalt, "Real-time body tracking with one depth camera and inertial sensors," in *2013 IEEE International Conference on Computer Vision*, IEEE, Dec. 2013.

[156] B. Pueo, "High speed cameras for motion analysis in sports science," *Journal of Human Sport and Exercise*, vol. 11, no. 1, 2016.

[157] C. Poynton, *Digital Video and HD: Algorithms and Interfaces*. Elsevier, 1 ed., 12 2002.

[158] H. Strasburger, I. Rentschler, and M. Juttner, "Peripheral vision and pattern recognition: A review," *Journal of Vision*, vol. 11, pp. 13–13, Dec. 2011.

[159] E. Chaplin, C. Karatzios, and C. Benaim, "Clinical applications of virtual reality in musculoskeletal rehabilitation: A scoping review," *Healthcare*, vol. 11, p. 3178, Dec. 2023.

[160] U. G. Longo, A. Carnevale, F. Andreoli, I. Mannocchi, M. Bravi, M. S. H. Sassi, F. Santacaterina, M. Carli, E. Schena, and R. Papalia, "Immersive virtual reality for shoulder rehabilitation: evaluation of a physical therapy program executed with oculus quest 2," *BMC Musculoskeletal Disorders*, vol. 24, Nov. 2023.

[161] E. Chaplin, C. Karatzios, and C. Benaim, "Clinical applications of virtual reality in musculoskeletal rehabilitation: A scoping review," *Healthcare*, vol. 11, p. 3178, Dec. 2023.

Chapter 11

List of Publications

11.1 Publications Related to the Dissertation

11.1.1 Journal publications

Jindřich Adolf, Peter Kán, Tiare Feutchner, Barbora Adolfová, Jaromír Doležal, Lenka Lhotská, “OffiStretch: Camera-based Real-time Feedback for Daily Stretching,” *Visual Computer, Springer*, Springer Science and Business Media LLC, May. 2024 DOI: 10.1007/s00371-024-03450-y

Jindřich Adolf, Yoram Segal, Matyáš Turna, Tereza Nováková, Jaromír Doležal, Patrik Kutílek, Jan Hejda, Ofer Hadar and Lenka Lhotská “Evaluation of functional tests performance using a camera-based and machine learning approach,” *PLoS ONE*, 18(11): e0288279., Nov. 2023. DOI: 10.1371/journal.pone.0288279

Jindřich Adolf, Jaromír Doležal, Patrik Kutílek, Jan Hejda and Lenka Lhotská “Single Camera-Based Remote Physical Therapy: Verification on a Large Video Dataset” *Applied Sciences*, 12, no. 2: 799., Jan. 2022. DOI: 10.3390/app12020799.

11.1.2 Conference Publications

Jindrich Adolf, Jaromír Doležal, Martin Macaš, Lenka Lhotská “Remote physical therapy: Requirements for a single RGB camera motion sensing,” International Conference on Applied Electronics (AE), pp. 1–4, 2021.

Yoram Segal, Yuval Yona1, Omer Danan1, Raz Birman, Ofer Hadar, Patrik Kutilek, Jan Hejda, Michaela Hourova, Pavel Kral, Lenka Lhotska, Jaromir Dolezal and Jindrich Adolf, “Camera Setup and OpenPose Software without GPU for Calibration and Recording in Telerehabilitation Use,” Camera Setup and OpenPose Software without GPU for Calibration and Recording in Telerehabilitation Use, 2021.

Jindřich Adolf, Jaromír Doležal, Patrik Kutílek, Michaela Hourov, Jan Hejda, Iva Milerská and Lenka Lhotská “Automatic Telerehabilitation System in a Home Environment Using Computer Vision.,” *Stud Health Technol Inform.*, vol. 26, pp. 3012–3019, Sep. 2020.

Patrik Kutílek, Jan Hejda, Lenka Lhotská, Jindřich Adolf, Jaromír Doležal, Michaela Hourová, Yoram Segal, Raz Birman, Ofer Hadar and Pavel Král “Camera System for Efficient non-contact Measurement in Distance Medicine” *19th International Conference on Mechatronics*, Mechatronika (ME), Prague, Czech Republic, 2020, pp. 1-6, 2020.

