FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Mobility Analysis for Frailty Syndrome Assessment

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Visual Computing

by

Stefan Doppelbauer
Registration Number 0656930

to the Faculty of Informatics

at the TU Wien

Advisor: PD. Dr. Martin Kampel
Assistance: Univ.Ass. Dr. Sebastian Zambanini
Dr. Rainer Planinc

Vienna, 24th August, 2017

Stefan Doppelbauer

Martin Kampel

Erklärung zur Verfassung der Arbeit

Stefan Doppelbauer
Tolleterau 132, 4710 St. Georgen bei Grieskirchen

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 24. August 2017

Stefan Doppelbauer

Acknowledgements

First, I want to thank my supervisor, Dr. Martin Kampel, for his advice and support. I also want to thank Dr. Sebastian Zambanini and Dr. Rainer Planinc for their input regarding the data acquisition and evaluation of this work. I also want to express my gratitude towards the staff from the nursing facility of Grieskirchen, Austria for their help in conducting the experiments and all the people who participated in the recording of my data.

Furthermore, I want to thank my beloved wife Mary Claire for her love, patience, understanding and support during the time of writing this thesis. And last, I want to thank my two beloved children Anna Louisa and Karl Raphael for cheering me up sometimes. Thanks to all of you.

Kurzfassung

Die menschliche Mobilität ist ein bedeutender Gesundheitsindikator, vor allem für potentiell gebrechliche, ältere Menschen. Die Analyse des menschlichen Gangs basiert gegenwärtig auf kostenintensiven oder intrusiven Technologien. Tiefenkameras wie die Microsoft Kinect bieten eine günstige und unaufdringliche Alternative zur Analyse der menschlichen Mobilität.

In dieser Arbeit wird Mobilitätsanalyse mittels Analyse des menschlichen Gangs und mittels automatischer Analyse des Timed UP & Go (TUG) Tests durchgeführt. Für beide Ansätze des Mobilitätsassessments wird jeweils eine Methode basierend auf Tiefendaten und eine Methode basierend auf Skelettdaten vorgestellt. Um die vorgestellten Ansätze zur Mobilitätsanalyse zu evaluieren, werden drei Datensätze aufgenommen und manuell markiert. Der erste Datensatz beinhaltet 234 Gangsequenzen von gesunden Erwachsenen, die auf vordefinierten Bodenmarkierungen gegangen sind. Der zweite Datensatz enthält 22 Gangsequenzen von 11 älteren Erwachsenen und der dritte Datensatz enthält 11 TUG Tests, die von älteren Erwachsenen durchgeführt werden.

Der erste Ansatz zur Mobilitätsanalyse untersucht den menschlichen Gang durch die Extrahierung von fünf raum- und zeitabhängigen Gangparametern. Die extrahierten Parameter sind unter anderem Schrittgeschwindigkeit, Schrittlänge und Schrittdauer. Als erstes wird der vorgestellte Ansatz zur Ganganalyse auf dem ersten Datensatz, der aus Gangsequenzen von gesunden Erwachsenen besteht, validiert. In diesem Datensatz werden alle Zeitpunkte für das Auftreten der Ferse und das Abheben der Zehen händisch markiert um den Fehler der extrahierten Gangparameter ermitteln zu können. Außerdem wurde die Auswirkung von 3 verschiedenen Gangpfaden, 3 verschiedenen Schrittweiten und 2 verschiedenen Gangrichtungen auf den Fehler der Schrittlänge untersucht. Als zweites wird die Methode zur Ganganalyse auf den zweiten Datensatz, der aus Gangsequenzen von älteren Erwachsenen besteht, angewendet. In diesem Datensatz werden die Anzahl der Schritte für jede Gangsequenz manuell markiert. Es zeigt sich, dass die vorgestellte Methode, die auf dem Stand der Technik basiert, sich nur schlecht für die Erkennung von Schritten von älteren Erwachsenen eignet. Mittels einer vorgeschlagenen Modifikation der Methode können die Schritte allerdings korrekt erkannt werden. Weiters wird der Status der Gebrechlichkeit der älteren Erwachsenen basierend auf Gangdaten, der benötigten TUG Zeit und eines validierten Fragebogens ermittelt.

Der zweite Ansatz zur Mobilitätsanalyse präsentiert eine Methode zur Kinect-basierten, automatischen Analyse des TUG Tests. Aktuell wird die benötigte Zeit für den TUG Test mit einer Stoppuhr von medizinischem Personal gemessen. Durch die automatische Analyse der benötigten Zeit kann der Arbeitsaufwand für das medizinische Personal gesenkt werden. Die vorgestellte Methode erkennt nicht nur die benötigte Zeit für den Test, sondern auch den Beginn und das Ende von sechs Phasen des TUG Tests. Diese werden erkannt, indem die Bewegung der Person basierend auf der Geschwindigkeit des Schwerpunkts bzw. des zentralen Schultergelenks im Skelettmodel analysiert wird. Der vorgestellte Ansatz zur Automatisierung des TUG Tests weist deutlich niedrigere Fehler bei der Messung der benötigten Zeit, verglichen mit der Stoppuhr, auf. Weiters zeigt ein Vergleich mit einer früheren Methode, dass der vorgestellte Ansatz auf dem aufgenommenen Datensatz exaktere Ergebnisse, für die benötigte TUG Zeit und die Erkennung der TUG Phasen, zurückliefert.

Abstract

Human mobility is an important health indicator, especially for older adults potentially transitioning to frailty. Currently, the analysis of human mobility is based on expensive or intrusive technologies. Depth camera devices, such as the Microsoft Kinect, have been demonstrated to be a valid low-cost alternative for assessing a person's mobility.

In this work, mobility assessment is approached based on the analysis of human gait and the automated analysis of the Timed Up & Go (TUG) test. For both mobility assessment approaches, one method for depth and one method for skeleton data is proposed. In order to evaluate the proposed mobility analysis approaches, three human mobility datasets have been acquired and manually labeled. The first dataset features 234 walking sequences of healthy adults, who walk over predefined floor markers. The second dataset is a gait dataset of 11 older adults that contains 22 walking sequences and the third dataset consists of 11 TUG tests from elderly participants.

The first human mobility approach analyzes human gait by extracting five spatiotemporal gait parameters from walking sequences. The extracted parameters are gait speed, stride length, step length, stride time and step time. First, the proposed gait analysis approach is validated on the healthy adults gait dataset. The dataset is manually labeled for heel strike and toe-off events to allow measuring the error of both spatial and temporal parameters. Moreover, extracted step lengths are evaluated under 3 different walking paths, 3 different step sizes and 2 different walking directions. Second, the gait analysis approach is applied on the elderly gait dataset to validate the approach under real-world conditions. An evaluation based on the number of steps demonstrates, that current state-of-the-art gait analysis approaches fail to correctly identify steps of older adults. A modification is proposed in order to successfully detect steps of older adults and evaluated on the elderly gait dataset. Moreover, the frailty status of the elderly participants is assessed based on extracted gait information, TUG score and a validated questionnaire.

The second human mobility approach presents a Kinect-based automated analysis of the TUG test. Currently, a medical assistant has to manually measure the time a patient takes to complete the test with a stopwatch in order to obtain a TUG score. By automatizing the assessment of the test, workload of medical staff is reduced. In addition to automatically determining the TUG score, the proposed TUG analysis approach detects the start and end of six TUG phases. TUG events are detected by analyzing the movement of the subject based on the velocity of the centroid or spine shoulder

joint. The proposed TUG analysis approaches show significantly lower TUG time errors compared to the manually measured time using a stopwatch. Moreover, comparison with a state-of-the-art approach shows improved accuracies for both the detection of TUG score and TUG events.

Contents

CHAPTER 1

Introduction

1.1 Introduction

Frailty is a geriatric condition that limits the ability of daily life activities and thereby increases the risk of falls, disability, hospitalization and mortality and the need for healthcare services [86]. It is one of the most problematic expressions of aging [59]. It is highly prevalent among the older population. Around 5% of people aged over 60 years [291], 10% aged over 65 and between 25% and 50% over 85 suffer from frailty syndrome [59].

1.1.1 Motivation: Why detect Frailty?

Human populations are growing old at an accelerated speed. The world population of 65 years or older was estimated in 2004 at 461 million people [142] and in 2015 at 617 million [115]. Predictions for the next 10 years estimate an additional 236 million people aged 65 and older, further accelerating the worldwide phenomenon of older populations [115].

According to findings and evidences, frailty is not an irreversible process [299, 185]. Frail elderly who are receiving medical care have been shown to have less cognitive or functional decline, possess lower mortality rates and experience fewer falls [40]. Moreover, exercise programs and increased activity have been shown to reduce adverse health outcomes and to delay the progression of frailty [232, 208, 153]. Other potential treatments for frailty include caloric and protein support, vitamin D and reduction of polypharmacy [185].

Hence, the early diagnosis of frailty may reduce its severity. This does not just benefit the individuals, but also relieve the burden for their families and the society [40]. Moreover, dropping the need for healthcare services reduces iatrogenic disability [152], which is the effect of reduced independence after hospitalization.

The most comprehensively researched model to determine the capabilities of frail elderly persons is called comprehensive geriatric assessment (CGA) [289]. It is an interdisciplinary diagnostic process including medical, psychological, social and environmental components. However, it is a resource intensive process and therefore complex and expensive. New methods of frailty assessments are required to find equally reliable but more efficient ways for routine care [59].

1.1.2 Frailty: Definitions, models, detection and controversies

There is no agreement about the definition of frailty, but it is widely recognized as a geriatric syndrome, that leads to functional decline among older adults [279].

Definition Many attempts have been made to describe frailty, e.g. the following: Frailty is a condition or syndrome which results from a multi-system reduction in reserve capacity [43]. Frailty is considered as the failure to integrate the complex responses required to maintain function [179]. Frailty is a state that is multifactorial and that implies vulnerability [225]. Frailty is the consequence of accumulated age-related defects in different physiological systems [299]. Frailty is a state of increased vulnerability to poor resolution of homoeostasis [40].

The following definition of Schuurmans et al. [240] is rated the highest by several experts [101]: Frailty is a loss of resources in several domains of functioning, which leads to a declining reserve capacity for dealing with stressors.

Frailty models Two principal models of frailty are considered by Clegg et al. [59]: The first and most frequently used model [37] is the phenotype model of Fried et al. [86]. It describes frailty based on five characteristics: Unintentional weight loss, weakness, poor endurance, slowness and low physical activity level. Having one or two characteristics defines intermediate subjects. Those with three or more characteristics are considered frail. The second model is the Frailty Index (FI) [227]. It defines frailty based on a cumulative deficit model. FI is computed as the equally weighted presence or absence of 92 parameters, e.g. diseases, disabilities, symptoms, signs or abnormal laboratory values. While FI has the advantage that the status of frailty is gradable, the sheer number of parameters require an extensive and elaborate assessment. Therefore both the phenotype model and FI may also be considered complementary, as the former is suited for an immediate frailty identification and the latter is suited for a comprehensive geriatric assessment [47].

Frailty instrumentation A large number of frailty measuring procedures have been proposed [67] e.g. standardized questionnaires, physical assessments or single measurements of gait speed or grip strength [59]. Popular questionnaires include the Groningen Frailty Indicator (GFI) [254], the Tilburg Frailty Indicator (TFI) [103] or the FRAIL scale [275]. The main advantage of questionnaires is that they are easy to administer (GFI, TFI and FRAIL scale consist of 15, 25 and 15 questions, respectively) and do not

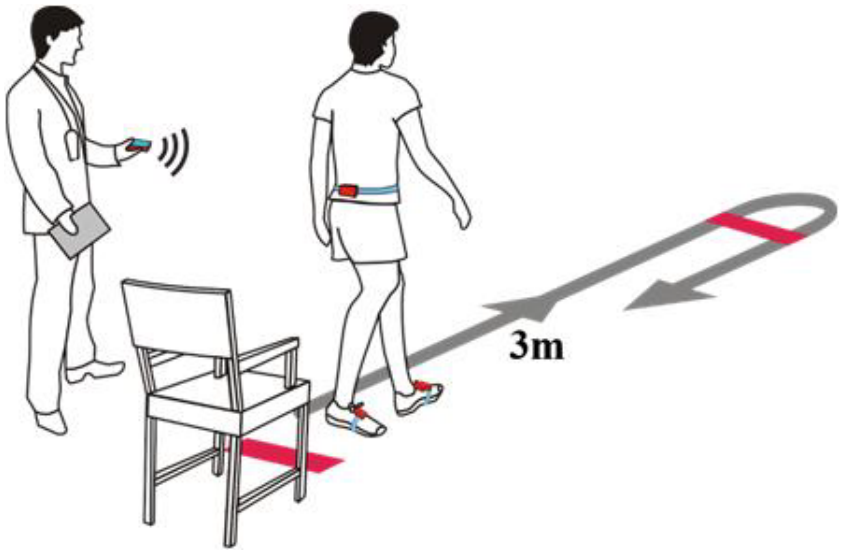


Figure 1.1: Illustration of the Timed Up & Go (TUG) test [287].

require additional equipment. The biggest disadvantage of questionnaires is that they depend on the personal valuation of the subject and therefore lack objective validity.

An alternative approach for measuring frailty is through physical assessments. A well-known and simple test for assessing mobility is the Get-Up-and-GO test [171]. It consists of a subject standing up, walking three meters, turning around and returning to sit back down on the chair (see Figure 1.1). A clinical expert observes the subject and rates the mobility on a scale from one to five.

Due to its subjective and imprecise scoring system, a modified version was proposed, the Timed Up & Go (TUG) test [215]. The score in the TUG test is computed based on the time taken for completing the test. There is no established norm for the TUG time cutoff for identifying frailty. Rockwood et al. [226] choose a 19-second cutoff based on the slowest 20% of TUG times from participants of the CSHA study [227], one of the largest population-based studies to measure the TUG. Other cutoff values in the literature are 20 seconds (for the Edmonton frail scale [228]) and 30 seconds [113]. The TUG test is widely considered to be a good indicator for frailty and future falls [246]. For example, having a high duration for the TUG test completion is most closely related (together with having any functional dependence) to needing institutionalized care after an operation [223]. However, its ability to identify prefail individuals and female individuals with a high fall risk is still controversial [236, 269]. Moreover, there seems to be no advantage in using TUG over gait speed when it comes to identifying frailty [236]. Furthermore, TUG is not able to assess the fatigue effect due to its short duration [177].

Besides TUG, there are a large number of other physical assessments. Short Physical Performance Battery (SPPB) [109] consists of a series of seven balance, gait speed and chair stand tests. Berg Balance Scale (BBS) [27] consists of a series of 14 balance

tests which are performed consecutively. Another functional evaluation test is the 30-s chair-stand (CST) test [131]. It consists of counting the number of sit-stand-sit cycles an individual is able to complete within 30 seconds of the test. More information on the various geriatric assessments can be found in the work of Vanswearing and Brach [276]

An advantage of physical assessments is that in a clinical setting a performance test is considered more relevant than a self reported questionnaire because it informs the clinician about actual functioning. [67]. A number of frailty measures also combine both questionnaires and physical assessments, e.g. the Edmonton Frail Scale (EFS) [228], which requires the patient to answer 10 questions and perform a TUG test.

Frailty controversies Frailty measurements are currently mainly used in research settings and have little use in medical practice [202]. Frailty screening remains controversial [40], since there is no consensus about a frailty definition [185] and no gold standard to evaluate the accuracy of frailty grading between different approaches.

Researches agree on the inclusion of aspects regarding strength, balance, nutrition, endurance, mobility, physical activity and cognition, but differ on adding components from the psychological or social domain [100]. There is a correlation between physical frailty and psychological problems as the cited characteristics (phenotype model) weight loss, weakness, poor endurance, slowness and low physical activity level also apply e.g. to depression. Therefore components from the psychological or social domain are included in commonly used frailty measure questionnaires, e.g. EFS, TIF and GFI. But because of these components, the majority of proposed frailty measures are considered subjective or partially subjective as they contain self-reported items [37]. A future definition of frailty cannot label a person frail simply because of self-perceived state of exhaustion, mood problems, sadness, loneliness, etc.

Self-reported health is both biased [274] and culture-dependent [134]. While it has been shown that subjective measures such as questionnaires are able to identify frail individuals, e.g. [207, 102, 184], it is currently unknown how well they perform due to a lack of a gold standard.

More evaluations are necessary in order to quantitatively assess not just how reliable the existing approaches are but also how reliable and predictive their individual components are. Van Campen states that frailty is an accumulation of not only physical but also psychological and social deficits [273]. But a vision of frailty as a mental issue would be a major shift in its concept since it is common sense that frailty is predominantly a physical deficiency. It affects not only depressed, but also psychologically healthy and optimistic personalities. Therefore, social and psychological components should be reconsidered in favor of objective components like a patient's mobility or strength.

Hence there is a need for objective and automatized assessments of frailty. A number of objective approaches have been proposed to assess frailty. One approach is to measure a subject's grip strength based on a grip-ball or Jamar dynamometer [280]. Another approach is the measurement of physical activity through wearable devices [191] or

presence sensors [237]. In this work, frailty is objectively and automatically assessed through the analysis of human gait and the analysis of the TUG test.

1.2 Detection of frailty using Kinect

Kinect and other depth sensors such as Asus Xtion, Leap Motion Controller, estimate the distance of objects in a given scene for each pixel. The first version of Kinect, the Kinect v1, is officially called Kinect for Xbox 360 and was released in 2010. It is a depth sensor based on the principle of structured light. It consists of an infrared projector and an infrared CMOS sensor. The former is used to project an irregular pattern of dots onto the scene. Depth is measured based on triangulation [309]. The more recent Kinect v2, officially referred to as Kinect for Xbox One and released in 2014, is based on the time-of-flight principle: Distances are obtained by measuring the phase of transmitted and received modulated light [243, 107].

Due to low cost and simplicity of setting up the sensor it shows great potential for various medical applications, e.g. physical therapy or Ambient Assistent Living (AAL) applications. Potential applications most relevant for this work are monitoring of human gait trough the extraction of spatiotemporal gait parameters and automatic analysis of physical assessments.

Kinect offers the advantage of unobtrusively measuring a given scene without the need for body-worn sensors or additional modifications of the environment. Moreover, Kinect provides detailed scene information which enables a deeper analysis of gait and locomotion compared to other unobtrusive sensors such as ambient presence sensors. Furthermore, the privacy concerns of older adults to video-based monitoring systems are alleviated when the monitoring is only based on silhouettes [69].

1.2.1 Gait Introduction

Two main phases in the gait cycle are considered: Stance phase, where the reference foot is on the ground and swing phase, where the reference foot swings in the air to prepare for the next footstep. Around 60% of the gait cycle is occupied by stance phase and the remaining 40% is occupied by swing phase. The gait cycle has traditionally been further divided into eight events. Five of them occur during stance phase [277].

1. *Heel strike* occurs when the heel of the swinging foot hits the ground. It is considered as the beginning of a gait cycle. The body's center of mass reaches its lowest point during heel strike.
2. *Foot-flat* is reached when the plantar surface of the foot touches the ground and the weight is transferred on the reference foot. It is also referred to as loading response phase [167].
3. *Midstance* occurs when the swinging foot swings past the standing reference foot. The body's center of mass reaches its highest point during midstance.

4. *Heel-off* occurs when the heel is raised from the ground.

5. *Toe-off* is the phase when the foot leaves the ground and marks the end of the stance phase.

The phases of the gait cycle are illustrated in Figure 1.2. A large number of spatial and temporal parameters can be computed from the gait cycle, e.g. gait speed, step length, step time, cadence, etc. They are summarized as spatiotemporal gait parameters. A longer list of parameters can be found in the work of Muro et al. [189]. They list 24 gait parameters that are relevant for medical applications, sports analysis and recognition purposes.

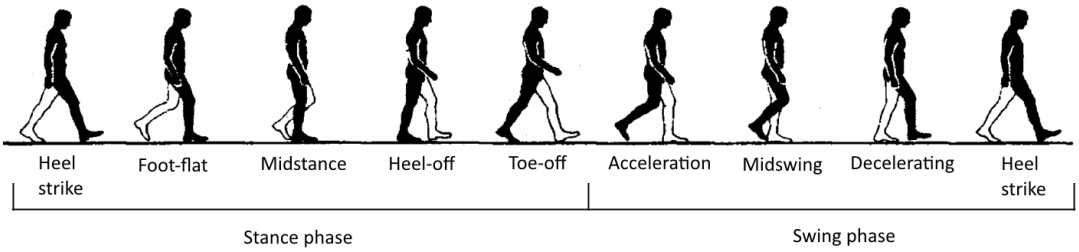


Figure 1.2: Gait cycle phases [68, 302].

1.2.2 Relation between frailty and spatiotemporal gait parameters

Spatiotemporal gait parameters have been linked to older adults, frail adults and adults with a history of falls in gait-related studies. The relation of frailty with a subset of spatiotemporal gait parameter is reviewed as follows.

Gait speed

Gait speed is reported as one of the most useful criteria for the identification of physical frailty [98, 230]. Furthermore, gait speed is considered as a robust predictor of future falls and mobility impairments, hospitalizations and death [48, 183]. Moreover, the assessment of gait speed alone is sufficient to predict adverse events in older person's lives [183]. This is further confirmed by Guralnik et al. [108], who report that gait speed alone performs almost as well in predicting incident disability than a full SPPB.

Vanswearing and Brach [276] compare 16 measures of physical function including gait speed regarding their appropriateness for older adults, their practical aspect and psychometric properties. They point out that the continuous nature of the measurement of gait speed allows to recognize minimal changes, which can be an early sign of physical decline. The authors also prefer gait speed over the TUG test as there exist extensive amount of comparative data for gait speed which help to determine its clinical meaningfulness.

Cadence

The data from Winter et al. [292], Elble et al. [78] and Kressig et al. [149] show almost no difference in cadence between young, healthy older adults and older adults transitioning to frailty.

Montero-Odasso et al. [182] report a smaller value in cadence for frail individuals compared to non-frail individuals (118.3 steps per minute for nonfrail, 101.2 steps per minute for frail subjects). Considering that Elble et al. [78] report a similar value of cadence (104 steps per minute) for healthy older adults as Montero-Odasso et al. [182] for frail adults, cadence seems unrelated to frailty based on the subset of reviewed studies.

Step length and stride length

Wolfson et al. [293] compare gait parameters of older adults with a history of falls and older adults with no fall history. They report a significant reduction in stride length for the fallers (0.53 m) compared to the non-fallers (0.82 m). Significant reductions in stride length for frail individuals are also confirmed by Maki [168], Montero-Odasso et al. [182] and Woo et al. [294]. However, it can easily be inferred that gait speed and stride length are highly correlated as cadence changes only slightly between non-frail and frail individuals.

Gait Variability

While gait velocity received the most research attention among the frailty related gait parameters, there is growing evidence that frailty is also linked with gait variability.

Hausdorff et al. [114] observe that increased gait variability is associated with future falls among community-living older people. The authors also mention that stride time variability shows a strong correlation with other measures such as strength, balance, gait speed, functional status and mental health, but these other measures do not distinguish future fallers from non-fallers.

Montero-Odasso et al. [182] report that the differentiation between three frailty status groups (frail, prefrail and non-frail) on gait variability is more evident under fast pace walking compared to usual walking. They also postulate that the increase in gait variability in older adults with frailty reflects the inconsistency of the neuromuscular system to maintain a steady gait. They conclude that future research is necessary to determine if gait variability is useful as a measure of frailty in older individuals.

Barak et al. [18] compare the walking pattern of elderly people who experienced a fall recently (”fallers”) with elderly people with no history of falls (”nonfallers”). The fallers show less stable gait patterns and greater gait variability compared to the nonfallers. The authors conclude that increased variability in gait may be an important gait risk factor regarding falls.

Gait parameter	normative values 70+ [120]	non-frail [182]	transitioning to frailty [149]	prefrail [182]	frail [182]	fearful fallers [168]
Gait speed (cm/s)	110.2 ± 17.3	124.2 ± 13.0	97 ± 23	95.2 ± 20.7	79.5 ± 19.4	66 ± 19
Cadence (steps/min)	108.8 ± 11.4	118.3 ± 6.7	105.7 ± 12.7	106.3 ± 9.1	101.2 ± 21.1	
Step length (cm)	60.7 ± 7.8		63.7 ± 6.5	55.9 ± 7.7	50.7 ± 7.7	
Stride length (cm)	122.1 ± 15.5	127.6 ± 13.9	111 ± 11	108.8 ± 18.3	98.7 ± 16.3	83 ± 16
Stride time (s)	1.10 ± 0.11	1.02 ± 0.06		1.14 ± 0.11	1.21 ± 0.14	
Double support time (%)	31 ± 6	28.4 ± 3	32.1 ± 5.8	31.7 ± 4.7	34.3 ± 4.8	19.8 ± 5.5
Step length variability (%CV)	5.8 ± 5.3	5.0 ± 1.5		6.6 ± 2.9	6.7 ± 1.7	
Stride length variability (%CV)	4.3 ± 3.9	4.0 ± 1.5		5.1 ± 2.8	5.7 ± 2.2	
Stride time variability (%CV)	4.4 ± 4.2	2.3 ± 1.1		2.99 ± 1.4	3.8 ± 2.0	

Table 1.1: Comparison of mean and standard deviation of gait parameters in older adults based on normative values and values for non-frail, pre-frail and frail subjects within the literature. Data for normative values is taken from Hollman et al. [120] and averaged over all age and gender groups based on their population count.

Comparison of non-frail and frail adults

Table 1.1 shows mean and standard deviation for several gait parameters from different studies analyzing gait of older adults at usual walking speed. In the selected studies, the studied subset of adults ranges from normative and non-frail older adults to frail adults and adults with a history of falls. Based on this data, the most obvious change with more progressed frailty is the reduced gait speed, step length and stride length. Moreover, an increased stride time, a slightly reduced cadence and an increased variability for step length and stride length can be observed. Double support time and stride time variability do not distinguish the different groups of older adults within this subset of data. More information on studies regarding the relation of spatiotemporal gait parameters and frailty can be found in the survey of Schwenk et al. [242].

1.2.3 Contribution of this thesis

In this work, frailty detection approaches based on human gait analysis and TUG test analysis are implemented and evaluated. Both approaches utilize the Kinect v2 depth sensor.

The human gait analysis approach detects frailty based on five spatiotemporal gait

parameters. These parameters are gait speed, step length, stride length, step time and stride time. Two gait datasets are acquired as a part of this work.

The first gait dataset consists of 18 walking sequences of 13 healthy adults for a total of 234 sequences. The 18 walking sequences consist of back and forth walking using three different step sizes and three different walking paths. Spatial ground truth is obtained based on markers attached to the floor. Temporal ground truth is obtained by manually labeling heel-strike and toe-off events for all walking sequences. Extracted spatiotemporal gait parameters are evaluated using the mean and standard deviation and the mean absolute error for each parameter. Moreover, the effect of walking path and step size and the effect of walking towards or away from the sensor are evaluated for the step length parameter.

The second gait dataset is acquired from 11 elderly volunteers aged 85-95 who live in a nursing facility. Each participant walks three meters back and forth for a total of 22 walking sequences. The number of steps during each walking sequence was manually labeled. The proposed gait analysis approach is evaluated on the elderly gait dataset based on the number of steps per walking sequence.

The TUG analysis approach detects frailty by automatically analyzing the TUG test. This is achieved based on a computer vision approach that analyzes the movement of the subject based on the velocity of the centroid or spine shoulder joint. Currently the TUG test requires a medical assistant to use a stopwatch to measure the time a subject needs to complete the test. With the proposed approach, the TUG time is measured automatically and no interaction from an assistant is required. Moreover, additional parameters are measured from the completion of the TUG test, e.g. time taken for sit-to-stand, walking, turning, stand-to-sit, etc. In order to evaluate the proposed approach, a real-world dataset of TUG recordings from 11 elderly participants is acquired and manually labeled. The set of extracted TUG parameters is evaluated on the acquired dataset based on the manual video annotation. Signed difference, absolute mean error and standard deviation as well as precision and recall are used as evaluation metrics.

Furthermore, the potential frailty status of the elderly volunteers is evaluated based on three different data sources: Spatiotemporal gait parameters from their walking patterns, TUG score from their TUG assessments and an additional score from a frailty questionnaire that has been repeatedly validated within the literature. While the acquired dataset lacks the required quantity to make statistically meaningful conclusions, it shows real-world examples of how each approach would classify the frailty status of older adults.

1.2.4 Organization

This thesis is organized as follows: Chapter 2 describes the state-of-the-art in several areas related to human gait and depth sensors.

The methodology for the gait analysis and TUG automatization approaches are discussed in chapter 3. First the silhouette extraction algorithm based on depth data is described

and for each choice, alternatives from related works are discussed. Then the extraction of footfalls and the analysis of various gait signals from both depth data and skeleton data are specified. The last part of the chapter covers the proposed TUG automatization approach.

Chapter 4 presents the results of this work. First the accuracy of extracted spatiotemporal parameters using different gait signals is compared and evaluated with a manually labeled ground truth. Second, the approach is evaluated on real-world gait data from 85 to 95 year old. Necessary adjustments are suggested and discussed based on the walk of the elderly participants. Third, the two proposed TUG automatization methods are evaluated based on TUG recordings from the elderly participants and compared against a third state-of-the-art method.

Chapter 5 concludes the thesis. The most significant observations made during the evaluation are discussed and an outlook for future studies is presented in the final chapter.

CHAPTER 2

Related work

This chapter presents the state-of-the-art regarding human motion analysis, gait recognition and Kinect-related applications. The related works are structured into the following four subsections:

1. **Sensor Modality:** The range of sensors that are used to gather human motion information is outlined and their advantages and disadvantages are discussed. The remaining subsections focus on the computer vision approach.

2. **Kinect specification and viability:** Available low-cost depth sensors are discussed and compared. After that, Kinect is discussed from a hardware perspective and a software perspective. From a hardware perspective the depth measurement error is reviewed based on previous experiments. From a software perspective, studies which extract spatiotemporal or kinematic parameters from Kinect are reviewed and compared. A conclusion is made at the end of the subsection whether Kinect is suitable for this work.

3. **Gait Recognition and gait analysis:** First the development of gait recognition during the last two decades are summarized and previous and current taxonomies are discussed. Then gait recognition and gait analysis based on Kinect are reviewed. Moreover, relations and differences between gait recognition, general medical gait analysis and gait analysis for frailty detection are discussed.

4. **TUG Automatization:** Existing approaches for Kinect-based analysis and automatization of the TUG test are reviewed.

2.1 Sensor modality for Human Gait Analysis

2.1.1 Motion Analysis Taxonomy

A common classification of motion analysis techniques distinguishes three categories: Image processing or machine vision, floor sensors and wearable sensors [189, 89, 71]. Approaches based on machine vision are markerless approaches based on either color or depth data. Floor sensors include all types of floor-based sensors, e.g. force platforms, baropodometric mats, walking mats, etc. Wearable sensors include all body-worn sensors and devices, e.g. pressure and force sensors, inertial sensors, active and passive markers, electromyography, etc.

Tao et al. [267] divide human gait analysis into kinetics, kinematics and electromyography. Gait kinetics studies the forces and moments that impel the movement of human limbs. Gait kinematics describes lower extremity movements of joints and segments independently of their mass and forces. Bontrager [68] reviews several gait instrumentation techniques used for rehabilitation. He distinguishes between picture video, temporal gait, foot pressure, motion, force and electromyography. Picture video is a simple recording viewed by a clinical expert. Temporal gait includes techniques that measure the temporal aspects, e.g. footswitches. Techniques based on pressure sensors, e.g. pressure mats and pressure insoles, are categorized as foot pressure. Similarily, force plates, force measuring sandals or force measuring walking aids belong to the force category. Techniques that measure the kinematic aspects of gait, e.g. the movements of the lower limbs, are categorized as motion. Wearable sensors, e.g. electrogoniometers, and marker-based motion capture systems are classified as motion techniques.

Frenken [84] groups gait analysis approaches based on their sensor modality. Similar to Tao et al. [267], sensors are grouped into kinetic and kinematic sensors and additionally, into body-worn and ambient sensors. Body-worn sensors are used for EMG, accelerometers or marker-based solutions. Ambient sensors are generally unobtrusive and may not even require active participation. Markerless camera-based gait analysis falls into the ambient and kinematic categories. Table 2.1 shows the taxonomy of Frenken.

Best and Begg [30] distinguish between kinematics/motion, Kinetics/force, pressure mats/pressure insoles and electromyography. Their kinematics/motion category contains footswitches, gait mats, marker-based and magnetic motion analysis systems and wearable devices.

Surer and Kose [265] review several gait analysis technologies in a clinical context. Marker-based and markerless systems, inertial measurements, force platforms and electromyography are evaluated regarding applications and limitations. The authors describe several markerless approaches based on color cameras and point out their advantages of not requiring intrusive markers or special hardware.

It should be noted that taxonomies made from a medical and practical perspective generally do not consider markerless vision-based gait recognition, e.g. the works of Best and Begg [30] or Bontrager [68], while taxonomies which use a scientific or technical

	Body-worn	Ambient
Kinetic	1. **Pressure and force sensors** in shoes	1. **Pressure and force sensors** on the ground in treadmills in furniture in walkers
Kinematic	1. **Time of flight** ultrasound 2. **Visual** marker based 3. **Electrical impulse** electromyography 4. **Inertial forces** e.g. accelerometers, gyroscopes body-worn in clothing 5. **Bending forces** electro-goniometer	1. **Time of flight** RADAR LIDAR 2. **Visual** marker less fluoroscopic 3. **Presence sensors** home automation RFID

Table 2.1: Classification of gait analysis approaches based on their sensor modality. [84]

perspective put it as one of their major categories [189]. This suggests that markerless gait recognition, while being an important research topic, still has catching-up to do to close the gap with existing gait analysis methods. However, current motion analysis technologies, while being accurate and reliable, are usually not non-intrusive, require special interaction from the subject or they require modifications of the environment. Therefore there is great potential in a vision-based motion analysis solution, as none of the currently deployed technologies offer an approach that may run unobtrusively and independently in the background, while providing the same quality of information.

2.1.2 Wearable inertial sensors

Gait analysis systems based on wearable sensors attach various motions sensors to the subject's body, e.g. accelerometers (measures acceleration along sensitive axis), gyroscopes (measures orientation from angular velocity), force sensors, strain gauges (measures stretching), inclinometers (measures angle of slopes) or goniometers (measures

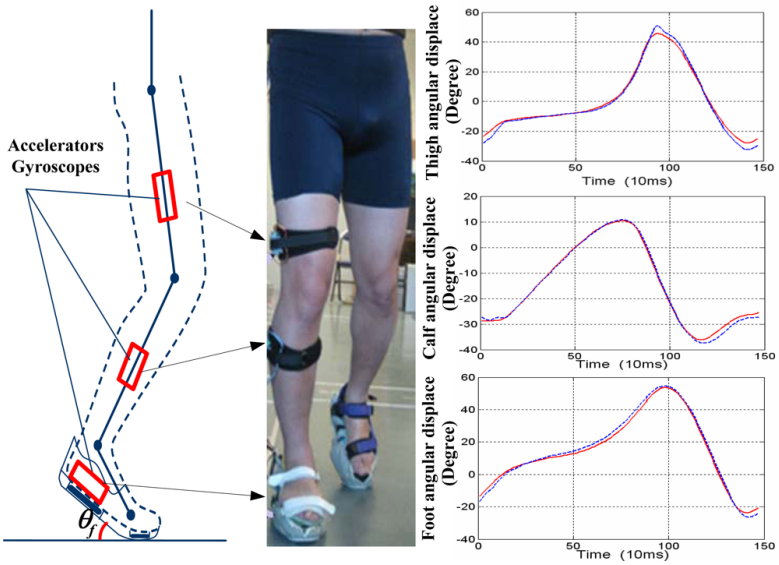


Figure 2.1: Thigh, calf and foot angles obtained from a set of wearable accelerators and gyroscopes [267].

angular changes) [267]. Figure 2.1 illustrates a set of inertial sensors attached to a person's lower limbs and their corresponding signals.

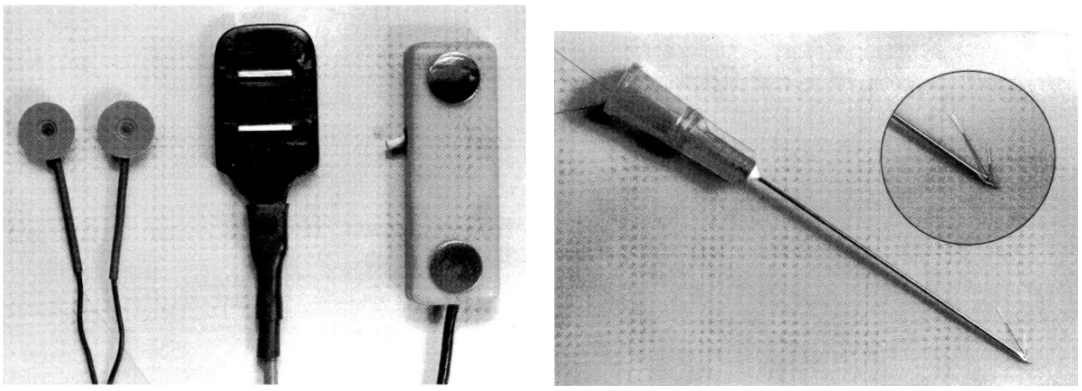
Wearable inertial sensors suffer from short battery life [75] and the need to be worn by the subject, which makes them unsuitable for long-term monitoring [282]. Furthermore, it has been shown that many older adults consider wearable devices to be invasive and inconvenient [70]. Attempts of improving the usability have been made, e.g. by using an accelerometer-enabled smartphone [266, 195], a smartwatch [130] or by placing the sensors in the subject's shoe [308].

2.1.3 Electromyography

Electromyography (EMG) measures electrical impulses that are used for muscle contraction. Muscle activity is detected either from electrodes that are attached to the skin (surface EMG, see Figure 2.2a) or with needle electrolytes that is inserted into the muscle (intramuscular EMG, see Figure 2.2b) [68, 174]. Both types are only suitable for a laboratory setting [50]. Figure 2.3 shows an example of applying EMG to observe lower limb muscle activity during pedaling.

2.1.4 Optical motion capture

Marker-based motion capture systems attach wearable markers onto the subject to facilitate identification and localization of body parts. Markers can be active, e.g. light emitting diodes (LEDs), or passive, e.g. retroreflective markers or fluoroscopic markers [13]. These markers are detected by a system of circularly arranged cameras



(a) (b)

Figure 2.2: Example of two sets of EMG utensils: (a) surface EMG. (b) intramuscular EMG [68].

(usually between 6 to 50 [39]), that are placed in a recording studio. In case of passive markers, infrared or visible lights are mounted beside the cameras and their light gets reflected back into the cameras by the retroreflective markers. The reflection of the light facilitates marker detection. In case of active markers, the light does not get reflected by the markers but is instead radiated by LEDs, which are attached to the markers [39]. An example for an active marker system is illustrated in Figure 2.4. Optical motion capture is widely used for computer graphics animations, physical therapy or athlete training optimization [4] and is already in use for decades [39]. Figure 2.4 illustrates an example of extracting a skeleton model based on body-worn active markers.

Commercial optical marker-based motion systems have to be discriminated from markerless approaches, as they avoid many computer vision problems by using a controlled environment and special hardware [99]. Examples for commercial marker-based motion capture systems are VICON [1] or Motion Analysis [2].

2.1.5 Electronic walkways

Electronic walkways or walking mats are embedded with pressure sensors that allow the detection of footfalls as a person walks over the walkway. Gait parameters such as walking speed, cadence, step length and step times are extracted from the footfalls using a compatible software. While electronic walkways are used for medical gait analysis [31, 182, 64, 278], they require an instructor and a clinical setting [148] and are costly [307]. Examples for commercial systems of electronic walkways are GAITrite [3],

[1] https://www.vicon.com/, Accessed 2017-05-25
[2] http://www.motionanalysis.com/, Accessed 2017-05-25
[3] http://www.gaitrite.com/, Accessed 2017-05-15

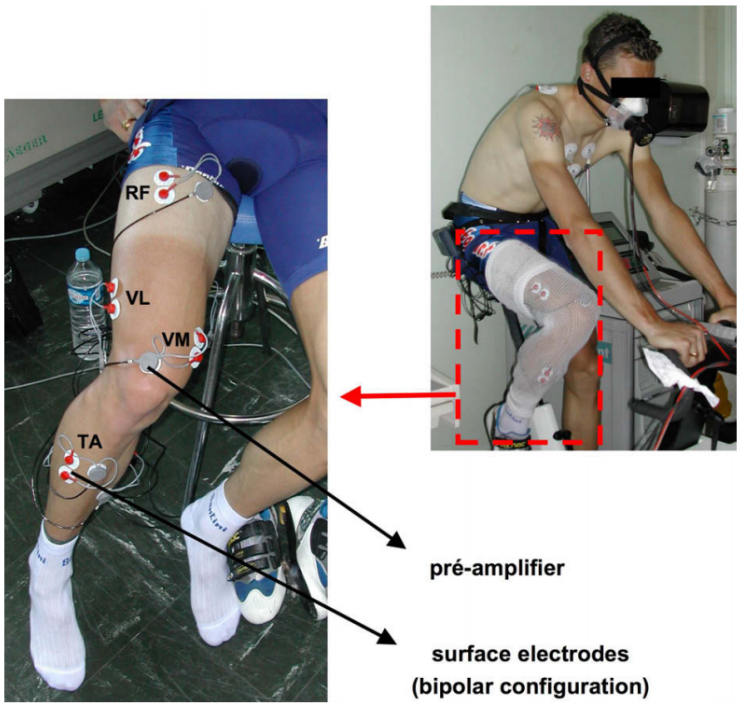


Figure 2.3: Example of a surface EMG applied on lower limbs during pedaling [122].

Tekscan Strideway System [4] or ProtoKinetics Zeno Walkway [5]. Figure 2.5 shows examples of commercial electronic walkways. Besides walkways, pressure sensors may also be mounted within various other objects, e.g. in shoes [16], walking aids [272] or furnitures [126].

2.1.6 Health Smart Homes

Home automation systems or health smart homes distribute various ambient senors throughout a person's home environment to monitor the person's activities, health, etc. Different sensors can be used for this purpose, e.g. pressure sensors, pressure mats, smart tiles, passive and active IR sensors, sound sensors, magnetic switches or optical and ultrasonic systems [237]. Except for a set of IR sensors and optical/ultrasonic systems, which allow estimation of gait speed, smart home sensors are only able to measure the current location and possibly the current activity of the subject. Figure 2.6 illustrates an example for a health smart home system that combines ambient and body-worn sensors to monitor the health and to track the activities of the subject.

Home automation systems provide potential unobtrusive long-term in-home monitoring

[4] https://www.tekscan.com/products-solutions/systems/strideway-system, Accessed 2017-07-15
[5] http://www.protokinetics.com/zeno-walkway/, Accessed 2017-05-15

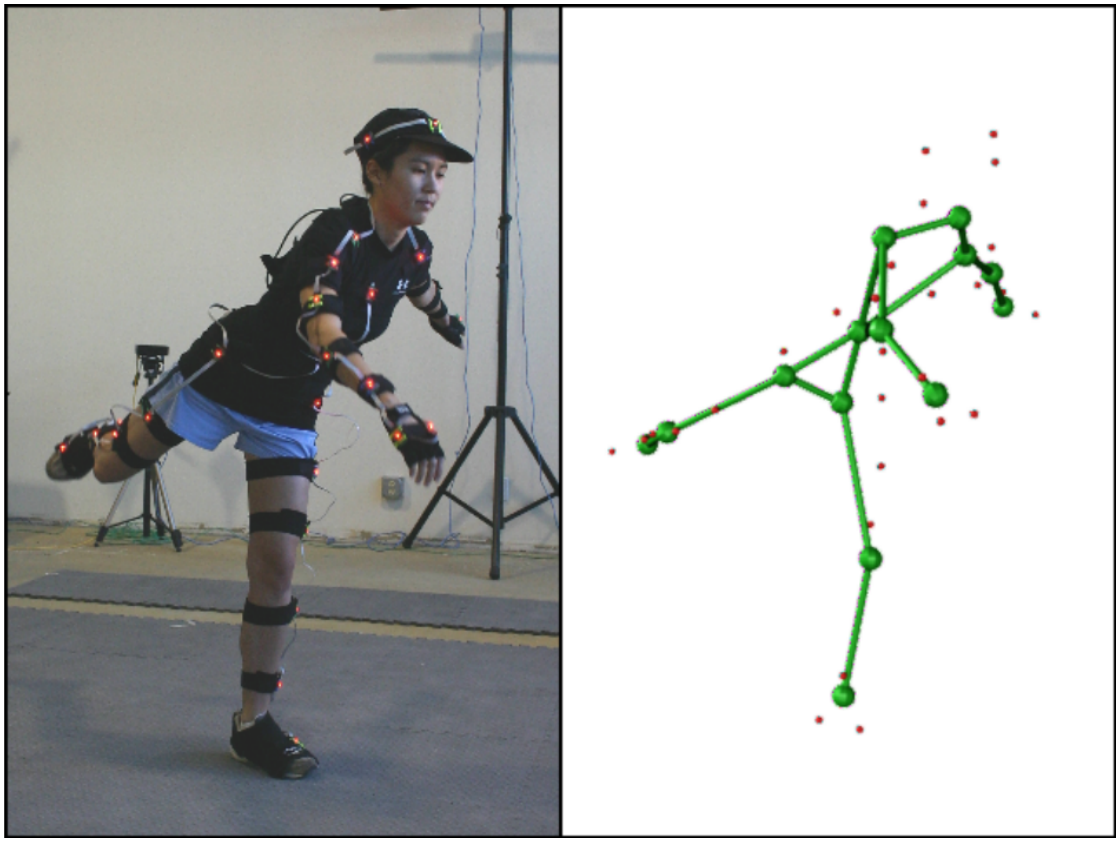


Figure 2.4: Illustration of an optical motion capture system based on active markers and the extracted corresponding skeletal structure [143].

of frail or prefrail persons. Kaye et al. [140] measure walking speed and physical activity (based on number of walks per day) from four passive infrared motion sensors in a row. Pavel et al. [204, 203] use passive IR motion sensors, contact switches and active radio frequency identification (RFID) to infer gait velocity and its variability. While physical activity and walking speed are well-suited to estimate frailty, other gait-related changes cannot be identified with ambient smart home sensors. Moreover, a large number of sensors is required to to track the movement and the activities of the subject.