Jindřich Adolf, Peter Kán, Bejamin Outram, Hannes Kaufmann, Jaromír Doležal and Lenka Lhotská “Juggling in vr: Advantages of immersive virtual reality in juggling learning,” *VRST: ACM Symposium on Virtual Reality Software and Technology*, 2019

Jindřich Adolf, Jaromír Doležal and Lenka Lhotská, “Affordable Personalized, Immersive VR Motor Rehabilitation System with Full Body Tracking,” *Stud Health Technol Inform.*, 261:75-81. PMID: 31156094, 2019.

11.2 Other Publications

Lenka Lhotská, Jan Husák, Jakub Stejskal, Martin Kotek, Jaromír Doležal, Jindřich Adolf “Role of virtual reality in the life of ageing population,” *Czech Technical University in Prague, Faculty of Transportation Sciences*, Neural Network World, 2022.

Jiří Pětník, Lenka Lhotská, Jaromír Doležal and Jindřich Adolf, “Behavioural Data Modeling: A Case Study in IoT.,” *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies* , vol. 26, pp. 3012–3019, 2019.

Patrik Kutilek, Petr Volf, Jan Hejda, Pavel Smrcka, Jindrich Adolf, Vaclav Krivanek, Lenka Lhotska, Karel Hana, Radek Doskocil, Jiri Kacer, Ludek Cicmanec “Non-contact Measurement Systems for Physiological Data Monitoring of Military Pilots During Training on Simulators: Review and Application,” *International Conference on Military Technologies*, 2019.

Lenka Lhotská, Jindřich Adolf, Jaromír Doležal “Virtual Reality in Research and Education: A Case Study,” *29th Annual Conference of the European Association for Education in Electrical and Information Engineering (EAEEIE)*, IEEE 2019.

Lenka Lhotska, Jaromír Doležal, Jindřich Adolf, Jiří Potůček, Miroslav Křížek, Baha Chbani “Personalized monitoring and assistive systems: case study of efficient home solutions, 49:19-28. PMID: 29866952, 2018” *pHealth 2018*, pp. 2394-2399.

Jindřich Adolf, Martin Macaš, Lenka Lhotská, Jaromír Doležal “Deep neural network based body posture recognitions and fall detection from low resolution infrared array sensor, 2018” *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2394-2399.

Miroslav Macík, Kateřina Pražáková, Anna Kutíková, Zdeněk Míkovec, Jindřich Adolf, Jan Havlík and Ivana Jíleková “Breathing friend: tackling stress through portable tangible breathing artifact,” *Human-Computer Interaction, Springer International Publishing.* , INTERACT 2017 (pp. 106–115).

Appendix A

Acknowledgements of Contribution and Authorship

ACKNOWLEDGEMENT OF CONTRIBUTION AND AUTHORSHIP

Date: September 20th, 2023

Title of the Manuscript/Research: Single Camera-Based Remote Physical Therapy: Verification on a Large Video Dataset, 2022 doi: 10.3390/app12020799

"I, Jindřich Adolf, affirm that the main results and findings of the paper titled "Single Camera-Based Remote Physical Therapy: Verification on a Large Video Dataset" are the results of my research and work. This statement is made in the context of my doctoral thesis submission."

The following individuals have made significant contributions by providing their expertise, assisting with the creation and review of the manuscript, and managing various aspects of the project:

The following individuals have been instrumental in various capacities during the course of this research and are rightfully acknowledged as co-authors:

Ing. Jaromír Doležal, Ph.D.
doc. Ing. Patrik Kutílek, MSc., Ph.D.
Ing. Jan Hejda, Ph.D.
doc. Ing. Lenka Lhotská, CSc.

We, the undersigned co-authors, acknowledge and affirm that while the foundational results and contributions of the aforementioned manuscript have been primarily generated by Jindřich Adolf, our considerable involvement in offering expertise, assisting in manuscript creation and review, and project management has been pivotal to the project's success. This confirms our recognition and inclusion as major contributors and co-authors.

We support Jindřich Adolf's use and submission of this research as a component of the doctoral thesis. Furthermore, we confirm that our roles and contributions have been transparently and accurately credited in the manuscript.