2.2 Kinect specification and viability

In this work the Kinect v2 sensor is utilized for the mobility analysis approach. In order to verify this choice, depth sensing technologies and a number of low-cost depth sensors are reviewed. Moreover, Kinect-based studies that analyze the depth measurement error and studies that analyze the error on extracted mobility parameters are discussed. The latter are discussed in more detail, as these studies resemble the gait analysis experiments in this work.

(a) (b) (c)

Figure 2.5: Three examples of commercial walkway systems: (a) GAITrite. (b) Tekscan Strideway System. (c) ProtoKinetics Zeno Walkway.

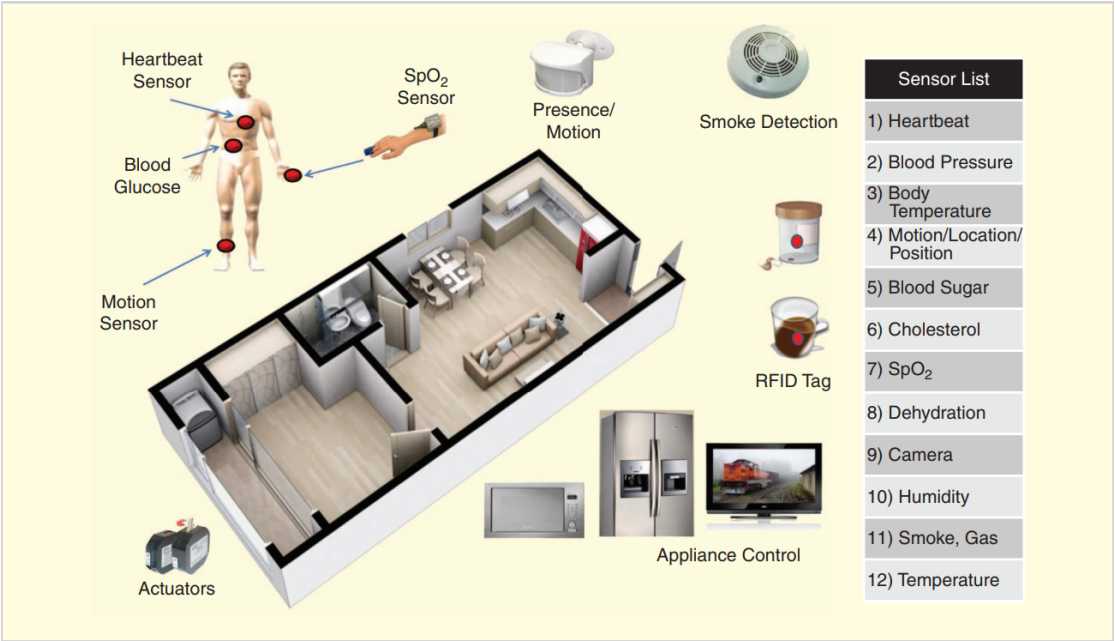


Figure 2.6: U-Health smart home system, which consists of a set of ambient sensors, body-worn sensors and actuators. [5].

2.2.1 Depth sensing technologies

Three main depth sensing technologies are found in computer vision research: Stereo cameras, structured light and time-of-flight (ToF) cameras [52]:

Stereo cameras: Stereo vision reconstructs a scene based on a calibrated binocular camera system [231] (see Figure 2.7a). It is inspired by human vision and its early research

dates back to the 1960s [52]. Depth information is extracted using the triangulation principle [123] and the binocular disparity [161], which is found after matching corresponding points in a pair of stereo pictures. Disadvantages of stereo vision are its inability to find matches in homogeneous object regions [147] and the high computational costs [123, 231] of solving the correspondence problem, which refers to the matching of corresponding pixels. Stereo vision is commonly used for gait recognition, e.g. [104, 125, 161].

Structured light: Structured light is the projection of a light pattern under geometric calibration on an object [189]. The idea of structured light is based on stereo vision: Instead of using two cameras, one camera is replaced by a light source that projects a known pattern, called structured light [52]. Depth information is extracted by analyzing the distortion of the observed pattern compared to the original projected pattern [234]. The working principle of structured light is shown in Figure 2.7c and 2.7d. Examples for depth sensors based on structured light are the original Kinect v1 or Asus Xtion [106]. An advantage of structured light devices over ToF devices is that they are cheaper to produce [52]. Disadvantages include problems with motion scenes, transparent and reflective surfaces and superposition of the light pattern with reflections [189].

Time-of-flight and LIDAR: Both LIDAR and TOF cameras measure the time difference between emission and arrival of a transmitted pulse and its reflection [144]. Since TOF cameras do not have an integrated mechanical scanner, they belong to a subclass of LIDAR called non-scanning or scanner less LIDAR [124].

There are two functional principles for the technology of ToF cameras [147]. One ToF technology makes use of modulated, incoherent light, and is based on phase measurement using standard CMOS or CCD technology (see Figure 2.7b), e.g. SwissRanger SR4000 [198] or Kinect v2. The other employs optical shutter technology, e.g. Zcam [124, 301]. While ToF cameras have problems with reflective surfaces and generally have low resolutions, 3D acquisition is real-time and, unlike stereo vision and structured light, no depth calculation is required [189].

Rotating multi-beam Lidar systems (RMB-LiDAR) provide a 360° FoV of the scene. The vertical resolution depends on the number of the sensors and the horizontal angle resolution depends on the rotation speed [24]. Both 2D LiDAR laser range scanner [199, 268] and RMB-LiDAR [90, 23] are used for gait analysis.

It should be noted that there are several other camera based depth sensor technologies available, e.g. depth-from-texture, depth-from-focus, depth-from-shading or depth-from-motion [263]. While not measuring depth, infrared thermography, which creates visual images based on surface temperatures, may also be used to extract silhouettes and gait information [300].

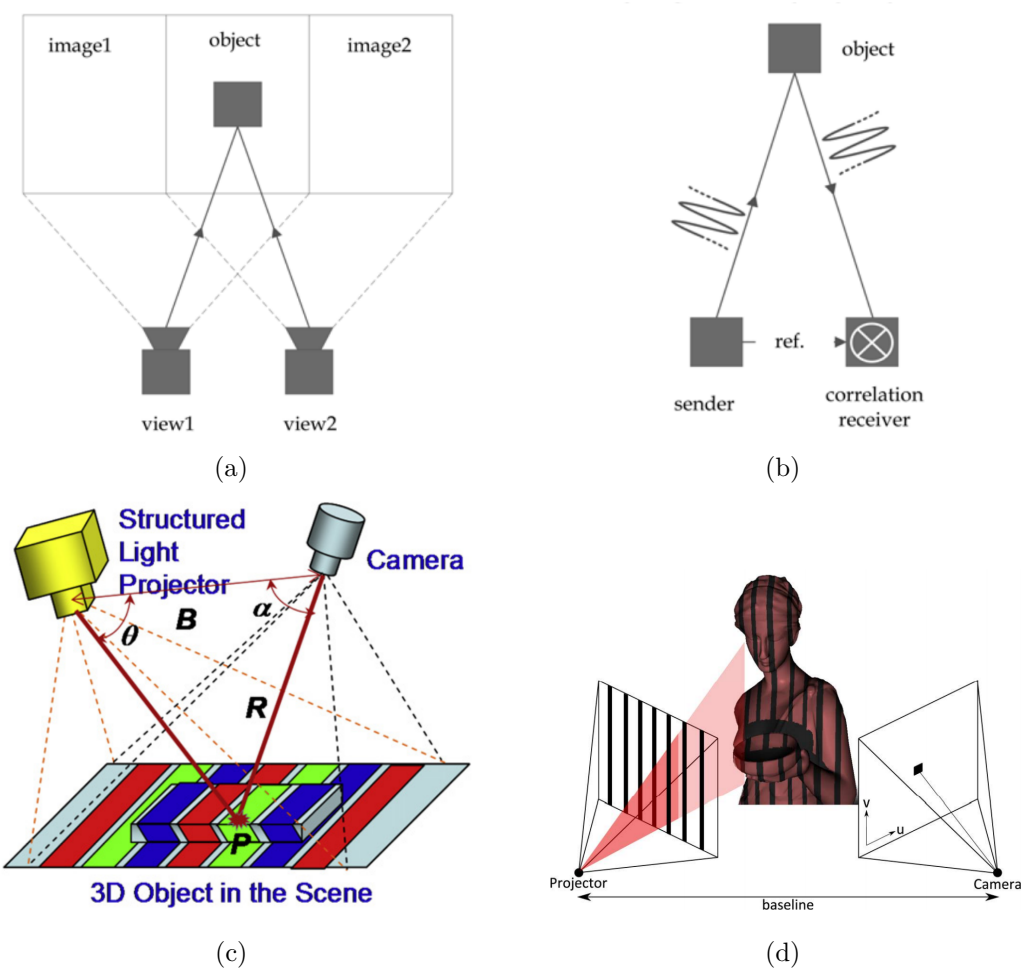


Figure 2.7: Working principles of three 3D sensing technologies: (a) Stereo vision [123]. (b) Time-of-flight [123]. (c-d) Structured light [94, 234].

2.2.2 Comparison of low-cost depth sensors

Mousavi and Khademi [186] compare four depth sensors (Leap Motion Controller [6], Asus Xtion [7], Creative Senz3D [8] and Kinect v1). They report that the Leap Motion Controller should be used for hand tracking with individual fingers, the Creative Senz3D for close range tracking and Kinect v1 or Asus Xtion for tracking of full body motion. Similar to the Creative Senz3D, the SoftKinetic DepthSense camera [9] is most suited for close range tracking. The DepthSense 325 sensor is limited to a range of 1.5 meter [60]. While the DepthSense 525 sensor driver offers extra modes for higher range, those modes run with

[6] https://www.leapmotion.com/product/desktop, Accessed 2017-06-22
[7] https://www.asus.com/3D-Sensor/Xtion_PRO/, Accessed 2017-06-22
[8] https://us.creative.com/p/web-cameras/creative-senz3d, Accessed 2017-06-22
[9] https://www.softkinetic.com/Products/DepthSenseCameras, Accessed 2017-06-22

(a) (b) (c)

(d) (e) (f)

Figure 2.8: Low-cost depth sensing devices: (a) SoftKinetic DepthSense 325. (b) Asus Xtion Pro. (c) Leap Motion Controller. (d) Creative Senz3D. (e) Microsoft Kinect v1. (f) Microsoft Kinect v2.

low frames per second (2.5 meters at 15fps and 4 meters at 6fps [9]). The discussed depth sensors are depicted in Figure 2.8.

Gonzalez-Jorge et al. [106] perform accuracy and precision tests to compare the Kinect v1 with ASUS Xtion. They use a standard artifact based on 5 spheres and 7 cubes. Both sensors show a decrease in accuracy with larger range, fitting to a second order polynomial. Outdoor measurements with sunlight are not possible for either sensor. Both sensors show metrical potential for low-range indoor applications that do not have very high accuracy requirements. Later, Gonzalez-Jorge et al. [107] compare Kinect v1 with Kinect v2 using the same artifact. Accuracy is superior for the Kinect v2 at both 1 m and 2 m range (-12 mm at 1 m range and -25 mm at 2 m range for Kinect v1, always $<$ -5 mm for Kinect v2). Precision is similar at 1 m range for both sensors and better at 2 m range for the Kinect v2. Measurement range is 4 m for the Kinect v2 and 6 m for the Kinect v1. Amon and Fuhrmann [7] compare the face tracking ability of the Kinect v1 and the Kinect v2 with their respective SDKs. They reported better performance in nearly all tested measurements for Kinect v2.

Specifications for Asus Xtion, Kinect v1 and Kinect v2 are shown in Table 2.2. While all three have sensors for RGB, depth and a microphone, the Kinect 2 sensor offers the largest field of views and the largest resolutions for both color and depth stream.

Summary of depth sensor comparison

Kinect v2 shows better accuracy than its predecessor [107], Kinect v1, while Kinect v1 and Asus Xtion show similar performance [106]. Advantages of Kinect v1 over Kinect v2

Depth Camera	Sensors	Depth Field of View	Depth Resolution	Color resolution	FPS	Depth range
Asus Xtion	RGB, depth and microphone	58° H, 45° V	320×240	640×480	30	$0.8 - 3.5$ m
Kinect 1	RGB, depth and microphone	57.5° H, 43.5° V	320×240	640×480	30	$0.4 - 6$ m
Kinect 2	RGB, depth and microphone	71° H, 60° V	512×424	1920×1080	30	$0.5 - 4.5$ m

Table 2.2: Technical specifications of three depth cameras suited for tracking full body motion [249, 106, 107].

are the larger depth range and that it comes with a tilt motor offering $\pm 27°$ of vertical tilt. Other depth sensors, such as the Creative Senz3D and the Leap Motion Controller are not considered as they are not recommended for tracking of full body motion [186] and the SoftKinetic DepthSense is only suitable for close range [60].

2.2.3 Kinect Validation

In order to evaluate the performance of Kinect as a tool for mobility analysis a literature research was done. "Kinect" and "Accuracy" are used as keywords and Google scholar is used as a platform. Retrieved papers are deemed relevant if they focused on either studying the hardware capabilities or the software capabilities of Kinect. Papers with a software focus include all papers that use depth data or skeleton tracking of the Kinect SDK or a similar framework (OpenNI, OpenKinect, etc.) to extract mobility parameters. Papers with a hardware focus include all papers that analyze the performance of Kinect sensors based on the raw sensor streams.

Hardware perspective

Khoshelham and Elberink [141] analyze the accuracy and resolution of the Kinect v1 depth stream for indoor applications. They report that both the error of depth measurements and the depth resolution increase quadratically with distance. At the maximum range of 5 meters the error increases from a few millimeters to 40 millimeters and the resolution is as large as 70 millimeters per pixel. The authors suggest that data should be acquired within 1 to 3 meters for mapping applications.

Smisek et al. [249] examine the Kinect v1 regarding depth measurement resolution and error properties. Moreover, they compare the accuracy of the Kinect v1 with a stereo camera and a TOF camera when measuring planar targets. The stereo camera performs the best (1.6 mm mean error), the Kinect is slightly inferior (2.4 mm mean error) and the TOF camera performs the worst (27.6 mm mean error) among the three .

Stoyanov et al. [262] evaluate the accuracy of a Kinect v1 against two TOF cameras (SwissRanger SR-4000 and Fotonic B70) in an uncontrolled indoor environment. An accurate standard actuated laser range is used as a ground truth. Within 3.5 meters, the performance of the Kinect is comparable to the actuated laser sensor with the two TOF cameras performing slightly worse. In their experiments with high range scans the performance of the Kinect drops significantly behind the two TOF cameras. However, none of the three sensors are comparable with the actuated laser in the high range experiment.

Software perspective

Springer and Yogev [250] evaluate the literature regarding Kinect-based gait assessment. They review 12 studies, which compare spatiotemporal gait parameters and kinematic gait parameters obtained from a Kinect with a gold standard. They find good validity only for a few spatiotemporal parameters (gait speed, step length and stride length) and poor validity for all kinematic parameters. One of the limitations of their study is that they do not differ between studies that used a Kinect v1 from studies which used a Kinect v2. Furthermore, they remark that most studies are performed with healthy adults in laboratory settings and that methodology varies greatly within the studies. Thus they suggest that the methodological procedure needs to be standardized and further research in real-world settings, involving people with pathological gait, is required.

Mousavi and Khademi [186] review studies focusing on Kinect-based rehabilitation. They find out that Kinect is an acceptable tool for rehabilitation applications, but still has several limitations due to occlusion and noise. They suggest to apply Kalman filtering, sensor fusion and calibration to mitigate these issues.

Obdržálek et al. [197] compare the pose estimation (skeleton model) of the Kinect SDK with a marker-based motion capture system during the performance of six physical exercises. While performing the exercises, the subjects are either seated or standing beside the chair. In controlled body postures, e.g. during standing and exercising arms, the Kinect shows similar accuracy than the motion capture system. However, in general poses including sitting and partial occlusions from the chair, the pose estimation error was about 10 cm. However, they reported problems of the Kinect skeleton tracking with occluding body parts, non-distinguishing depths (e.g. arms close to the body) and other objects in the scene. Other problems are caused by the variable limb lengths, which led to situations with non-anthropometric kinematic models.

Bonnechere et al. [34] compare segment lengths and joint angles obtained with the Kinect v1 and with a wearable marker based system (Vicon). They use data from subjects performing a deep squatting motion. Their results show poor accuracy for the majority of segment lengths (especially arm, hand, trunk and thigh segments). The authors conclude that the Kinect can be used for medical games, but its usefulness for biomechanical applications is limited.

Clark et al. [57] evaluate the accuracy of the Kinect v1 against a marker-based system during three postural control tests. They report excellent concurrent validity and conclude that the Kinect has the potential for a wide range of clinical applications. More recently, Clark et al. [56] compare spatiotemporal gait parameters obtained from the skeletal model of the Kinect v1 sensor with a marker-based system (3DMA). Spatiotemporal parameters are computed from toe-off and ground contact events, which are detected based on a velocity threshold in the movement direction. Foot swing velocity is estimated as the peak velocity during swing phase. They report good agreement for gait speed, step length and stride length (correlation coefficient > 0.9), but poor overall agreement for swing velocity, step time and stride time. The authors conclude that the Kinect offers significant advantages for assessing most spatiotemporal gait parameters, but the parameters must be chosen carefully as some parameters cannot be accurately measured.

Weber et al. [288] investigate the tracking accuracy of Kinect v1 compared to an optical marker based system. While they report excellent accuracy for the human gait relevant angles, they also reported that Kinect does not provide reliable segment lengths. As a remedy, they suggest to apply the Levenberg-Marquardt algorithm to ensure that the measured segment lengths are constant to prevent variation in segment length over time.

Paolini et al. [200] validate their foot tracking method which is based on a color camera and a Kinect v1 sensor. A colored 3D patch is attached to the tracked foot and the color information is utilized to track the foot. For their evaluation experiment twelve healthy subjects walk on a treadmill under three different walking conditions. They report excellent correlation between the foot tracking of their method and a marker-based reference system.

Llorens et al. [162] compare the accuracy and jitter of marker-based optical, electromagnetic and skeleton tracking from a Kinect v1. Participants are recorded during the activities of a virtual reality rehabilitation system. The skeleton tracking performs second behind the marked-based optical system regarding accuracy and performs worst regarding jitter.

Chang et al. [50] evaluate the performance of Kinect v1 during motion tracking compared to an optical marker-based system. Two participants perform six different types of upper limb movement. They conclude that Kinect's accuracy is competitive with the marker-based system.

Yang et al. [305] investigate the reliability of standing balance assessment with Kinect v1. The authors find excellent test-retest reliability and good concurrent validity when compared to a marker-based motion analysis system.

Pfister et al. [209] compare a Kinect v1 with a VICON marker-based system regarding angular displacement, stride timing and tracking ability. For their experiment twenty subjects walk on a treadmill with three different velocities. They observe the highest accuracy of Kinect for the stride parameters and worse accuracy for kinematic knee and hip parameters. Variability is lower at the slowest gait velocity. Another observation is that the tracking shows problems when the knees are crossed. They reported that on

average, 8 to 18% of steps are missed by the Kinect v1 tracking. They summarize that the Kinect is an acceptable tool to measure stride times, but currently unacceptable for clinical measurements. However, it may be clinically acceptable with some improvements regarding hip and knee measurements.

Dutta [77] evaluate the usage potential of the Kinect v1 for posture and movement assessments in the workplace. He compares the accuracy of the relative positions of four cubes placed within the Kinect's field of view with an optical marker-based system (VICON). He reports root-mean-squared errors between 5.7 and 10.9 millimeters and concludes that the Kinect may be used for ergonomic assessments with further development.

Baldewijns et al. [14] calculate step length and step time from Kinect depth image sequences. They apply the player detection algorithm from the SDK to distinguish between pixels belonging to the person and to the background. Afterwards they apply connected component analysis and calculate the mean position of all pixels from the biggest foreground object to obtain the center of mass. The X- and Y-coordinates of the center of mass are then used to determine the step length and the step time. Evaluation with an electronic walkway (GAITRite) shows good accuracy for walking sequences and poor accuracy for single steps.

Stone and Skubic [255] compare the accuracy of walking speed, stride time and stride length between two Kinect v1 sensors and a web camera system. They report good accuracy with measurement errors ranging from 0.1% to 6.1%. Their results show better performance for the Kinect placed within walking direction (looking towards the subject) compared to the Kinect positioned perpendicular to the walking direction. The authors also use the same experimental setup to measure gait variabilities for stride length, stride time and stride velocity [259]. The web camera system is more accurate for stride length variability, but the Kinects show superior accuracy for detecting stride time variability and stride velocity variability.

Regazzoni et al. [222] compare an RGB sensor system consisting of six RGB cameras and a RGB-D system consisting of two Kinects v1 regarding accuracy of joint positions and tracking of fast movements. They report that the overall performance is around the same between the two systems and that both systems are not recommended for applications requiring high precision.

Geerse et al. [93] evaluate a multi-Kinect v2 gait analysis system against a marker based system. Various gait parameters, such as walking speed, cadence, step length, stride length, step width, step time and stride time, are extracted from the 10 meter long gait sequences. They report that both gait parameters and the time series of the body points match well with their control evaluation system.

Eltoukhy et al. [79] investigate the accuracy of spatiotemporal gait parameters obtained from a Kinect v2 sensor when compared with an infrared camera motion analysis system. The authors conclude that the Kinect v2 sensor has the potential to be an effective

clinical tool for sagittal plane knee and hip joint kinematics as well as spatiotemporal parameters such as step characteristics.

Mentiplay et al. [173] measure the accuracy of spatiotemporal and kinematic parameters obtained with a Kinect v2 sensor. They report excellent concurrent validity for spatiotemporal parameters and poor to modest validity for kinematic parameters such as ankle flexion, knee flexion or hip flexion.

Hotrabhavananda et al. [121] evaluate the accuracy of stride length, stride time and walking speed obtained from both the Kinect v2 skeletal model and the depth image stream. Like Stone and Skubic [255], correlation coefficient time series is used to extract footstep locations from the depth data. Gait parameters are measured during the performance of the TUG test and the Figure of 8 Walk Test and comparison is done with a marker-based motion analysis system. The authors report an absolute mean percentile error of less than 3% for both the skeleton and the depth data. They also report higher accuracy for stride length and stride time measurements using the Kinect depth stream and slightly higher accuracy for the walking speed using the skeletal model.

Staranowicz et al. [252] examine the accuracy of stride length and stride duration obtained from the skeletal model of a Kinect v1. They mount the Kinect on a mobile robot which follows the observed subjects and compare their results with a marker based system (VICON). They report good accuracy with an average stride length error of 1.4 centimeters and an average stride duration error of 0.018 seconds.

Abiddin et al. [1] measure the accuracy of right hip angle and right knee joint angle obtained from the skeletal model of a Kinect v1 compared to an optical marker based system (VICON). They report a correlation coefficient of 0.996 for right hip angles and 0.987 for right knee angles, which is equal to an absolute mean difference of 0.7 degrees for right hip angles and 0.68 degrees for right knee angles. They conclude that Kinect can be an effective tool for to measure gait parameters depending on the required accuracy level.

Auvinet et al. [9] measure heel-strike accuracy of the Kinect sensor under the assumption that it can be estimated from extreme values of the knee joints distance. Knee joint distances are found by estimating the knee height in the depth stream as a constant fraction of the total height of the subject. Optoeletronic data is used to confirm the assumption of determining heel-strikes as extreme values of knee joints.

Dolatabadi et al. [74] evaluate spatiotemporal gait parameters using two Kinect v2 sensors and an electronic walkway (GAITRite). They report correlation coefficients from 0.9 to 0.98 for all gait parameters and conclude that the Kinect v2 is a valid clinical tool for measuring selected parameters.

González et al. [105] compare a wearable inertial system and a Kinect v1 with an electronic walkway (GAITRite) as a reference system. They report slightly lower relative and absolute errors for the wearable inertial system compared to their Kinect system.

Furthermore, the Kinect system is evaluated from two different heights (65cm and 240cm) and no substantial difference regarding the Kinect sensor height is found.

Bonnechère et al. [33] compare segment lengths and joint angles obtained with the skeletal model of the Kinect v1 with a marker based system (VICON). Parameters are obtained from subjects performing a deep squatting motion. The authors report poor accuracy for segment lengths, especially during motion, and good accuracy for most joint angles.

Fernández-Baena et al. [81] compare the accuracy of joint angles between Kinect v1 and an optical marker-based system (VICON). Even though the authors report mean accuracy errors ranging from 5° to 17°, they conclude that Kinect is useful for clinical applications. Like other authors, they suggest to improve Kinect accuracy by imposing a fixed segment length constraint.

Xu et al. [298] evaluate the accuracy of heel strike and toe-off events with a Kinect v1 sensor and a marker based system (Optotrak Certus System) during treadmill walking. Heel strike occurrences are determined as distance maxima between the ankle joint of the front foot and the hip center joint. Toe-off events are found as distance maxima between the ankle joint of the rear foot and the hip center joint. Their results show better accuracy for heel strike timing (0.18 to 0.30 frames) than for toe-off event timing (-2.25 to -2.61 frames).

Gieser et al. [97] evaluate the accuracy of the Kinect v1 skeletal model for rehabilitation exercises. Subjects are asked to simply walk around and move their arms. The authors find significant differences between the Kinect and an optical marker-based system (VICON) with an average difference of at least 5 cm per joint.

Funaya et al. [87] examine the accuracy of a Kinect v1 (OpenNI) with a marker based system (MAC3D) for assessing balance disorders. Six healthy subjects are recorded during the performance of several balance assessments. They propose their own system for finding temporal and spatial correspondences between the two sensor systems and they report adequate accuracy for standard balance tests.

Schmitz et al. [239] study the accuracy and repeatability of joint angles of the Kinect v1 sensor compared to a marker-based system (Motion Analysis Corporation) and an inclinometer. The authors find comparable accuracies for the Kinect-based system and the marker-based system for parameters such as flexion-extension, ab-adduction and axial rotation. The Kinect-based system is more accurate for estimating adduction angles and the marker-based system is better at estimating abduction angles.

Seung-kook Jun et al. [133] compare the joint angle accuracy of the Kinect v1 skeleton model with a marker-based motion capture system (VICON) during squatting movements. The authors observe that the performance of the Kinect v1 critically depends on the subject being in the recommended field of view and that only one subject can be seen at a time. Significant errors are reported for the Kinect system and they conclude that the Kinect has only limited usefulness for medical applications. They are able

to improve Kinect's accuracy with post processing such as Kalman filtering [137] or kinematic calibration.

Staranowicz et al. [253] evaluate their multi-Kinect skeleton fusion algorithm based on two spatiotemporal gait parameters: Stride length and stride width. In their experiment, three subjects walk in a straight line towards a Kinect v1 and a Kinect v2 sensor. Ground truths for stride length and stride width are obtained trough stick tape markers on the floor. For stride width, their fusion method performs the best, with the Kinect v2 being second and the Kinect v1 being third. For stride length, the Kinect v2 outperforms both their fusion method and the Kinect v1 with a significant difference between the Kinect v2 and the Kinect v1 (0.5 cm versus 2.3 cm mean deviation).

Wiedemann et al. [290] analyze the accuracy of 8 joint angles during a set of static posture including 14 seated and two standing postures. They report low median error rates for upper body and knee angle joints and large error rates for neck and hip angles. VICON is used as a gold standard.

Table 2.3 and Table 2.3 show the discussed papers on the accuracy of the Kinect v1 in chronological order. Datasets, experiment setups and extracted parameters vary greatly within the discussed studies. While this makes comparing the results of individual papers difficult, trends can still be recognized:

- The reported measurement errors for the extracted parameters are low enough to consider Kinect v1 good or adequate for the majority of applications. Only a few authors deem the sensor to be only limitedly useful.
- Several papers report problems due to varying and unreliable segment lengths [288, 81, 34, 33]. Segments lengths are the distance of adjacent joints and these distances are not treated as a constant by the Kinect SDK. For an arbitrary segment, a positional error of either joint causes a variation in segment length. In order to prevent variation in segment lengths over time several authors suggest to apply a constraint on segment lengths after their lengths are known.
- There is no noticeable difference for the height of the sensor placement [105]. However, only one study examines the effect of sensor height on the error of extracted parameters. Moreover, as the official recommendations suggest a sensor height between 0.6 and 1.8 meters [10], this should be further evaluated in future studies.
- Skeleton joints show poor performance regarding jitter [162].
- As already pointed out by Springer and Yogev [250], the Kinect is more reliable for the estimation of spatiotemporal parameters than for the estimation of kinematic parameters.

[10] https://support.xbox.com/de-AT/xbox-360/accessories/sensor-placement

Method	Year	Experiment Setup	Ground truth	Parameters	Conclusion
[255]	2011	Straight walking path recorded from two angles	MOCAP	Walking speed, stride time and stride length	good accuracy; worse performance for the Kinect placed perpendicular to the walking path
[259]	2011	Straight walking path recorded from two angles	MOCAP	Variabilities of stride time, stride length and stride velocity	Sufficient accuracy for variability of stride velocity, worse than web-camera system for stride length variability
[81]	2012	Rehabilitation exercises	MOCAP	joint angles (OpenNI skeleton)	useful for clinical applications
[197]	2012	six physical exercises	MOCAP	joint accuracy	similar accuracy to MOCAP for controlled poses; significant errors for general poses
[50]	2012	upper limb motor tasks	MOCAP	motion tracking performance	accuracy competitive to marker-based system
[34]	2012	deep squatting motion	MOCAP	segment lengths & joint angles	limited usefulness for biomechanical applications
[77]	2012	Four cubes placed in front of the Kinect	MOCAP	root-mean-squared distance error (cm)	error between 5.7 and 10.9 millimeters; potential to perform ergonomic assessments
[57]	2012	postural control tests (forward reach test & standing balance test)	MOCAP	distance reached, trunk flexion angle, joint accuracy & trunk flexion angle	excellent concurrent validity and potential to assess postural control in a clinical setting
[288]	2012	treadmill walking	MOCAP	segment lengths & joint angles	good accuracy for joint angles, unreliable segment lengths
[133]	2013	squatting motion	MOCAP	joint angles	limited usefulness; improved accuracy after applying Kalman filter
[252]	2013	robot following and recording walking subject	MOCAP	stride length and stride duration	good average accuracy
[33]	2013	squatting motion	MOCAP	joint angles & segment lengths	bad accuracy for segment lengths, good accuracy for joint angles
[56]	2013	straight 2.5 meter walkway towards the Kinect	MOCAP	spatiotemporal parameters	good agreement for gait speed, step length and stride length; poor agreement for swing velocity, step time and stride time
[87]	2013	Four different balance tests	MOCAP	distance between joints and markers	adequate accuracy
[97]	2014	arbitrarily walking around	MOCAP	joint accuracy	significant errors ($\geq$ 5 cm per joint)
[14]	2014	Straight walking path with four walking patterns	GAITRite	step length & step time	good accuracy for walking sequences worse accuracy for single steps
[209]	2014	treadmill walking	MOCAP	angular displacement, stride timing and tracking ability	good accuracy for stride parameters; bad accuracy for kinematic knee and hip parameters; currently unacceptable for clinical measurements
[239]	2014	A jig with a ball-and-socket joint to simulate a leg was placed in multiple static postures	MOCAP	joint angles	similar accuracy to marker-based system

Table 2.3: Chronological list of studies regarding Kinect v1 accuracy, based on a software perspective (part 1)

Method	Year	Experiment Setup	Ground truth	Parameters	Conclusion
[305]	2014	three standing balance tests	MOCAP	average velocity of the center of mass and position variability	Calibration with linear equations enables Kinect to reliably evaluate standing balance
[200]	2014	treadmill walking	MOCAP	absolute difference & correlation coefficient	excellent correlation with marker-based system
[222]	2015	free body movements, e.g. crossing arms, kneeling down, etc.	RGB sensor system	motion tracking under different conditions, e.g. slow movement, fast movement	good agreement with reference system; not recommended for applications with high precision requirements
[298]	2015	treadmill walking	MOCAP	heel strike & toe-off events	good accuracy for heel strike, bad accuracy for toe-off timings
[9]	2015	treadmill walking	opto-electronic	heel-strike timing	good accuracy
[162]	2015	virtual reality rehabilitation system	MOCAP & electro-magnetic	accuracy and jitter of joints	three times the accuracy error of the optical marker-based solution, bad results regarding jitter
[105]	2016	Two Kinects placed on different sensor heights perpendicular towards a 4.27 meter walkway	GAITRite	step time and stride time	slightly worse accuracy than inertial sensors; no substantial difference regarding the Kinect sensor heights
[1]	2016	Five meter walkway	MOCAP	hip, knee and ankle joint angle	effective tool for gait analysis

Table 2.4: Chronological list of studies regarding Kinect v1 accuracy, based on a software perspective (part 2)

- Marker-based motion capture (MOCAP) systems are by far the most popular choice as a ground truth. However, one approach also uses floor markers, which is the choice for obtaining a spatial ground truth used in this work.

Table 2.5 shows the six discussed papers for Kinect v2 accuracy. As shown in the table, conclusions are generally favorable with the exception of the accuracy of kinematic parameters [173]. Only one paper (Staranowicz et al. [253]) uses both Kinect v1 and Kinect v2 in their experiment and compared their performance for the computation of spatiotemporal gait parameters. Kinect v2 performs better than Kinect v1 for both estimated parameters.

It should be noted that, compared to Kinect v1, the accuracy of Kinect v2 seems understudied as it received considerably less attention within the scientific community. While a large amount of papers are found for the accuracy of Kinect (see Table 2.3 and Table 2.4), only a few papers of those focus on Kinect v2 accuracy (see Table 2.5).

Several reviewed studies show that the Kinect is a viable tool for the extraction of spatiotemporal gait parameters. Based on this observation, the measurement error for the spatiotemporal parameters extracted in this work is expected to be reasonably low. Moreover, when working with the Kinect SDK skeleton model, possible improvements should be considered. These include applying a constraint for segment lengths or applying a Kalman filter on the position of the joints. While other improvements have been tested,

Method	Year	Experiment Setup	Ground truth	Parameters	Conclusion
[93]	2015	Straight 10-meter walkway with 4 Kinect v2 placed every 2.5 meters	MOCAP	spatiotemporal parameters	excellent absolute accuracy
[173]	2015	comfortable and fast paced walking trials towards the Kinect	MOCAP	spatiotemporal and kinematic	good accuracy for spatiotemporal parameters, bad accuracy for kinematic parameters
[253]	2015	straight line towards two Kinects (1 Kinect v1 & 1 Kinect v2)	stick tape floor markers	stride width and stride length	Kinect v2 outperformed Kinect v1 for both stride width (slightly) and stride length (significantly)
[290]	2015	14 seated and 2 standing body postures	MOCAP	8 Joint angles	good accuracy for upper body angles and knee angles; poor accuracy for neck and hip angles
[74]	2016	Straight walking towards the Kinect under different walking conditions (normal speed, walking while counting backwards, fast paced walking)	GAITRite	spatiotemporal parameters	valid clinical tool
[121]	2016	Timed Up & Go test	MOCAP	stride length, stride time and walking speed	Better accuracy for stride length & stride time using the depth stream and better accuracy for walking speed using the skeleton model
[79]	2017	treadmill walking	MOCAP	spatiotemporal and kinematic	potential to be an excellent clinical tool

Table 2.5: Chronological list of studies regarding Kinect v2 accuracy

e.g. using multiple Kinect sensors or using a sensor fusion approach, there is no clear consensus about their potential error reductions. However, these potential improvements are not evaluated within the scope of this work.

2.3 Gait recognition and gait analysis

Gait recognition refers to the identification of a person from its walking style [244]. Other research domains are covered in their respective surveys, e.g. recognition of activities [217, 4], faces [127], gestures [205], hand pose estimation [80], etc.

While gait recognition is a different area of application than vision-based human gait analysis, it can be seen that the pipeline of steps applied for both gait recognition and gait analysis is very similar. The main difference between both applications is that gait recognition applies a classifier on the extracted gait parameters (also called gait signature), while this is not necessarily true for gait analysis.

2.3.1 Overview: Past and Present

One of the earliest surveys of vision-based motion analysis was done by Aggarwal et al. [3] in 1995. They describe methods of articulated and elastic non-rigid motion and classify them as either approaches with or approaches without a priori shape models. This classification relies on whether prior knowledge of the object shape is incorporated into the motion analysis technique. For model-based approaches, they distinguish between stick figure representation and volumetric models.

Cedras and Shah [46] identify two main steps for gait recognition: Creation of a motion model based on the extracted motion information and matching the input with the previously constructed model. They distinguish three types of motion information: Region-based features, optical flow and trajectory-based features.

Aggarwal and Cai [2] review three categories: Body structure analysis, tracking and activity recognition. Body structure analyis, which is highly similar to pose estimation, is split into model-based and model-free. They further divide approaches based on the way they represent the human body: As a Stick figure, a 2D contour or a 3D volume. Tracking is divided into single-view and multi-view. Gavrila [91] distinguish 2D approaches from 3D approaches. For 2D, approaches are split depending on whether or not they use a shape model.

The taxonomy of Moeslund and Granum [180, 181] groups approaches based on the subsequent phases of human motion analyis: Initialization, tracking, pose estimation and recognition. Pose estimation is further divided into model-free, indirect model-based and direct model-based. Tracking, pose estimation and recognition highly overlap with the three categories proposed by Aggarwal and Cai [2].

Wang et al. [285] use similar categories as [2, 180, 181]: Human detection, human tracking and human behavior understanding. Human tracking is further subdivided into model-based, region-based, active contour-based and feature-based.

Wang and Singh [284] review two phases of gait recognition: Tracking and motion analysis. Both tracking and motion analysis are split into two categories: Model-based and model-free (model-free tracking was called motion-based tracking).

Menier et al. [172] distinguish three categories for markerless motion tracking: Learning-based, model-free and model-based methods. Wang et al. [283] review gait recognition methods prior to 2011, focusing on gait image representation, feature dimensionality reduction and gait classification. They distinguish gait image representations as either model-based or model-free. Ji and Liu [128] review human detection, pose estimation and behavior understanding, focusing on view-invariant approaches.

Lee et al. [154] consider model-based and model free gait recognition approaches. Both categories are further divided into two subcategories: Model-based approaches are divided based on the features used, either waveform derived features together with image features or using time as an implicit feature by using the set of images directly. Model-free approaches are subdivided into temporal correspondence and spatiotemporal motion

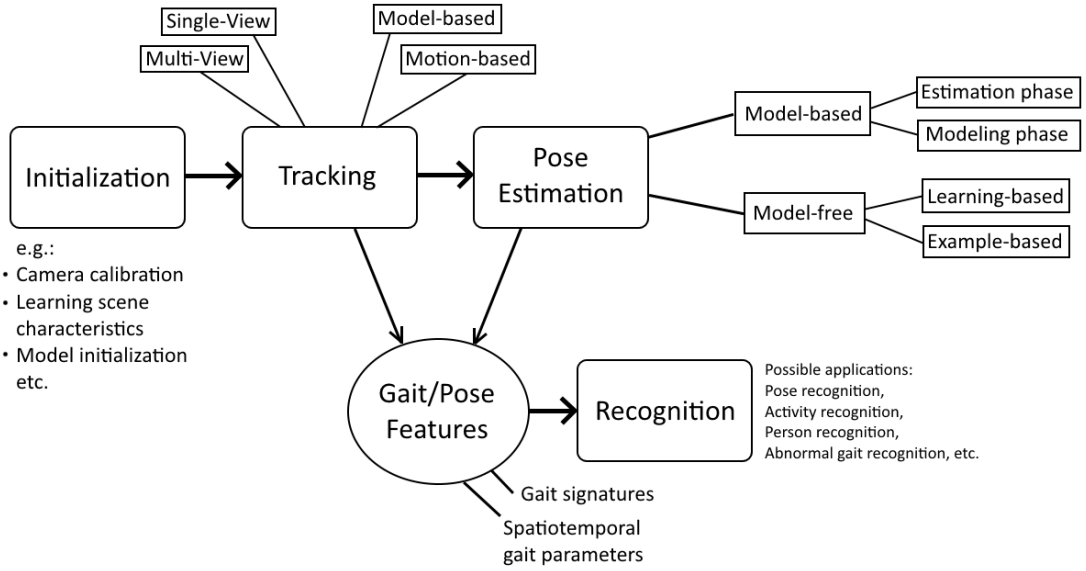


Figure 2.9: A general structure for human body motion analyis and its recognition applications.

summary. Temporal correspondence refers to comparing features between the test input and existing patterns, e.g. spatial comparison on a frame by frame basis. Instead of comparing individual frames, spatiotemporal motion summary refers to comparison of summarized motion features, e.g. by utilizing statistical approaches.

Surveys generally distinguish based on the dimensionality of the approach [2, 91], based on the phases of motion analysis [180, 181, 284] or based on the usage of shape models [216, 233, 154].

Due to the complexity of gait recognition, scientific papers specialize on a small part, e.g. system initialization and camera calibration, segmentation and tracking, pose estimation, feature extraction, classification and recognition, etc. For a viable gait recognition system, all of these steps have to be considered. Figure 2.9 shows the general structure of a motion analysis system, based on several motion analysis surveys [180, 284, 216].

2.3.2 Gait recognition

Two major types of gait recognition approaches are considered, model-based and model-free.

Model-free gait recognition:

Model-free approaches analyze the motion of the subject or the shape made during the walk [32]. They consist of the following steps: Subject detection, silhouette extraction, feature extraction and classification [244]. Gait features are typically extracted from

silhouette shapes, contours or the whole motion of the human body [283]. Different types of features may be extracted, e.g. based on silhouette shape [286], silhouette width vector [136], Fourier descriptors [187], optical flow [159], etc. or based on the dynamics information such as temporal alignment [235, 297, 244].

The majority of gait recognition research is done on model-free approaches [154]. For human identification, model-free approaches have the advantage of being generally faster and less sensitive to the quality of the gait sequences [244]. Model-free approaches are also called motion-based [286], feature-based [51] or appearance-based [297] approaches.

One specific example for a model-free gait features is gait energy image (GEI) [169], which is simply the mean body silhouette in the sagittal plane over several gait cycles. GEI encodes the continuous changes of pose during human walking as a moving probability graph [151, 158]. Higher values within the GEI correspond to regions where the human body appears frequently during the walking motion [158]. GEI is used as a gait feature representation in several model-free gait recognition methods, e.g. Bashir et al. [19]. Other model-free approaches include encoding silhouettes, e.g. using central moments [38] or Hu moments [229], or encoding contours, e.g. based on shape contexts [21].

Model-based gait recognition:

Model-based gait recognition approaches fit a structural model, a motion model or both to the person in every frame of the walking sequence [32, 303]. A structural model describes the topology of the human body, e.g. head, torso, knees and ankles. Parameters, e.g. static parameters such as lengths of body parts or dynamic parameters such as stride length, angular velocities or trajectories, are measured on the estimated body model. The process of creating a body model is called pose estimation [180, 128]. A motion model describes the kinematic or dynamic aspects of the motion of each body part [303]. In other words, a motion model describes how the person moves [32]. An example for employing a motion model-based approach is the work of Cunado et al. [62, 63]. They represented the periodic thigh angular motion during gait as Fourier series (see Figure 2.10f).