Signed,

Ing. Jindřich Adolf ______________ Date: 20.9.2023

Ing. Jaromír Doležal, Ph.D. ______________ Date: 20.9.2023

doc. Ing. Patrik Kutílek, MSc., Ph.D. ______________ Date: 21.9.2023

Ing. Jan Hejda, Ph.D. ______________ Date: 21.9.2023

doc. Ing. Lenka Lhotská, CSc. Lhotská ______________ Date: 21.9.2023

ACKNOWLEDGEMENT OF CONTRIBUTION AND AUTHORSHIP

Date September 20th, 2023

Title of the Manuscript/Research: Evaluation of functional tests performance using a camera-based and machine learning approacht, 2022 doi: 10.1371/journal.pone.0288279

"I, Jindřich Adolf, affirm that the main results and findings of the paper titled "Evaluation of functional tests performance using a camera-based and machine learning approach" are the results of my research and work. This statement is made in the context of my doctoral thesis submission."

The following individuals have made significant contributions by providing their expertise, assisting with the creation and review of the manuscript, and managing various aspects of the project:

The following individuals have been instrumental in various capacities during the course of this research and are rightfully acknowledged as co-authors:

Yoram Segal
Mgr. Matyáš Turna
PhDr. Tereza Nováková, PhD.
Ing. Jaromír Doležal, Ph.D.
doc. Ing. Patrik Kutílek, MSc., Ph.D.
Ing. Jan Hejda, Ph.D.
Prof. Ofer Hadar
doc. Ing. Lenka Lhotská, CSc.

We, the undersigned co-authors, acknowledge and affirm that while the foundational results and contributions of the aforementioned manuscript have been primarily generated by Jindřich Adolf, our considerable involvement in offering expertise, assisting in manuscript creation and review, and project management has been pivotal to the project's success. This confirms our recognition and inclusion as major contributors and co-authors.

We support Jindřich Adolf's use and submission of this research as a component of the doctoral thesis. Furthermore, we confirm that our roles and contributions have been transparently and accurately credited in the manuscript.

Signed,

Ing. Jindřich Adolf ____________ Date: 20.9.2023

Yoram segal ____________ Date: 29/11/2023

Mgr. Matyáš Turna ____________ Date: 22.1.2024

PhDr. Tereza Nováková, PhD. ____________ Date: 13 01 2024

Ing. Jaromír Doležal, Ph.D. ____________ Date: 25.09.2023

doc. Ing. Patrik Kutílek, MSc., Ph.D. ____________ Date: 21.9.2023

Ing. Jan Hejda, Ph.D. ____________ Date: 21.9.2023

Prof. Ofer Hadar ____________ Date: 5.1.2024

doc. Ing. Lenka Lhotská, CSc. ____________ Date: 21.9.2023

ACKNOWLEDGEMENT OF CONTRIBUTION AND AUTHORSHIP

Date: May 9th, 2023

Title of the Manuscript/Research: "OffiStretch: Camera-Based Real-Time Feedback for Daily Stretching Exercises."

"I, Jindřich Adolf, affirm that the main results and findings of the paper titled "OffiStretch: Camera-Based Real-Time Feedback for Daily Stretching Exercises." are the results of my research and work. This statement is made in the context of my doctoral thesis submission."

The following individuals have made significant contributions by providing their expertise, assisting with the creation and review of the manuscript, and managing various aspects of the project:

The following individuals have been instrumental in various capacities during this research and are rightfully acknowledged as co-authors:

Ing. Jindřich Adolf
Mag. Dr. Peter Kán
Tiare Feuchtner. PhD, MSc.
MgA. Barbora Adolfová
Ing. Jaromír Doležal, Ph.D.
doc. Ing. Lenka Lhotská, CSc.

We, the undersigned co-authors, acknowledge and affirm that while the foundational results and contributions of the aforementioned manuscript have been primarily generated by Jindřich Adolf, our considerable involvement in offering expertise, assisting in manuscript creation and review, and project management has been pivotal to the project's success. This confirms our recognition and inclusion as major contributors and co-authors.

We support Jindřich Adolf's use and submission of this research as a component of the doctoral thesis. Furthermore, we confirm that our roles and contributions have been transparently and accurately credited in the manuscript.

Signed,

Ing. Jindřich Adolf ____________ Date: 9.5.2024

Mag. Dr. Peter Kán ____________ Date: 14.5.2024

Tiare Feuchtner. PhD, MSc. ____________ Date: ____________

MgA. Barbora Adolfová ____________ Date: 11.5.2024

Ing. Jaromír Doležal, Ph.D. ____________ Date: 14.5.2024

doc. Ing. Lenka Lhotská, CSc. ____________ Date: 14.5.2024