Different types of structural models are used (see Figure 2.10a- 2.10e). The level of details varies between just the subjects' height to the combination of structural and dynamic information of the subject [180]. Examples for simple structural models are a five segment stick figure [196] or the aspect ratios between certain limbs [112]. Examples for detailed models are the works of Gavrila and Davis [92] and Plänkers and Fua [214]. Gavrila and Davis use two components to represent the human body in 3D: A stick figure to represent the skeletal structure and a representation of the surrounding flesh using volumetric primitivies. Plänkers and Fua [214] use metaballs to represent the flesh of the subject and a polygonal structure to represent the skin, in addition to a kinematic model that represents the bone structure. The estimated pose is generally not very detailed, typically only positions of major limbs or a rough description of the human body are

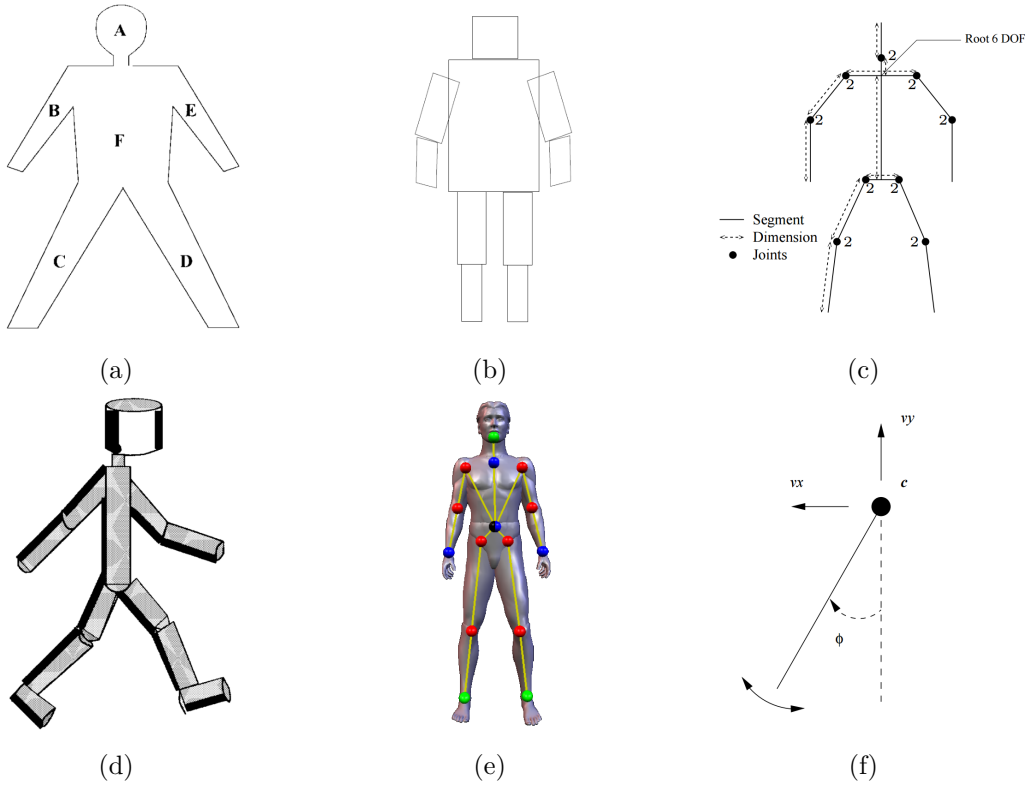


Figure 2.10: Different body models used in model-based approaches: (a) 2D ribbon model [2, 157]. (b) 2D Cardboard model [132]. (c) Stick figure or skeletal model [172]. (d) Volumetric model [2, 119]. (e) 3D surface model [45]. (f) Motion model [63].

used as a pose representation [180] Elaborate, surface-based shape models are instead found for computer graphics applications, e.g. free-viewpoint video [45].

One of the earliest model-based gait recognition methods is the approach by Niyogi and Adelson [196]. They create an XYT image cube from a walking sequence using time as the third dimension. Then active contour snakes are used to detect spatio-temporal patterns of a person's walk in the XT plane. Applying this procedure to multiple levels in the XT plane creates bounding contours of the walking person. A simple five-stick body model is fitted to the motion of the bounding contours. Gait information is then acquired from the periodic angle signals from upper and lower leg. A drawback of this approach is its limitation to side view.

BenAbdelkader et al. [22] estimate two spatiotemporal parameters, stride length and cadence, for human identification. Both parameters are estimated from the periodicity of gait. They use the width of the walker's bounding box as a function over time to estimate the periodicity. Stride length is estimated from the periodicity and the total distance traveled. Cadence is estimated from the periodicity and the total walking time.

Model-based methods have the advantage of being invariant to viewing angle and scale [283] and being more robust to occlusion and noise [244].

The described structure and taxonomies for gait recognition also applies to gait recognition based on depth data. The gait analysis methods implemented in this work are based on depth data and skeleton tracking. Both approaches are model-based. Gait analysis based on skeleton model is clearly model-based, as it utilizes the ankle joint coordinates provided by the pose estimation step. While the gait analysis approach based on depth data does not estimate the whole human pose, it estimates the position of the feet and is therefore also model-based. Hence the advantage of model-based methods also applies to the gait analysis methods implemented in this work. Therefore it is expected that the viewing angle does not influence the error of the extracted spatiotemporal parameters.

Pose Estimation Pose estimation estimates the configuration of the human body in an image or image sequence [233].

3D model-based pose estimation is the most widely investigated approach for pose estimation [128], due to the fact that 2D human models are only suitable for motion parallel to the image plane [216].

Two types of pose estimation approaches are distinguished, model-based and model-free [216, 128]. Some authors consider additional categories, e.g. Moeslund and Granum [180] divide model-based into direct model use and indirect model use. Sarafianos et al. [233] consider part-based pose estimation, a subcategory of model-based approaches, and hybrid approaches, a combination of model-based and model-free approaches.

Poppe [216] divide model-based pose estimation into a modeling and an estimation phase. The modeling phase constructs a function that returns the image likelihood, based on a set of parameters describing body shape, body appearance and camera. The estimation phase minimizes the error between projection of the human body model and image observations. He distinguishes two types of body models, kinematic models and shape models. Kinematic models describe the human body with segments linked by joints. Shape models may use 2D patches, e.g. rectangular patches, volumetric shapes, e.g. spheres, or surfaces to describe the human body. Model-free pose estimation approaches are classified as learning-based and example-based. Learning-based approaches use training data to learn a function from image space to pose space. Example-based approaches store exemplars together with their pose description in a database. A pose estimate for an input image is obtained by performing similarity search and interpolation of candidate poses.

Poppe [216] describes five image descriptors used for pose estimation in model-based approaches: Silhouettes and contours, edges, 3D reconstructions, motion and color and texture. Perez-Sala et al. [206] consider similar descriptors as Poppe [216]: Silhouettes and contours, intensity and color and texture, motion information, logical information and depth information. From the list of available image descriptors, only 3D reconstruction and depth information are relevant for this work.

A scene can be reconstructed in 3D when stereometry, a depth sensor or multiple cameras are used. Two common techniques for using multiple RGB cameras are volume intersection and a voxel-based approach. For volume intersection [36, 42] the silhouette of the subject is first extracted in multiple, calibrated camera views. Then the silhouettes from each camera view get projected into a discretized volume space and a human body model can be fitted to the resulting volume. Another possibility is a voxel-based approach [53, 175, 260]. Stone and Skubic [260] extract silhouettes from each of two RGB cameras. They compute 3D objects formed by the intersection of the projection of a single connected component from each silhouette image per camera. Voxel objects are classified as either human or non-human based on two features: Connected component usage and position corrected volume. In order to allow long-term observation, they use an update procedure based on the frame-to-frame motion estimated from the overlapping blocks: Blocks which contain motion are prevented from updating, blocks which contain non-human voxels update quickly and background pixels update slowly. They also detect footfalls in order to estimate spatiotemporal gait parameters [257]. Voxels are projected to the ground plane and then grouped into clusters, which represent footfalls.

Stereometry uses triangulation to calculate the depths of corresponding points of a calibrated pair of cameras. This approach resembles human vision and is used for human pose estimation, e.g. Haritaoglu et al. [112].

Depth information can be directly used for 3D reconstruction to obtain a point cloud of the subject. Furthermore, several new depth-specific descriptors have been proposed due to the popularity peak after the release of Microsoft Kinect, e.g. Gabor filters over depth maps [220] or saliency of depth maps [212]. The additional depth information facilitates human pose estimation. One approach of Schwarz et al. [241] uses a graph-based representation of the 3D point cloud obtained from depth data to compute geodesic distances between different body parts. The information is then combined with the optical flow computed from two successive intensity image frames in order to estimate a skeleton model of the subject. Pose estimation from depth frames is even feasible from single depth images without considering temporal information, e.g. the approach of Shotton et al. [245], which is used for the pose estimation of Kinect v1. Figure 2.11 illustrates the basic pipeline used for the Kinect v1 skeletal tracking approach.

In this work the pose estimation of the Kinect v2 SDK is utilized for gait analysis. It fits a skeleton model with 25 joints on the subject within the field of view. It is important to note that a number of alternative pose estimation methods exist and may perform better in some scenarios, e.g. when sitting. For example, Cippitelli et al. [55] propose the use of anthropometric models to locate the position of 6 joints in side view. Evaluation with a marker based system shows lower errors of the trajectories compared to the corresponding trajectories obtained with the Kinect SDK. However, in this work a view-invariant pose estimation approach is required and therefore the Kinect v2 SDK is chosen.

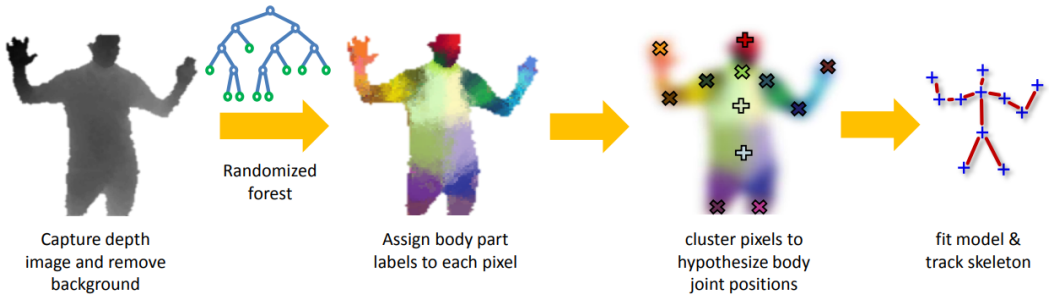


Figure 2.11: Basic pipeline of the Kinect v1 skeletal tracking system [146].

2.3.3 Gait recognition based on Kinect

Depth sensors such as Kinect have the advantage of providing a depth value for every pixel of the scene [52], offering a 2.5D view of the scene. Depth information, restricted to a single viewpoint, is referred to as 2.5D [264]. Depth imagery simplifies various computer vision tasks, e.g. background subtraction or contour detection [52].

The Kinect v1 and the Kinect v2 sensors have been a popular choice for both gait recognition and gait analysis based on depth information. The following scientific publications show efforts that have been exerted for gait recognition based on Kinect.

Gabel et al. [88] compare the accuracy of gait parameters they extracted from the skeleton model produced by the Kinect SDK with parameters extracted from wearable sensors. They extract the direction of progress and the walking speed from the change in position of the center of mass. Stride time and stride length are extracted based on a state machine consisting of a HEEL, TOE and SWING state.

Preis et al. [218] extract thirteen biometric features consisting of 11 segment lengths, step length and walking speed from the Kinect v1 skeletal data. Gianaria et al. [95] use the Kinect skeleton model to extract a set of anthropometric features and gait features. Their feature set consists of segment lengths, distances between joints, variances of joint movements and two spatiotemporal gait parameters, gait speed and stride length. Stride length is obtained as the distance between two stationary positions of a foot.

Ball et al. [15] recognize 4 individuals applying unsupervised clustering on skeleton data obtained from the Kinect using K-means with an accuracy of 43.6%. Yang et al. [304] recognize humans based on gait features, including a novel set of relative distance-based features, using a Kinect v1 and KNN classification.

Jiang et al. [129] use the Kinect v1 skeleton model to obtain bone lengths as a static feature and joint angles as a dynamic feature for human recognition. Nearest neighbor was used as a classifier. Munsell et al. [188] perform action classification and human identification based on the Kinect v1 skeleton model and SVM. They extract a high-dimensional feature set consisting of joint coordinates, radial and elevation motion patterns and proportions between joints.

Milovanovic et al. [178] store 3D keypoints of the Kinect skeleton model of each frame of a walking sequence in a single image row. The resulting image contains the whole joint coordinate information transferred into a 2D image representation. Person identification is done using content-based image retrieval techniques on the result images.

Sinha et al. [247] propose two types of gait features obtained from the Kinect skeleton data. They use area features, which represent the area of upper or lower body parts, and distance features, which represent the distances between the upper body centroid and different joints of upper or lower limbs. An adaptive neural network is used for human classification. They report an overall average accuracy of only 25.2%, which increased to 86% after they extract gait cycles manually.

Banerjee et al. [17] apply unsupervised clustering on shape descriptors such as bounding box information and image moments to distinguish the gait from two elderly residents and their visitors. Possibilistic C-Means clustering is used to identify the gait patterns from the residents and temporary visitors, which are considered outliers.

Hofmann et al. [117] adapts Gait Energy Images (GEI) [169] to depth information from a Kinect v1. In contrary to GEI obtained from color information, using the depth information provided by the Kinect offers a more reliable way to obtain an accurate silhouette. For their nearest-neighbor classification they calculate a feature descriptor from the silhouette by applying histograms of oriented gradients (HOG) [65], averaging the gradient histograms over a gait cycle and applying dimensionality reduction.

Liu et al. [160] propose a side-view gait recognition method using joint angles of the Kinect v2 skeleton model. Dikovski et. al. [72] evaluate different feature sets obtained from the skeletal data of a Kinect v1 for recognizing humans based on their gait. They report lower body parts having better gait information than upper body parts and features like height or joint distances performing better than angular features.

Andersson et al. [8] use 20 anthropometric and 60 gait features extracted from a Kinect v1 to identify individuals with a KNN classifier. Gait features consist of joint angles and spatiotemporal gait parameters such as stride length, gait cycle time and gait velocity. Anthropometric features consist of 19 segment lengths and the subject's height. They observe that when used separately, anthropometric attributes are significantly more useful than gait attributes.

Ahmed et al. [6] observe the spatio-temporal changes of relative angles among different pairs of skeletal joints for Kinect v2 gait recognition.

The discussed Kinect-based gait recognition approaches are summarized in Table 2.6. The majority of the approaches utilize the skeleton model provided by the Kinect SDK pose estimation. Therefore the majority of gait recognition features are directly related to the skeleton model, e.g. joint trajectories, joint angles or segment lengths.

Article	year	sensor	Depth data/Skeleton data	Gait recognition features
Ball et al. [15]	2012	Kinect v1	skeleton data	18 lower limb joint angle features
Hofmann et al. [117]	2012	Kinect v1	depth data	depth-based gait energy image computed from silhouette
Banerjee et al. [17]	2012	Kinect v1	depth data	bounding box information and image moments computed from silhouette
Munsell et al. [188]	2012	Kinect v1	skeleton data	joint coordinates, radial motion patterns and proportions between joints
Preis et al. [218]	2012	Kinect v1	skeleton data	segment lengths, step length and walking speed
Gianaria et al. [95]	2013	Kinect v1	skeleton data	segment lengths, stride length, walking speed, distances between several joints and variance of joint movements
Sinha et al. [247]	2013	Kinect v1	skeleton data	area of upper or lower body parts and distance between upper body centroid and several limb joints
Milovanovic et al. [178]	2013	Kinect v1	skeleton data	joint coordinates stored in image
Dikovski et. al. [72]	2014	Kinect v1	skeleton data	7 sets of features consisting of segment lengths and joint angles
Jiang et al. [129]	2014	Kinect v1	skeleton data	segment lengths and joint angles
Ahmed et al. [6]	2015	Kinect v2	skeleton data	relative joint angle features
Andersson et al. [8]	2015	Kinect v1	skeleton data	gait features and anthropometric attributes
Yang et al. [304]	2016	Kinect v1	skeleton data	relative coordinates of the skeleton joints
Liu et al. [160]	2016	Kinect v2	skeleton data	joint angle features from side-view

Table 2.6: Overview of Kinect-based gait recognition approaches.

2.3.4 Gait analysis for medical purposes

The field of health care based on depth imaging includes a large amount of applications, e.g. rehabilitation, automatizing patient setup in diagnostic imaging, sleep monitoring, assistive technology for blind or handicapped people [20]. Applying Kinect for physical therapy and rehabilitation is covered in a survey by Mousavi and Khademi [186].

Researches deploy Kinect for gait analysis for the detection of gait pathologies such as gait asymmetry, the detection of Parkinson's disease, Multiple Sclerosis and general gait abnormalities. The following papers show how Kinect can be utilized for this purpose.

Kastaniotis et al. [139] use the Kinect to record both healthy subjects and Multiple

Sclerosis patients performing the 2-minute walk tests. They are able to estimate the walked distance from the depth data (with minor errors) and distinguish between the healthy subjects and the patients (88% correct classifications). They use Linear Discriminant Analysis as a classifier on a view invariant representation of the skeleton information provided by the Kinect SDK.

Prochazka et al. [219] distinguish between healthy, elderly subjects and individuals with Parkinson's disease based on mean stride lengths (normalized by leg lengths of the subjects) obtained from the skeleton model of the Kinect v1. They use a flat stride length threshold for the classification and achieved 91.7% accuracy.

Rocha et al. [224] use a Kinect v1 to discriminate between three Parkinson's disease subjects and three control subjects. They calculate 34 different gait parameters from which the variance of the central shoulder velocity is chosen as it best distinguishes between three states of Parkinson's disease.

Parajuli et al. [201] propose a system distinguishing between normal and abnormal gaits as well as sitting and walking. They record one dataset for each activity using a Kinect v1 positioned at 1 meter height. They choose an SVM classifier and experimented with different parameter inputs to improve the result. Tucker et al. [271] ensure medication adherence among Parkinson's disease patients by observing the gait patterns using a Kinect v1. Their method makes use of the skeleton model obtained from four different walking directions.

Dao et al. [66] use the skeleton data obtained from a Kinect v1 to classify between normal and abnormal gait using a linear SVM. Bera et al. [25] inspect the gait pattern of elderly people to detect pain within selective body joints. They apply several post-procession and data reduction steps on the coordinates of the joints obtained from a Kinect v1 before classifying the data using a linear SVM. Ndayikengurukiye and Mignotte [193] distinguish normal gait from abnormal gait by visualizing gait as a high-frequency spectral energy map and applying a number of classifiers.

Auvinet et al. [10] analyze the difference in body volume between normal and pathological walk using a set of three Kinect v1 sensors. Each camera is calibrated before applying volume reconstruction and volume analysis to obtain the gait parameters. Later Auvinet et al. [11] use a Kinect v1 to record subjects walking on a treadmill for the measurement of gait asymmetry. They convert frames of the depth stream into a point cloud and reproject them onto an orthographic plane. Left-right differences are assessed by using horizontal flipping on key depth maps, which are averaged, ideal representations of gait cycles. Their method is validated using healthy subjects and subjects simulating the gait pathology with an asymmetric sole. More recently, Auvinet et al. [12] propose a gait asymmetry index based on the longitudinal spatial difference between lower-limb movements during gait cycle, which they obtain using Kinect. Their asymmetry index performs better at distinguishing asymmetrical gait compared to common spatiotemporal gait parameters for healthy subjects.

Kaenchan et al. [135] use a system consisting of three Kinects v1 to analyze the upper-

body posture and detect unbalanced gait. Nguyen et al. [194] present a pose estimation method based on the Kinect v1 depth stream for motion asymmetry detection. Dolatabadi et al. [73] distinguish between healthy and pathological gait using the trajectories of the skeleton joints of the Kinect v2 as features for their classification algorithm.

Chaaraoui et al. [49] use a skeleton-based spatiotemporal feature called joint motion history for abnormal gait detection. The feature is built by normalizing the skeletons, tracking the motion over a segment of frames in order to combine spatial and temporal information and applying dimensionality reduction. Both Kinect v1 and Kinect v2 are used in their evaluation.

Clark et al. [58] examine the reliability of a gait analysis system for the assessment of people who suffered from a stroke. They extract seven spatiotemporal gait parameters using a Kinect v1 sensor and compare them with various clinical assessments, e.g. the TUG test. Even though they report a slightly higher variance for the Kinect-derived parameters compared to the clinical assessments, the overall scores are similar on all measures. They conclude that Kinect-derived parameters are highly reliable, but also very redundant.

Leightley et al. [156] recognize motions such as chair rise, tandem balance or walking 4 meters from Kinect v2 skeleton data. They first group skeletal joints into five groups, one for each human extremity plus head and torso. Then a feature set is computed from the five body groups based on Euler angles, Euclidean distance, body lean angle and center-of-mass. A feature reduction step is applied before classifying the motion as one of five motion categories. After the classification of the motion type, they also compare the feature vector with mobility models generated from healthy subjects in order to classify a person's mobility as either good or poor.

Ye et al. [306] use the skeleton model of Kinect v2 for the classification of gait phases between stroke patients and healthy subjects. They identify nine gait phases between two consecutive heel strikes with the same foot. Gait patterns are characterized based on 12 proposed gait parameters, e.g. the ratio between the distance of the feet and the height of the hips or the angle between knee and hip during thigh raise. A neural network based on a non-linear autoregressive exogenous (NARX) model is used for the classification step.

Table 2.7 shows a summary of the described medical gait analysis approaches. Since human gait is related to a large number of medical conditions, the applications of the reviewed approaches are manifold. However, the procedure of the approaches is similar. Based on either Kinect depth or skeleton data a set of gait parameters is extracted and optionally, the parameters are used for classification.

2.3.5 Frailty detection and fall risk assessment

A specific medical gait analysis application is the detection of frailty and fall risk. The goal is to identify frail subjects or potential fallers and separate them from vigorous

Article	Year	Application	Sensor	Depth data/Skeleton data	Gait features
Auvinet et al. [10]	2011	Gait asymmetry	3 Kinect v1	depth data	Gait frequency obtained by applying Fourier transformation on the centroid movement
Auvinet et al. [11]	2012	Gait asymmetry	Kinect v1	depth data	Gait asymmetry index computed from key depth maps
Parajuli et. al [201]	2012	Abnormal gait	Kinect v1	skeleton data	Skeleton joint coordinates
Gabel et al. [88]	2012	Gait analysis	Kinect v1	skeleton data	Stride duration and arm angular velocities
Kaenchan et al. [135]	2013	Abnormal body posture	multiple Kinect v1	skeleton data	Upper-body tilt
Kastaniotis et al. [139]	2014	Multiple Sclerosis assessment	Kinect v1	skeleton data	Euler angles of selected limbs
Rocha et al. [224]	2014	Parkinson's disease assessment	Kinect v1	skeleton data	34 gait parameters, e.g. central shoulder velocity variance
Chaaraoui et al. [49]	2015	Abnormal gait	Kinect v1 and Kinect v2	skeleton data	Joint motion history computed from several skeleton frames
Auvinet et al. [12]	2015	Gait asymmetry	Kinect v1	depth data	Gait asymmetry index computed from comparing the spatial position of the left and the right legs at their respective step cycle
Prochazka et al. [219]	2015	Abnormal gait	Kinect v1	skeleton data	Normalized mean stride lengths
Dao et al. [66]	2015	Abnormal gait	Kinect v1	skeleton data	Gait features based on hip progression line, foot steps and X-rotation angle extracted from a Biovision Hierarchy file
Bera et al. [25]	2015	Joint pain detection	Kinect v1	skeleton data	Skeleton joint velocities
Tucker et al. [271]	2015	Medication adherence	Kinect v1	skeleton data	Large (1890-dimensional) feature set consisting of joint coordinates, joint velocities, joint accelerations and ratios between joints
Clark et al. [58]	2015	Stroke patient assessment	Kinect v1	skeleton data	7 spatiotemporal gait parameters, e.g. mean gait speed, step length, step length asymmetry, etc.
Leightley et al. [156]	2016	Abnormal gait	Kinect v2	skeleton data	Feature set from 5 body groups based on Euler angles, Euclidean distance, body lean angle and center-of-mass
Ndayikengurukiye and Mignotte [193]	2016	Abnormal gait	Kinect v1	depth data	high-frequency spectral energy maps
Ye et al. [306]	2017	Stroke patient detection	Kinect v2	skeleton data	12 gait parameters, e.g. ratio of foot distance and hip height, angle between knee and hip during thigh raise, etc.

Table 2.7: Summary of Kinect-based gait analysis approaches for gait analysis in a medical context.

walkers. An example for a fall prevention approach based on physical activity is the work of Planinc and Kampel [213]. They monitor the activity of elderly subjects based on alert lines, which describe the subject's inactivity during one minute intervals. If a significant amount of intervals of an alert line are outside the range $\mu \pm 2\sigma$ (μ denotes the mean and σ denotes the standard deviation of previous alert lines), a change in long-term mobility is detected and an alarm is generated. Fall prevention technologies are reviewed in a survey by Hamm et al. [110]. The authors focus on prevention and intervention systems regarding pre-fall prevention, post-fall prevention, fall injury prevention and cross-fall prevention. The following papers describe existing work on frailty detection and fall risk assessment based on gait analysis.

Phillips et al. [210] analyze the association of changes in in-home gait speed and stride length immediately before a fall with occurring fall events. The raw depth data of a Kinect sensor, which is located on a small shelf near the ceiling, is used to detect walks. A logistic regression model is used to predict falls based on gait parameters obtained from the walks within 30 days before a fall as well as random 30-day windows for residents with no falls. They observe a significant association of both gait speed and stride length changes with the occurrence of falls. Their model estimates, that a cumulative decline in gait speed of 2.54cm/s within 7 days in in-home gait speed increases the odds to undergo a fall within the next 3 weeks by 4.2 compared to a resident falling with no observed gait changes. For stride length they observe that a decrease of 2.5 cm over 7 days increased the odds of falling by 6.8.

Gianaria et al. [96] extract frailty-related gait parameters from the walking phase of a TUG test in front of a Kinect v2 sensor. The spatiotemporal and postural parameters are TUG time, covered distance, walking time, walking speed, swing time, double support time and torso inclination angle. Spatiotemporal parameters are estimated from the skeleton data. Their algorithm assigns each foot one of two states, depending if the foot is moving or standing still, based on a movement threshold. One of the drawbacks of this algorithm is that the Kinect pose estimation does not register the movement of the foot when it's occluded by the other foot. Torso inclination angle is estimated as the angle between walking direction and spine joints. Excessively tilted forward postures are related to increased fall risk. The obtained parameters are compared with TUG time and Tillburg Frailty Indicator score [103] for correlations.

Dolatabadi et al. [75] track gait parameters of an elderly subject with hip replacement surgery on one side with a Kinect v1. The spatiotemporal gait parameters step length, stance time per side, stride length and cadence are extracted over the course of 9 weeks. Swing and stance phases per foot are detected based on the movement of the ankle joint. Skeletal data with a center of mass outside of the Kinect view range is discarded. The authors find significantly smaller step lengths before the operation on the operated side and no changes on the non-operated side. They also notice an improvement in gait characteristics within the first 6 weeks after the operation.

Stone and Skubic [255] estimate walking speed, stride time and stride length for in-home fall risk assessment. They use background subtraction to create the subject's silhouettes

on each frame and use it to obtain a point cloud of the subject. Both feet are extracted from the point cloud using a fixed height threshold. The remaining feet point cloud is projected onto the ground and a correlation coefficient is computed for each frame. The resulting correlation coefficient signal is used to extract footsteps. Gait sequences, which have an erroneous step sequence, are filtered out and do not affect their evaluation. Later, Stone et al. [261] propose a system for automated health alerts in case of sudden changes of in-home gait parameters. Walking speed, stride length and stride time are continuously monitored based on their previously proposed approach. Alerts are generated when the measured parameters are outside of a threshold of their standard deviation. Their method is evaluated based on three retrospective case studies. More recently, Stone et al. [256] use average in-home gait speed as a metric for fall risk assessment. They evaluate the metric based on how well it can predict the score of a set of traditional clinical assessments with average gait speed performing the best.

Dubois and Charpillet [76] develop a fall risk assessment system based on the depth stream of a Kinect v1. First, a background subtraction algorithm is applied using the difference between the running average depth and the current depth values. Then the center of mass is calculated as the average of the remaining foreground pixels. The maxima of the vertical displacement of the center of mass during the walk are then used to obtain the positions of the feet during stance phases. The spatiotemporal parameters step length, step duration and walking speed are extracted based on the footfall positions and timings. For their evaluation they use four different walking setups: Normal walking perpendicular to the camera, small step walking perpendicular to the camera, walking with long skirts perpendicular to the camera and normal walking towards the camera. Comparison with an electronic walkway (GAITRite) as a reference system shows small errors for all walking setups except for walking towards the camera, where the error is significantly larger.

Table 2.8 shows a summary of the described papers for frailty detection and fall risk assessment based on Kinect-based gait analysis. Again, the functional principle resembles previous approaches on gait recognition and medical gait analysis. However, a few observations can be made regarding the differences. First, unlike previously discussed approaches on gait recognition and medical gait analysis, the majority of the approaches are based on the raw depth data instead of skeleton data. One reason is that the set of joints offered by the skeleton data is not needed for the extraction of basic spatiotemporal parameters such as gait speed, step length or step time. From the discussed approaches, only Gianaria et al. [96] utilize skeleton data since they extracted additional parameters regarding gait and posture such as swing time, double support time or torso inclination angle.

2.3.6 Gait analysis summary: Relation to this thesis

In this subsection, gait recognition and medical gait analysis approaches are reviewed. A general structure for gait recognition approaches is presented based on previous surveys regarding gait recognition based on color information. By discussing a large number of

Article	Year	Application	Sensor	Depth data/Skeleton data	Gait features
Stone and Skubic [255]	2011	Fall risk assessment	Kinect v1	depth data	gait speed, stride length, stride time
Dolatabadi et al. [75]	2013	Gait monitoring	Kinect v1	skeleton data	step length, stance time per side, stride length and cadence
Stone et al. [261]	2014	Health alerts	Kinect v1	depth data	gait speed, stride length and stride time
Dubois and Charpillet [76]	2014	Fall risk assessment	Kinect v1	depth data	step length, step time and walking speed
Stone et al. [256]	2015	Fall risk assessment	Kinect v1	depth data	average gait speed
Gianaria et al. [96]	2016	Frailty assessment	Kinect v2	skeleton data	TUG time, covered distance, walking time, walking speed, swing time, double support time and torso inclination angle
Phillips et al. [210]	2016	Fall risk assessment	Kinect	depth data	gait speed and stride length

Table 2.8: Summary of frailty detection and fall risk assessment through Kinect-based gait analysis.

Kinect-based approaches it is demonstrated, that the presented general structure for gait recognition also applies to Kinect-based gait recognition and gait analysis.

Moreover, approaches in gait recognition, medical gait analysis and gait-based frailty detection and fall risk assessment are reviewed. The majority of gait recognition and medical gait analysis approaches use skeleton data over depth data. Furthermore, both gait recognition and medical gait analysis commonly apply a classifier on the extracted gait parameters. However, this is not the case for frailty detection or fall risk assessment approaches. Due to the lack of a definition for frailty, there is no ground truth about the frailty status of subjects and therefore classification is ambiguous.

2.4 TUG analysis based on Kinect

Automatizing TUG is approached based on various sensor modalities, e.g. color cameras [29, 248], using wearable IMUs [116] or based on ambient sensors [85]. More information for automatizing TUG tests based on different sensor technologies can be found in the survey of Sprint et al. [251].

TUG automatization based on depth sensors has the advantage that TUG can be recorded unobtrusive, no additional equipment is required and there is no need for subjects to use body-worn sensors while performing the TUG test. The following results are deemed relevant based on a literature search for TUG automatization based on Kinect.

Vernon et al. [281] investigate the potential of Kinect v1 as a medical instrument for

TUG test analysis. The Kinect sensor is placed off-center of the walking area, facing the armchair. Due to their camera setup, part of the walking phase and the turning phase are not in the sensor field of view. Their method recognizes the following seven TUG events: Standing, first step, first stride, both times the participant was 2 m from the camera and the final sitting position. They identify start and end as the beginning and end of trunk joint movement. The standing event is found when the shoulder center joint reached peak height. First step and first stride are detected from ankle velocity. Based on these seven TUG events, a set of seven TUG parameters is extracted: Peak trunk flexion angle and peak trunk angular velocity during standing, first step length, first stride length, gait speed and turn time. Due to the lack of vision during the turn event, they roughly estimate turn time as the time between reaching the point 2 meters from the camera twice. Thirty individuals with stroke participate in their evaluation. Test-rest reliability is assessed using intraclass correlation coefficient, redundancy using Spearman's correlation and score prediction using multiple regression. Except trunk flexion angle, all of their Kinect-based TUG parameters are considered reliable. Most parameters are deemed redundant with TUG time, except first step length and trunk flexion angle. They conclude that utilizing the Kinect for TUG analysis provides additional information that may be predictive for changes in performance over time. The accuracy of the detected TUG event is not assessed in their work.

Kitsunezaki et al. [145] use Kinect to automatically measure TUG time and time for the 10 meter walk test. Their method measures TUG time based on the trajectories of the head and hip joint. Additionally, a cubic space is defined around the chair and conclusions are made about the pose and position of the subject based on the number of joints inside the cubic space. They experiment with three different positions for the Kinect sensor: in front of the chair, at the side of the chair and above the chair, looking down. Each position has some errors, but the position in front of the chair is chosen as the best. Their approach is evaluated using the stopwatch time from a physical rehabilitation expert as a gold standard. The average time difference between the stopwatch and their system is 0.33 seconds. Healthy adults are used for their evaluation.

Lohmann et al. [164] propose a TUG automatization method, sTUG, based on the skeleton data of two Kinect v1. Two Kinects are used to make sure the whole TUG area is covered. The two skeleton models are not merged, instead they simply pick either of the two skeleton models depending on the position of the subject. They use the trajectory of the shoulder center joint and its first and second derivative for the detection of TUG events. In addition to TUG time, the following 10 events are detected: Start moving, end uprising, start walking, start rotating, max turn, end rotating, start accelerating, end accelerating, end moving and start lowering. Their method is evaluated with 5 elderly and 4 healthy participants using a stopwatch and aTUG [85], an ambient sensor based TUG automatization approach. Detected TUG events match well with those from both aTUG and the manual video annotation. TUG time between their method and aTug differs consistently with approximately 1.5s difference. While aTug detects the start of the TUG test after the subject lifts from the chair, sTUG detects the forward motion

before uprising as the start.

Kargar et al. [138] analyze TUG test performance using a Kinect v1 placed at a height of 120 cm in front of the 12 elderly participants. They extract three features from the skeletal data: The total number of steps taken during the TUG test, the average duration of each step and the turn duration. For their TUG analysis, they recognize three different phases: Seated phase, walking phase, and turning phase. Seated phase is detected based on the distance between the hip joint and the camera. As soon as the person starts walking, the distance decreases and only reaches its constant maximum again after sitting down. Turning phase is estimated based on the absolute difference of the two elbow joints in x-direction (left and right from the sensor). After turning starts, the difference in x-direction decreases and reaches its minimum in the middle of turning. All remaining frames between seated phase and turning phase are considered walking phase. For their evaluation, a geriatric expert assesses the videos and categorizes subjects into potential fallers and non-fallers. The three extracted features are used to discriminate potential fallers from non-fallers using Bag of Words and SVM. The accuracy of the extracted TUG phases is not evaluated.

Cippitelli et al. [54] apply a sensor fusion approach to combine data from the skeleton model of Kinect v1 and an inertial measurement unit (IMU). They use their time synchronization method on the extraction of parameters of the TUG test. Data from Kinect is utilized for obtaining step lengths and cadence, the duration of the sit-to-stand (STS) phase, the duration of the back-to-sit (BTS) phase, the duration of the turning phase and the total tug time. Step length and cadence are estimated from peaks of the distance between both feet joints. The STS phase is detected based on the assumption that standing up requires to lean forward to put the center of mass over the feet. The movement of leaning forward is detected as a minimum followed by a maximum from the y-coordinate (height) of the head joint trajectory. The BTS phase is identified analogue by looking for a maximum followed by minimum followed by returning to default height. Turning phase is identified from shoulders and head joints. An orientation vector is computed from these three points and the angle between the orientation vector and the reference direction is used to estimate the beginning and the end of the turning phase. For their evaluation three trials are conducted by 20 healthy subjects and the mean and standard deviation for all obtained parameters were computed. However, the accuracy of their approach is not evaluated. Their recorded dataset is available online [11].

Table 2.9 shows a summary of the described TUG automatization approaches. It is observed that none of the previous approaches use Kinect v2 as a depth sensor and only one out of five approaches evaluate the error of individual TUG events. 3 out of 5 papers evaluate the results with a stopwatch and only one approach uses manual video annotation as a ground truth. All approaches use the skeleton model of the Kinect pose estimation. Four out of five approaches use the trajectories of specific joints, the distance between joints or the distance between a joint and the sensor as a feature for

[11] http://www.tlc.dii.univpm.it/blog/databases4kinect#IDTUG, Accessed 2017-07-12

Article	Participants	Data	TUG events	Feature	Study evaluation
Lohmann et al. [164]	4 healthy, 5 elderly	Kinect v1 skeleton	TUG time, 10 TUG events	velocity and acceleration of shoulder center joint; distance of left and right shoulder joints	mean difference of detected TUG time was -0.12 compared to -0.10 for stopwatch and 1.36 for aTUG [85]
Kitsunezaki et al. [145]	6 healthy adults	Kinect v1 skeleton	start and end (TUG time)	trajectories of head and hip; cubic box around chair	average TUG time error of 0.33 seconds compared to stopwatch
Kargar et al. [138]	12 elderly partici-pants	Kinect v1 skeleton	average step number, step duration and turn duration	distance between sensor and hip joint; distance of elbow joints	classification of fallers and non-fallers
Cippitelli et al. [54]	20 healthy subjects	Kinect v1 skeleton and IMU	TUG time and duration for STS, turning and BTS	trajectories of shoulder and head joints	mean and standard deviation of TUG events; dataset available online
Vernon et al. [281]	30 elderly with stroke	Kinect v1 skeleton	7 TUG events	trajectory of trunk joint and shoulder center joint	Spearman correlation of TUG time assessed with Kinect and stopwatch was 0.99; TUG parameters redundant; error of single TUG events not assessed

Table 2.9: Summary of TUG analysis approaches based on Kinect.

the extraction of TUG parameters. While Lohmann et al. [164] use the distance between two joints for the detection of turn events, all other TUG parameters are obtained by using the velocity or the acceleration of the shoulder center joint.

39454

CHAPTER 3

Methodology

This chapter describes the proposed approaches for gait analysis and TUG automatization. The first part of this chapter describes the techniques used to extract a gait signal from human walking and how spatiotemporal parameters are estimated from the gait signal and the person's centroid. This process is referred to as gait analysis. In this work, gait analysis is approached based on both skeleton and depth data obtained from a Kinect v2. The approaches are chosen from the review of related works from the previous section. The gait analysis approach based on depth data is based on the work of Stone et al. [255] and Hotrabhavananda et al. [121]. The approach based on skeleton data is based on the work of Clark et al. [56] and Gianaria et al. [96]. Adaptions are made due to differences in the dataset. The exact methodology used is described as follows.

The second part of this chapter describes the proposed approach for automatically analyzing the TUG test based on Kinect v2. Both skeleton and depth data of Kinect v2 are utilized for estimating the time needed to complete the TUG test and for the detection of six TUG phases. The proposed approach for depth data and skeleton data is based on sTUG, proposed by Lohmann et al. [164] and discussed previously in the related works. The proposed approach differs from sTUG, since the center shoulder velocity is used instead of the center shoulder acceleration. The exact methodology for the TUG automatization approach is described in this chapter.

For both the gait analysis and the TUG automatization approach, an assumption is made that there is only one person in the scene at any time and that the background environment is stationary. In order to be deployed in a dynamic real-world environment a robust multiple-person tracking approach and an update procedure for handling a dynamic setting are required.

(a) Examples for depth frames of different walking sequences.

(b) Depth frame samples depicting older people executing a TUG test.

3.1 Gait Analysis

The goal of the gait analysis approach is to extract a set of spatiotemporal parameters. It is shown in the related works that the extraction of spatiotemporal parameters of a person's walk allows to draw conclusions about the person's health status. In this work, gait speed, step length, stride length, step time and stride time are estimated from walking sequences recorded with a depth sensor. The 5 spatiotemporal parameters are extracted from both Kinect depth data and skeleton data. The approach based on depth data is described first.

3.1.1 Depth data

Figure 3.1a and 3.1b show examples for depth data frames in two different scenes. Darker pixels represent image points close to the sensor and lighter pixels represent image points farther away from the depth sensor. White pixels represent image points with no depth measurement. These pixels occur commonly in case of shiny object materials or around the border of the image.

Background model estimation

In related works, a background model is established from a set of background training images captured prior to recording the subject [255, 55, 121]. In the cited works, outliers from the background model are considered as foreground. However, no prior background frames are recorded for the datasets acquired in this work. Therefore the approaches used in the cited works are not valid for the recorded dataset and a different approach is required to obtain a background model.

Related work Piccardi [211] reviews several background subtraction methods including Running Gaussian Average, Temporal Median Filter, Mixture of Gaussians, Kernel density estimation, etc. Only single-valued background models are considered as there are no outdoor scenes which have constant background motion from wind, rain, etc. Single-

valued background extraction techniques such as running Gaussian average (RGA) and temporal median filter (TMF) provide sufficient accuracy for the purpose of this work. Their disadvantage is that they perform poorly in outdoor scenes with constant motion and they do not consider spatial correlations.

Comparing RGA and TMF Modified versions of running Gaussian average (RGA) and temporal median filter (TMF) are implemented. Typically, these methods update their background model based on the last n frames [211]. However, this is not suitable for short image sequences that require successful background subtraction starting with the first frame. Instead the whole image sequence Ω consisting of N images $I_k, k \in \{1, ..., N\}$ is considered for the computation of the background model and the background model is not updated for the successive frames. While this substantially increases computational effort and memory requirements, it allows utilization of later frames for the silhouette extraction of earlier frames. For longer image sequences, the image sequence Ω can be subsampled to decrease memory requirements, while still obtaining an adequate background model [61].

Running Gaussian Average Standard RGA computes μ_t using α as a weight that controls update speed [296].

$$\mu_t = \alpha I_t + (1 - \alpha)\mu_{t-1} \tag{3.1}$$

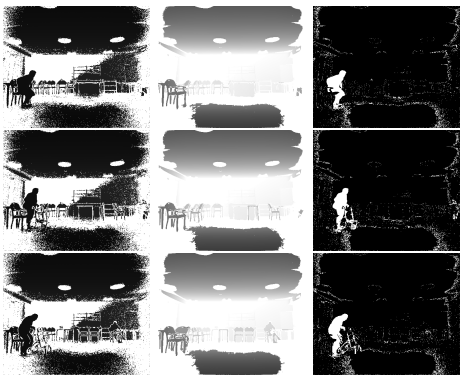
In the modified RGA μ_Ω and σ_Ω are simply the mean and standard deviation in the temporal dimension with each frame of Ω being weighed equally. Furthermore, the binary foreground image F_k is obtained from input image I_k by the inequality

$$F_k = \mu_\Omega - I_k > k; \sigma_\Omega \tag{3.2}$$

The term $|I_k - \mu_\Omega|$ in the original RGA is replaced with $\mu_\Omega - I_k$ as pixels are only considered foreground if they are closer to the camera than the background. Pixel with no depth measurement (see section 3.1.1) are ignored for the calculation of μ_Ω and σ_Ω and excluded for the calculation of the normalization factor $\frac{1}{N}$. If the number of pixels with no depth measurement in the temporal dimension exceedes the threshold $\frac{3N}{4}$, this particular pixel of μ_Ω is treated as ∞.

Temporal Median Filter Lo and Velastin [163] propose the idea of using the last n frames as a background model. Similar to the modified RGA, a modified TMF is applied to the whole image sequence Ω. Pixels with no measured depth are treated as ∞ for the purpose of median calculation. The calculation of the median is adapted to avoid artifacts caused by the mean of a scalar depth value and pixels with no depth measurements, which are treated as ∞ (see Figure 3.3).

(a) Foreground extraction based on running Gaussian average. From left to right: Input frame I_k, mean μ_Ω, standard deviation σ_Ω, foreground F_k



(b) Foreground extraction based on temporal median filter. Left to right: Input image I_k, median $\tilde{x}_\Omega$, foreground F_k

Figure 3.2: Comparison of running Gaussian average and temporal median filter for foreground extraction in regions where the subjects stays stationary over a significant amount of frames.

$$\tilde{x}(i,j) = \begin{cases} x_{\frac{N+1}{2}}(i,j), & N \text{ uneven} \\ x_{\frac{N}{2}}(i,j), & N \text{ even} \end{cases} \tag{3.3}$$

$\tilde{x}(i,j)$ represents the median of the set of corresponding pixels $x_k(i,j), k \in \{1,...,N\}$, which are sorted by their depth value. Foreground F_k of an input frame I_k is extracted by the inequality

$$F_k = \tilde{x}_\Omega - I_k > T \tag{3.4}$$

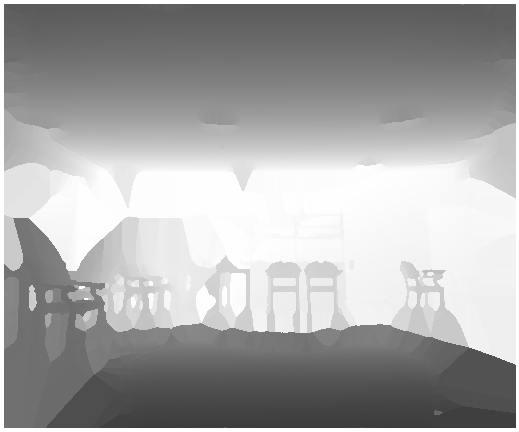
The depth threshold $T = 0.2$ m is used in this work. No-measured depth pixels (∞ values) are removed in a later step (see section 3.1.1).

Temporal median filter is chosen over running Gaussian average as it provides a more robust silhouette in regions where the subjects remains for a relatively long time. This is especially critical for image sequences of the TUG test, as elderly subjects spend a significant amount of time near the chair for standing up and sitting down. Figure 3.2 shows a few examples where RGA shows ghosting artifacts in the area near the chair but TMF does not. However, it should be noted that TMF shows the same ghost artifacts if the subjects spends 50% or more time in an area.

Depth hole filling

In order to obtain a robust silhouette of the subject, pixels on the background model $\tilde{x}_\Omega$ with no measured depth value are filled. Since pixels with no depth measurement are

(a) Artifacts occur when the amount of pixels with successful depth measurement and no depth measurement (white pixels) is equal in the temporal dimension.



(b) Estimated background model using the adapted median formula of equation 3.3.

Figure 3.3: Final estimated background model using TMF with standard median formula (left image) and adapted median formula (right image).

treated as ∞ in the TMF background model $\tilde{x}_\Omega$, any measured depth value for these pixels would be closer to the sensor and therefore be detected as foreground. Moreover, all pixels with no measured depth value that separate the subject's limbs are filled on the input frame I_k in order to avoid separated limbs.

Related work Signal filtering is a preprocessing step which can be used not just for removing noise but also for recovering missing depth information (holes) [111, 28]. Naive approaches use existing image filters used for monochromatic images such as Gaussian filters, average filters or median filters. While they are suitable to remove noise, they are not ideal for filling depth holes. Another drawback is that they only work for regions where the underlying depth data is in favor of the filter [111]. More sophisticated methods can be used in order to remove holes from a sequences of depth images.

Berdnikov and Vatolin [26] identify two characteristics of holes in depth data and their algorithm picks a filling scheme accordingly. While it works better than the naive approach, it does not consider additional information such as temporal information and depth and color values. Camplani and Saglado [44] apply a joint bilateral filter to fill holes in depth images. A bilateral filter has the advantage of preserving edges, due to weighing neighboring pixels based on both spatial and depth value distance. More specific, they use temporal information to iteratively a depth map model D_m, which is a consistent depth map of the scene. The consistency of each pixel in the depth map is stored in a separate consistency map C_{map}. Missing pixels in a new depth frame are replaced by their corresponding pixel in D_m if they are considered reliable enough based on its corresponding value in C_{map}. Then a joint bilateral filter is applied on the

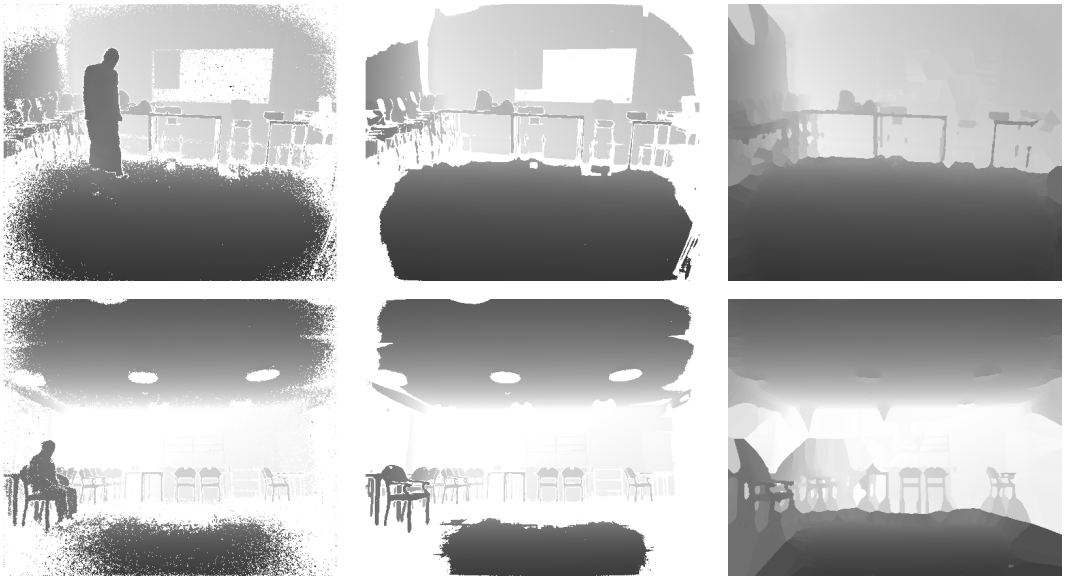


Figure 3.4: Left image: Single frame I_k. Center image: median $\tilde{x}_\Omega$. Right image: Final background model $\tilde{x}'_\Omega$ after applying hole-filling.

depth frame using both the pixel similarity in D_m and the similarity in the intensity domain. Missing pixels in D_m are filled over time based on reliable neighborhood pixels or corresponding non-missing pixels from new depth frames.

Several papers [176, 238, 221] consider the relation of RGB image and depth image and apply a sensor fusion approach for filling depth holes. Milani and Calvagno [176] detect segments in the RGB data and interpolated missing depth values based on the assumption that the transition of depth values is smooth within each segment. Schmeing and Jiang [238] improve the depth image using a superpixel segmentation algorithm to replace corrupted edges with correct ones. Qi et al. [221] adopt an RGB inpainting approach based on information fusion for hole filling in depth maps. Since depth images lack the texture information to detect object boundaries, boundaries are detected on the corresponding color image and their position is projected onto the depth image.

Hole-filling of the background model The discussed techniques also have their disadvantages. For example, sensor fusion techniques need a secondary data input, e.g. RGB or stereo, which is not available for the recorded data. The joint bilateral filtering approach [44] does not guarantee to fill all depth holes in a limited number of frames. Especially in corner regions their depth hole filling technique proceeds slowly. In this work Euclidean distance is used to replace all no-measured depth pixel in $\tilde{x}_\Omega$ with their nearest non-hole pixel. Figure 3.4 shows two examples of hole filling applied on $\tilde{x}_\Omega$ to obtain the final background model $\tilde{x}'_\Omega$.

Figure 3.5: Extracting silhouette S_k from frame I_k. From left to right: Original frame I_k, median filtered frame $\tilde{I}_k$, difference $\tilde{x}'_\Omega - \tilde{I}_k$, thresholding $\tilde{x}'_\Omega - \tilde{I}_k > T$, resulting silhouette S_k.

Silhouette Extraction

The depth holes in areas where body limbs, e.g. legs and trunk, are separated by a line of no-measured depth pixels, are filled by applying a 2D median filter with a 7×7 neighborhood. Depth hole filling in the remaining area is not necessary as only the silhouette is used for further processing. After applying the median filter, foreground extraction is applied based on the modified TMF of equation 3.4. The silhouette of the subject is then obtained as the blob with the largest number of foreground pixels. Identifying the largest blob as the main subject is a simple, but effective way to remove small blobs caused by noise. Extracting the biggest blob as the foreground object is applied in previous works, e.g. [150, 14].

Algorithm 3.1 shows pseudo code for the extraction of silhouettes $S_k, k \in \{1, ..., N\}$ from image sequence Ω. This algorithm works under the assumption that there is exactly one person in front of the depth camera during the whole video sequence. This is a reasonable assumption, which is applied in previous works, e.g. [150, 255]. However, extending this approach to detect silhouettes of multiple subjects is possible. Figure 3.5 shows examples of extracted silhouettes using algorithm 3.1.

Algorithm 3.1: Pseudo code for the extraction of silhouettes S_k from the image sequence $I_k, k \in \{1, ..., N\}$.

Input: A sequence of depth images $I_1, I_2, .., I_N$
Output: Binary silhouettes $S_1, S_2, ..., S_N$

```
Estimate background median x̃_Ω based on TMF;
Remove depth holes from background median x̃_Ω to obtain final background model x̃'_Ω;
for k ← 1 to N do
    Apply median filter to frame I_k to remove noise and minor holes between limbs;
    Extract foreground F_k from frame I_k and background model x̃'_Ω;
    Extract largest blob from F_k as the silhouette of the subject;
end
```

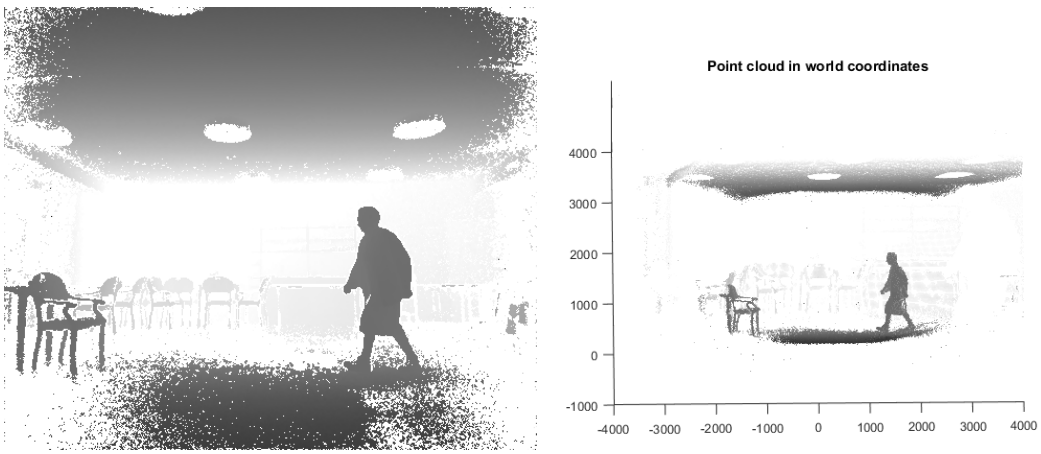


Figure 3.6: Left image: Original input frame. Right image: Resulting point cloud

Point cloud conversion

Since the silhouette does not have any 3D information of the subject, it has to be converted into a point cloud in order to determine the position and the walking posture. The point cloud is first obtained in camera coordinates and then transformed into world coordinates.

Transformation to camera coordinates Transforming the depth image I_k from image coordinates to camera coordinates creates a 2.5-D point cloud representation of the scene. Coordinates can be transformed from image coordinates (u, v) to camera coordinates (X_c, Y_c, Z_c) using intrinsic camera parameters [83].

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix} = \begin{bmatrix} \frac{(u-O_x)\ I_z}{f_x} \\ \frac{(v-O_y)\ I_z}{f_y} \\ I_z \end{bmatrix} \tag{3.5}$$

O_x, O_y denote the coordinates of the principal point. f_x, f_y denote the focal length in x and y direction and I_z is the depth value of pixel (u, v). $P_c = \{X_C, Y_C, Z_C\}$ denotes the corresponding point in camera coordinates. Figure 3.6 shows the point cloud conversion of a depth frame based on equation 3.5.

Transformation to world coordinates A transformation from camera coordinates to world coordinates is obtained by [83]

$$P_w = \begin{pmatrix} \mathcal{R} & t \\ 0^\intercal & 1 \end{pmatrix} P_c \tag{3.6}$$

P_w and P_c are points of homogeneous coordinates in world coordinates (P_w) and camera coordinates (P_c). $\mathcal{R}$ denotes the rotation matrix from camera to world coordinates and t is a translation vector. $\mathcal{R}$ and t have been estimated based on the position and inclination of the floor plane. The RANSAC variant MSAC [270] is used to estimate the floor plane $\mathcal{P}_{\text{floor}}$ from the point cloud obtained from the intrinsic parameter transformation step. Given the estimated floor plane $\mathcal{P}_{\text{floor}}$, $\mathcal{R}$ can be directly computed as

$$\mathcal{R} = I + [\vec{v}]_\times + [\vec{v}]_\times^2 \frac{1 - \vec{a} \cdot \vec{b}}{\|\vec{v}\|^2} \tag{3.7}$$

where I is the 3×3 identity matrix, $\vec{a}$ is the normal vector of the estimated floor plane in camera coordinates, $\vec{b} = (0, 1, 0)^\intercal$ is the normal vector of the floor plane in world coordinates, $v = \vec{a} \times \vec{b}$ and $[v]_\times$ is the skew-symmetric cross-product matrix of $\vec{v}$, which is defined as

$$[\vec{v}]_\times = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix}$$

Equation 3.7 is valid for all unit vectors $\vec{a}, \vec{b}$ except $\cos(\angle(\vec{a}, \vec{b})) = -1$. The case that $\vec{a}$ and $\vec{b}$ point in opposite directions is avoided by always picking the floor plane normal vector $\vec{a}$ with a positive y-coordinate (vertical coordinate $a_2 > 0$).

Alternatively, $\mathcal{R}$ can also be obtained from the three Euler angles (ϕ, θ, ψ) of the set of extrinsic coordinates, e.g. by $\phi = \arccos a_2$, $\theta = 0$ and $\psi = \arcsin a_1$ with $\vec{a} = (a_1, a_2, a_3)$ being the normal vector of the estimated floor plane $\mathcal{P}_{\text{floor}}$. It should be noted that the value of θ (rotation among the vertical vector) is independent of $\vec{a}$ and therefore ambiguous. A second plane, e.g. the plane of the opposite wall, would have to be estimated to determine θ.

Footprint extraction A point cloud representation of the subject in the scene is obtained by transforming all pixels of input frame I_k that are part of its corresponding silhouette S_k from image coordinates into world coordinates. The feet of the subject can be extracted by applying a vertical threshold T_{vert} on the point cloud of the subject. $T_{\text{vert}} = 0.5$ meter is used in this work. As pointed out by Stone and Skubic [255], the threshold needs to be large enough to obtain enough information when the subject is far from the sensor. Moreover, points that are very close to the ground are considered background during the background extraction algorithm (see section 3.1.1).

Footprint are obtained by projecting the remaining feet point cloud onto the ground plane. This can be achieved by simply cutting the vertical component of each point. The remaining points, which are above T_{vert}, are used to compute the subject's centroid. Figure 3.7 shows an overview of all previously described steps for obtaining footprints from a series of depth images.

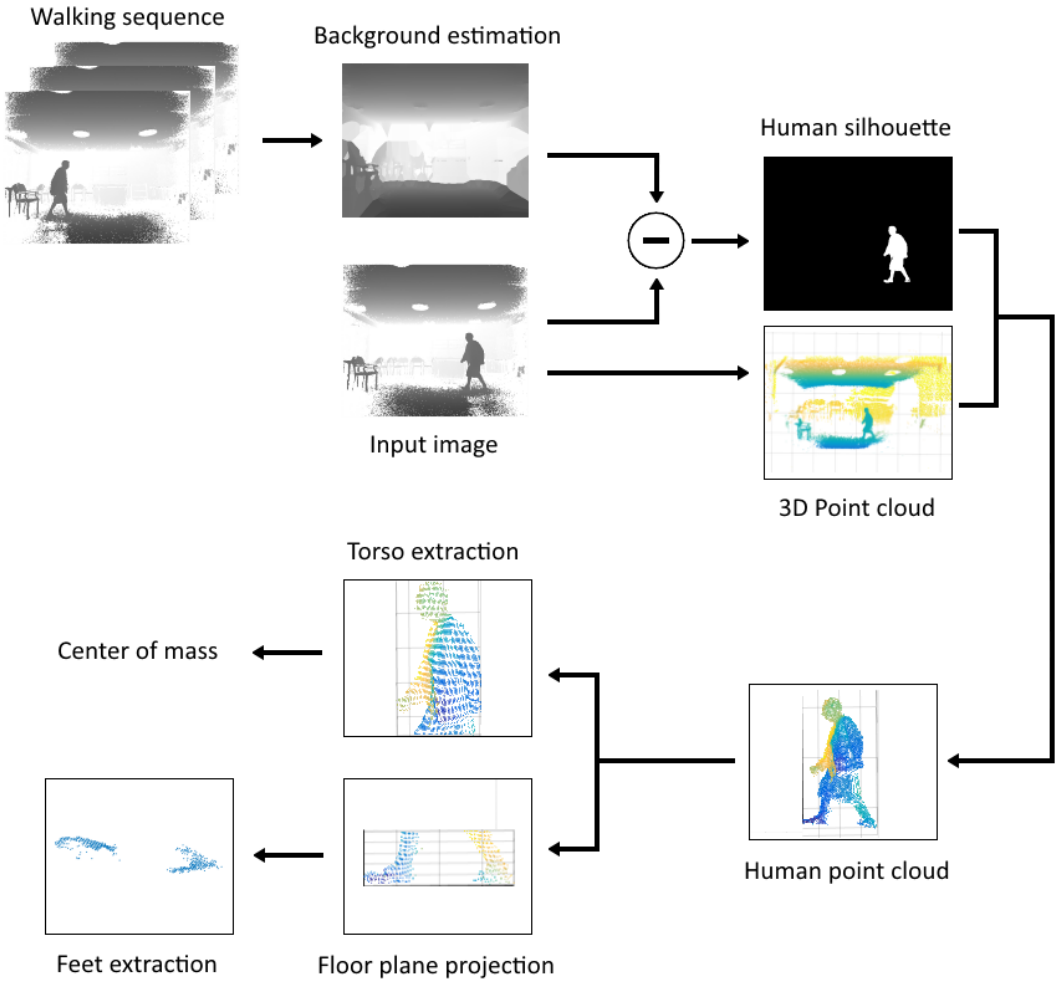


Figure 3.7: Overview of performed steps to obtain the subject's centroid and footprints from a depth image sequence.

3.1.2 Gait signals

The time series of a subject's footprints is used to compute several periodic gait signals. Four different gait signals for depth data and and one additional gait signal for skeleton data are computed and evaluated in this work. The extracted gait signals are based on correlation coefficient, distance of the feet, horizontal and vertical oscillation and ankle joint velocity.

Correlation coefficient time series

Extracting footsteps from the correlation coefficient time series (CCTS) is originally proposed in [255] and later also used in [121]. The idea is that the correlation coefficient

reaches its peak during a double support phase and reaches zero when the feet are parallel.

First, the projected footprint is normalized regarding position and rotation. This is done by subtracting the mean and rotating the footprint depending on the walking direction. The resulting normalized footprint has its centroid at the origin and is walking towards the positive z-axis. Then the following correlation coefficient is computed for each frame from the normalized ground-projected footprint.

$$p = \frac{\sum_{n=1}^{N} x_n y_n}{N} \tag{3.8}$$

x_n and y_n denote the coordinates of the $n^{\text{th}}, n \in \{1, ..., N\}$ point of the ground-projected point cloud of the subject's feet. The correlation coefficient of equation 3.8 is computed for each frame of a walking sequence to obtain the CCTS. As illustrated in Figure 3.8, peaks of the CCTS correspond to heel strikes and double support phases, while roots correspond to midstance phases. In the work of [255, 121], CCTS is filtered using a median filter and a moving average filter of a variable window size, which depends on the frame rate and the walking speed of the subject. Footsteps are then detected as minima and maxima of CCTS.

Feet distance time series

The idea of using the feet distance time series is a commonly used approach as it is applied in the work of Hotrabhavananda et al. [121], Prochazka et al. [219] and Cippitelli et al. [54] for the extraction of spatiotemporal parameter from Kinect skeleton data. Moreover, Auvinet et al. [10] detect double support from the anterior-posterior knee distance and Bobick and Johnson [32] consider the distance of the feet during the double support phase as a feature for gait recognition.

In this work, the positions of the two feet are estimated from the projected point cloud of the feet using unsupervised clustering. K-means clustering with 2 clusters is used to estimate the centroids C_1 and C_2 for both feet. Figure 3.9 shows an example of K-means clustering applied on the data of a single footstep. The distance of the feet is then estimated as the Euclidean distance of cluster C_1 and C_2.

$$d_{\text{feet}} = \|C_1 - C_2\| \tag{3.9}$$

Peaks of the feet distance time series relate to heel strikes and double support phases. Troughs relate to midstance phases, when both feet are the closest. This is illustrated in Figure 3.10.

Horizontal and vertical oscillation

Various spatial displacements such as horizontal and vertical oscillation can be observed during gait [190]. While one foot is in the air, the walking person puts his center of mass

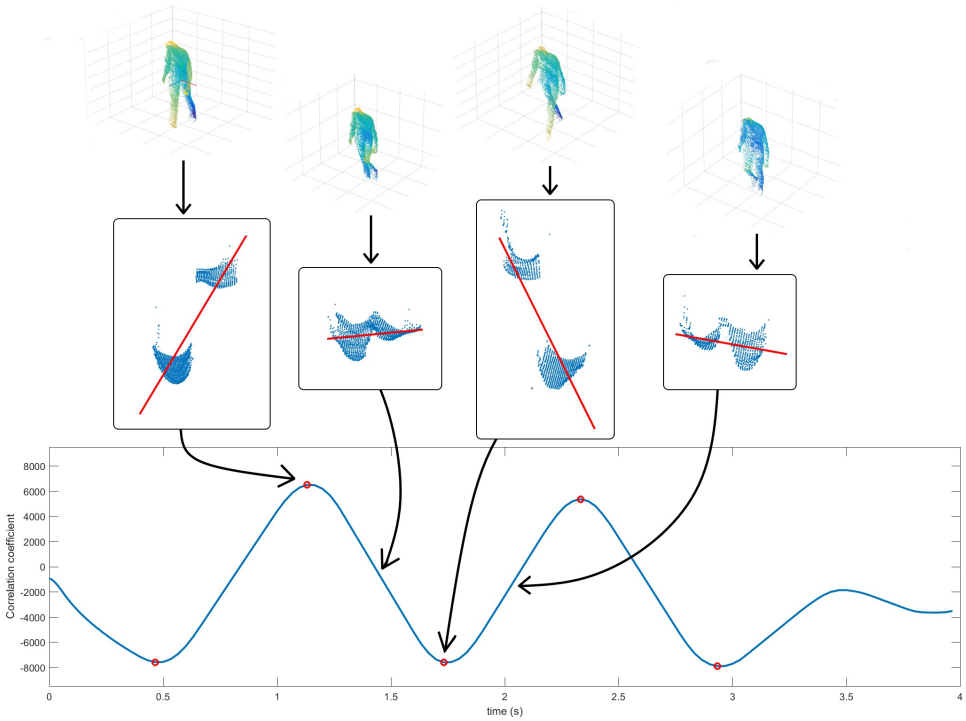


Figure 3.8: From point cloud of walking subject to CCTS: Upper row: Point cloud of silhouette. Center row: Ground-projected footprint (red line indicates correlation coefficient). Bottom row: filtered CCTS

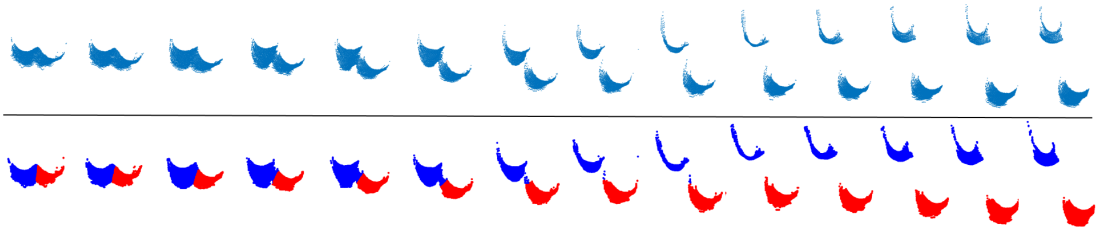


Figure 3.9: Extracting left and right foot positions using K-means clustering during a single footstep. Top row: ground-projected point cloud of the feet. Bottom row: Separation of feet trough K-means clustering.

on the standing foot in order to balance. Alternatively balancing on each foot creates a horizontal oscillation. The effect of vertical oscillation is caused by lowering the trunk during double support phase and raising the trunk during midstance phase in order to balance the weight. Both horizontal and vertical displacement is computed based on the trajectory of the center of mass of the subject's point cloud.

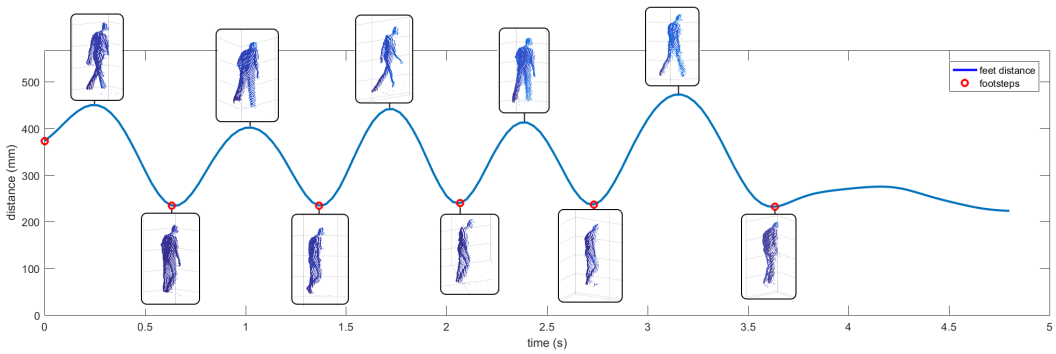


Figure 3.10: Feet distance: Peaks correspond to double support phases and troughs correspond to midstance phases.

Horizontal Oscillation The horizontal oscillation (HO) is computed by measuring the horizontal displacement during a straight walk. The distance from the ground-projected centroid C to the line with starting point P and end point Q is obtained by projecting $\vec{CP}$ on the normal vector of line $\vec{PQ}$.

$$d_{\text{HO}} = \frac{\vec{n} \cdot \vec{r}}{\|n\|} \tag{3.10}$$

where $\vec{n}$ is the (unnormalized) normal vector of line $\vec{PQ}$ and $r = \vec{CP}$ is a vector from point C to the first point of the line P. Figure 3.11 shows an example for the HO time series obtained from a walking sequence. It can be seen that peaks and troughs correspond to midstance phases of the respective foot and turning points correspond to double support phases. Peaks and troughs of the HO time series are used to estimate footstep timings and positions.

Vertical Oscillation Vertical oscillation (VO) describes the vertical displacement due to the pelvis moving up and down during walking. The vertical displacement is simply the vertical coordinates of the centroid c_y normalized by the mean vertical coordinate μ_{c_y}. Peaks of the VO time series correspond to midstance phases and troughs correspond to double support phases (see Figure 3.12).

Signal filtering

All gait signals are filtered using a median filter with the window size proposed by Stone and Skubic [255].

$$w = \frac{f * k}{v} \tag{3.11}$$

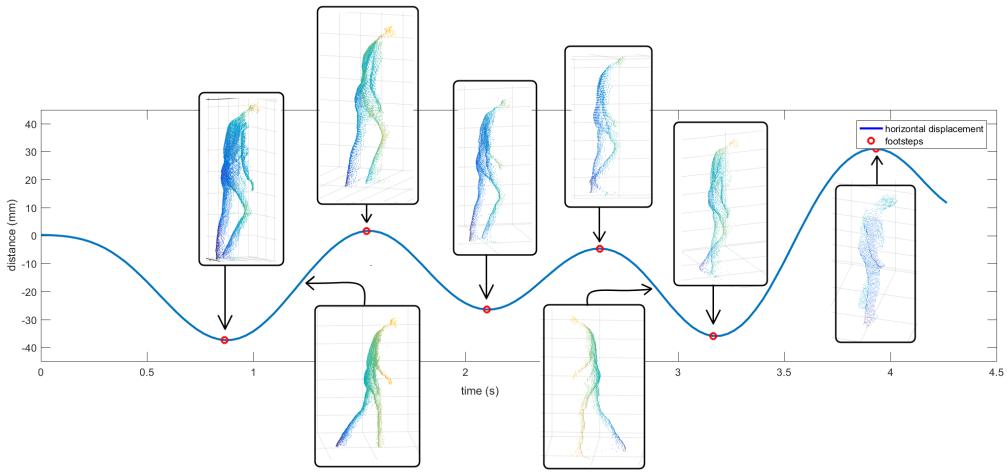


Figure 3.11: Horizontal oscillation during a walking sequence: Peaks and troughs correspond to midstance phases. Turning points correspond to double support phases since the weight is equally distributed when both feet are on the ground.

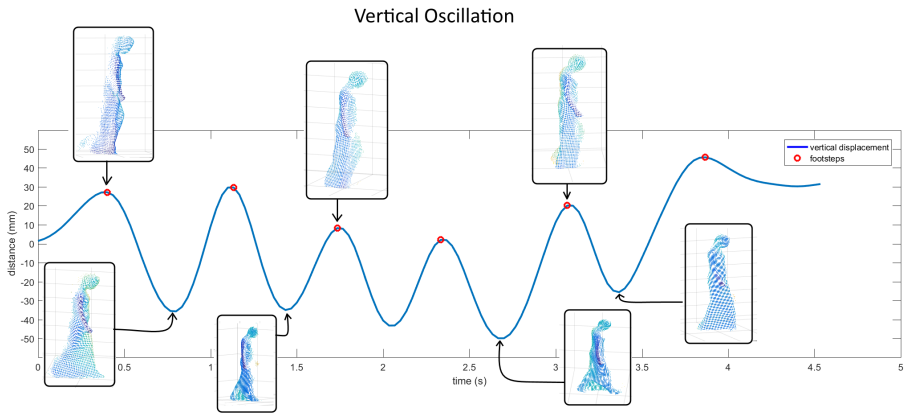
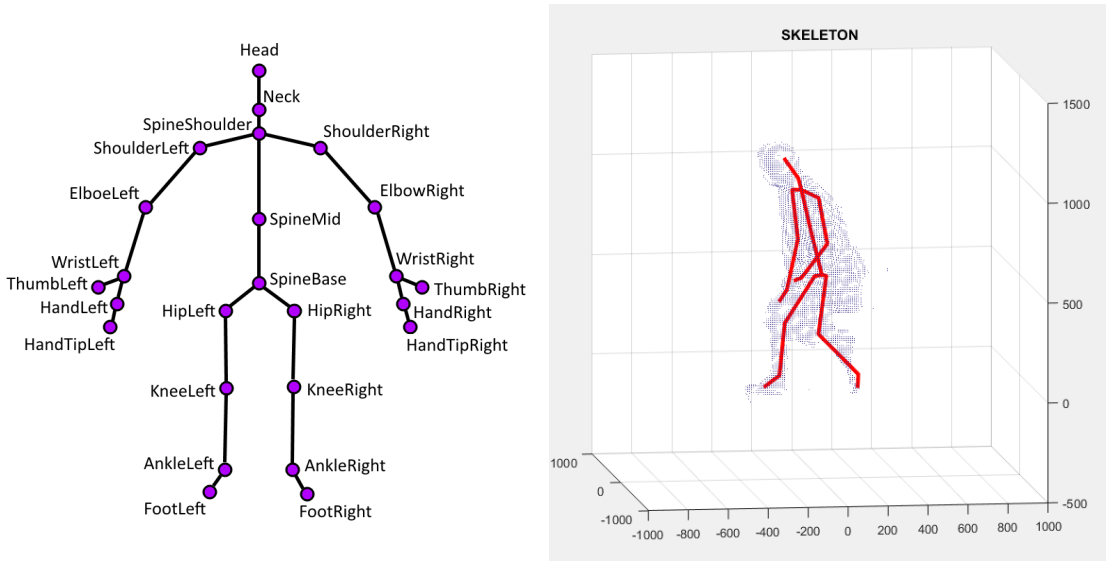


Figure 3.12: Vertical oscillation: Peaks correspond to midstance phases and troughs correspond to double support phases.

where v is walking speed in meters per second, f is frame rate and k is a constant parameter. The authors use $k = 16.6$ and inches per second as a metric for walking speed, therefore $k = 0.4216$ is used to adjust for the different metric. The authors point out, that k could potentially be adjusted based on the subject's height, but this is not considered as it would only have a minor effect.

After applying the median filter, a Gaussian filter is applied with $\sigma_{CC} = \frac{w}{2}$ for filtering the correlation coefficient signal. For all other gait signals, a smaller value $\sigma = \frac{w}{4}$ is used, as these signals have a doubled frequency. For CCTS, two consecutive peaks represent



(a) Kinect v2 skeleton model.

(b) Skeleton model and point cloud.

Figure 3.13: Left image: Joints of the Kinect v2 skeleton model. Right image: Skeleton model and point cloud of corresponding frame.

one stride, but only a single step for feet distance or vertical velocity.

3.1.3 Skeleton Data

Kinect v2 SDK provides pose estimation for a skeleton model that consists of the trajectories of 25 body joints. Kinect v2 skeleton model is illustrated in Figure 3.13a, while Figure 3.13b shows the skeleton model with its corresponding point cloud. Skeleton joints are filtered using a Gaussian filter with the same standard deviation σ_f as used for depth data.

The same gait signals based on correlation coefficient, feet distance, vertical and horizontal oscillation have also been extracted from the skeleton data. Similar to the work of Gabel et al. [88], the center of mass is computed as the mean of joints that belong to the human trunk, including spine, shoulder and hip joints. Correlation coefficient is estimated based on the inclination of the ground-projected positions of both ankle joints. As with depth data, the position is previously normalized by subtracting the mean followed by a rotation based on the walking direction. Feet distance is computed as the Euclidean distance of both ankle joints. Horizontal and vertical oscillations are estimated based on the trajectory of the center of mass.

Ankle Velocity

An additional gait signal based on ankle velocity is extracted from the skeleton data. This approach is similar to Clark et al. [56] and Gianaria et al. [96], who use velocity

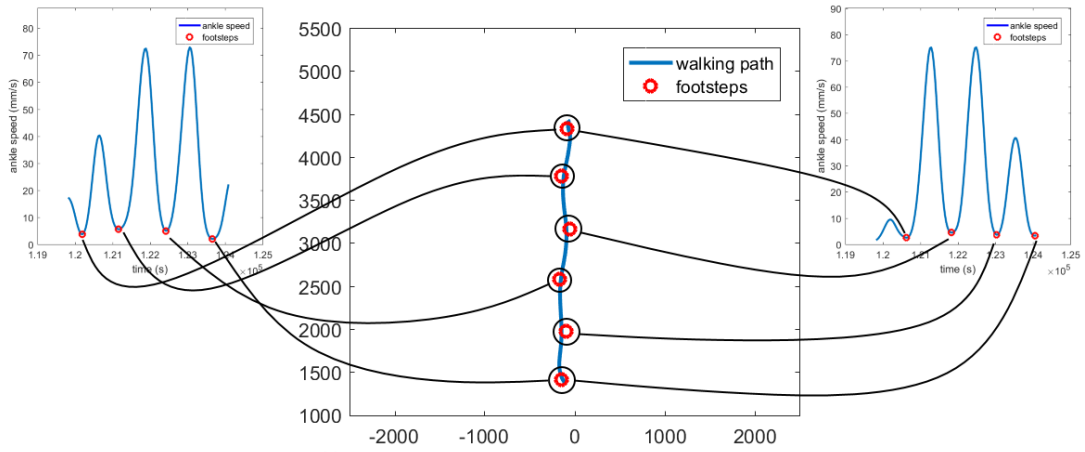


Figure 3.14: Merging footsteps detected on either left or right ankle velocity based on spatial proximity. Here the start and end position of the walking sequence are merged since they fall in the same area.

thresholds for the feet trajectories to extract footstep events. In the present thesis, ankle velocity for both ankles is determined as the derivative of the ground-projected ankle trajectory. Peaks correspond to swing phases and troughs correspond to midstance phases of the respective foot. The time when both feet velocities are low indicates double support time. Footsteps are detected from both left and right ankle velocity and get merged if the Euclidean distance of their locations is within a small threshold. Detecting the same footstep from both left and right signal occurs commonly on start and end positions as well as in situations when the stationary foot occludes the movement of the foot which is in swing phase. Figure 3.14 illustrates the extraction of footstep locations from both ankle velocities.

Signal gap filling

When the Kinect pose estimation fails no skeleton information is provided for the specific frame. This causes gaps in the trajectories and therefore also in the extracted gait signals. Small gaps in the signal are filled using spline interpolation. Cubic spline interpolation is used over other interpolation methods such as piecewise cubic hermite interpolating polynomial (PCHIP) or linear interpolation since the underlying signal is oscillatory. Figure 3.15 shows a comparison of the three mentioned interpolation techniques for filling signal gaps. Gaps smaller than 1.5 seconds are interpolated, while larger gaps are not changed. If more than 50% of the skeleton information of a walking sequence is missing the walking sequence is skipped.

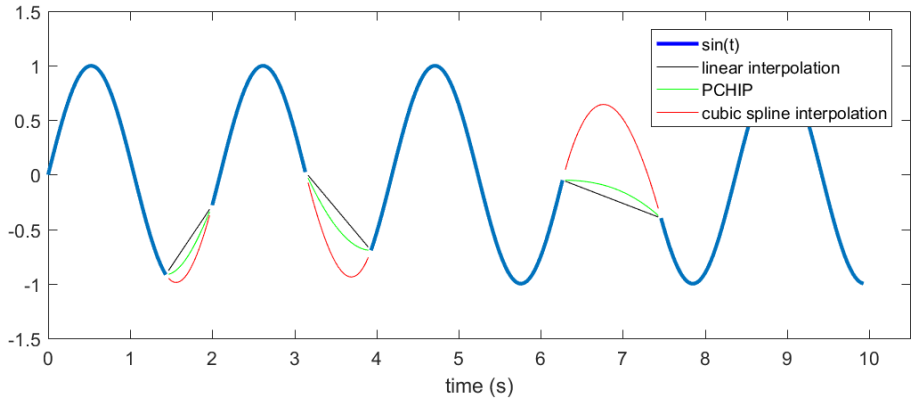


Figure 3.15: Comparison of linear interpolation, piecewise cubic Hermite interpolating polynomial (PCHIP) and cubic spline interpolation for fillings gaps caused by failed skeleton model extractions. Unlike the former two, cubic spline interpolation preserves the oscillating nature of gait.

3.1.4 Extraction of spatiotemporal parameters

Keypoints of gait signals such as minima, maxima, inflection points or roots can be used to extract time and spatial occurrence of regular gait events such as heel strikes, midstances, etc. Inflection points have the disadvantage that they occur commonly with minor curvature changes and are therefore not sufficiently stable. Roots have the disadvantage that a minor shift in amplitude towards the positive or negative side of the y-axis causes significant errors for the footstep detection. Therefore only minima and maxima are used for the detection of footsteps for all gait events.

Spatial gait parameters are obtained from the temporal occurrence of footsteps combined with the spatial trajectory of the movement of the ground-projected point cloud centroid. Step length is measured as the distance between two consecutive steps projected onto the walking direction in order to avoid inaccurately measuring lateral foot spacing as a part of step length. Stride length is measured as the distance between two consecutive footsteps of the same foot. Step times and stride times are measured based on the corresponding time stamps of detected footsteps. Gait speed is measured as distance traveled divided by time elapsed.

3.2 TUG Automatization

The goal of the proposed TUG automatization approach is to extract TUG time and the start and end of six TUG phases without any user input. The following assumptions have been made regarding the setup of the TUG and the sensor. Kinect is placed perpendicular to the 3 meter walkway on a table with medium height. The distance from the walkway to the Kinect is smaller than the maximum viewing range of the

sensor. Figure 3.16 shows a schematic placement of Kinect in a scene based on the two assumptions. While the following approach is described based on the assumption that Kinect is placed perpendicular to the walking direction of the subject, it can also be applied for a different setup by updating the coordinates based on the walking direction. A preprocessing routine that detects the initial setup is required in this case.

The six TUG phases that are detected from the described approach are chair rise C^1, walk #1 W^1, turn #1 T^1, walk #2 W^2, turn #2 T^2 and sit down C^2. The start and end event of each TUG phase are detected for a total of twelve detected TUG events. It should be noted that it is possible that these events overlap, e.g. turning and sitting down may be performed simultaneously by a particular subject. Moreover, there may also be a significant pause between two events, e.g. a subject may look back at the chair for several seconds to prepare for sitting down after turning.

TUG Automatization approach In order to detect twelve TUG events two generic functions have been defined. The first one, $\mathcal{G}(f(x), I_s, I_e)$, returns the beginning and the end of the longest consecutive occurrence of the logical condition $f(x)$ being true in the interval $[I_s, I_e]$. The second function $\mathcal{F}(f(x), n, I_s, I_e)$ returns the beginning and the end of the highest n peaks within the interval $[I_s, I_e]$. Unlike sTUG [164], events are not detected in the order they take place. Instead, events which are less challenging to detect, e.g. walking, are detected first followed by more challenging events, e.g. chair rise and sit down or turning. This has the advantage that the interval of occurrence can be narrowed down based on the information of previously detected events.

Skeleton data The derivative (velocity) of the trajectory of the spine shoulder joint ss' in x and y direction (ss'_x and ss'_y) is used for the detection of the walking, chair rise and sit down events. For turning, the distance between the left and the right shoulder joint in x direction $sd_x = |sl_x - sr_x|$ is used. While shoulders are close to parallel during walking, their distance in x direction (parallel to walking direction) reaches a maximum in the middle of turning. A Gaussian filter with $\sigma = 5.0$ is applied to both ss and its derivative ss'. A Median filter with window size 15 followed by a Gaussian filter with

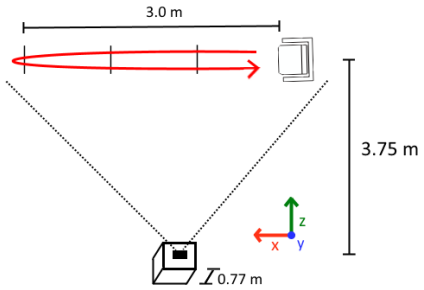


Figure 3.16: Scheme of the TUG recording scene; Kinect is placed perpendicular to the walking direction. *red:* x-axis, *blue:* y-axis, *green:* z-axis.

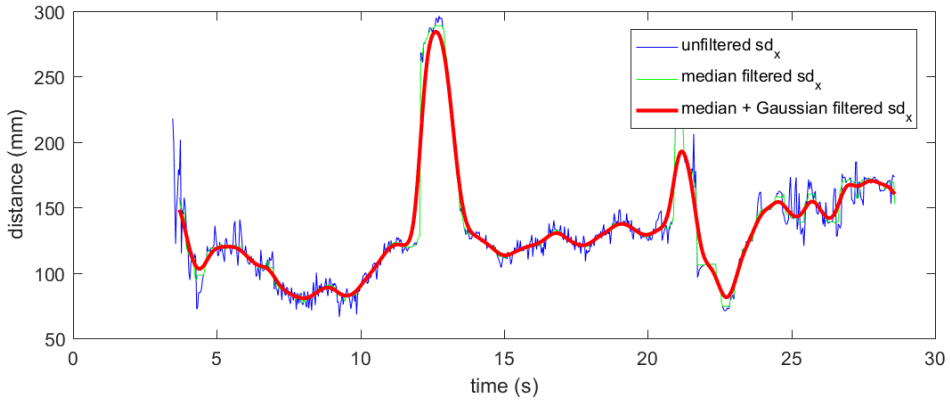


Figure 3.17: Applying a median filter and a Gaussian filter on sd_x; *blue:* unfiltered, *green:* median filter, *red:* median filter and Gaussian filter.

$\sigma = 5$ are applied to sd_x before the extraction of turn events. The median filter is used to remove outliers caused by noise from either shoulder joint (see Figure 3.17). The start and end of the following three TUG phases are detected based on the following criteria:

- *walking:* The first walking event W^1 takes place after rising from the chair. It is caused by a steady movement away from the chair in x-direction. It is detected by searching for $W^1 = \mathcal{G}(ss'_x > T, ss_{\text{start}}, ss_{\text{end}})$ with $T = \frac{1}{2}\sigma_{ss'_x}$. $\sigma_{ss'_x}$ denotes the standard deviation of the derivative of ss in x-direction. During the second walking event W^2 the subject returns to the chair. Similar to the first walking event, it is caused by a steady movement towards the chair in x-direction. It is detected by searching for $W^2 = \mathcal{G}(ss'_x < T, ss_{\text{start}}, ss_{\text{end}})$ with $T = -\frac{1}{2}\sigma_{s'_x}$.

- *chair rise and sit down:* The chair rise and sit down events mark the beginning and end of the TUG. Rising from the chair is detected as $C^1 = \mathcal{G}(ss'_y > 0, ss_{\text{start}}, W^1_{\text{start}})$. Sitting down is detected as $C^2 = \mathcal{G}(ss'_y < 0, W^2_{\text{end}}, ss_{\text{end}})$.

- *turning:* The first turn takes place between the two walks and the second turn is done right before sitting down. Turns are detected by looking for the two highest peaks in sd_x: $(T^1, T^2) = \mathcal{F}(sd_x, 2, W^1_{\text{start}}, C^2_{\text{end}})$. The peaks are ordered based on their time of occurrence and assigned accordingly.

Figure 3.19 illustrates the detection of the 6 TUG phases based on the skeleton data. The left side shows the detection of the TUG phases within the signal of the whole TUG sequence. The right side illustrates detected TUG phases with the corresponding TUG movements. Figure 3.20a shows additional examples of detected TUG events based on the skeleton approach.

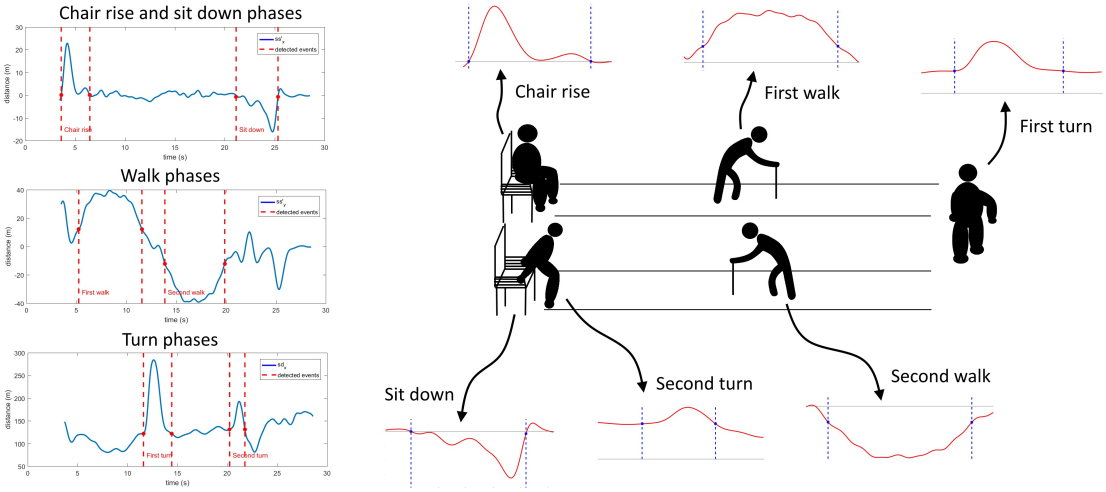
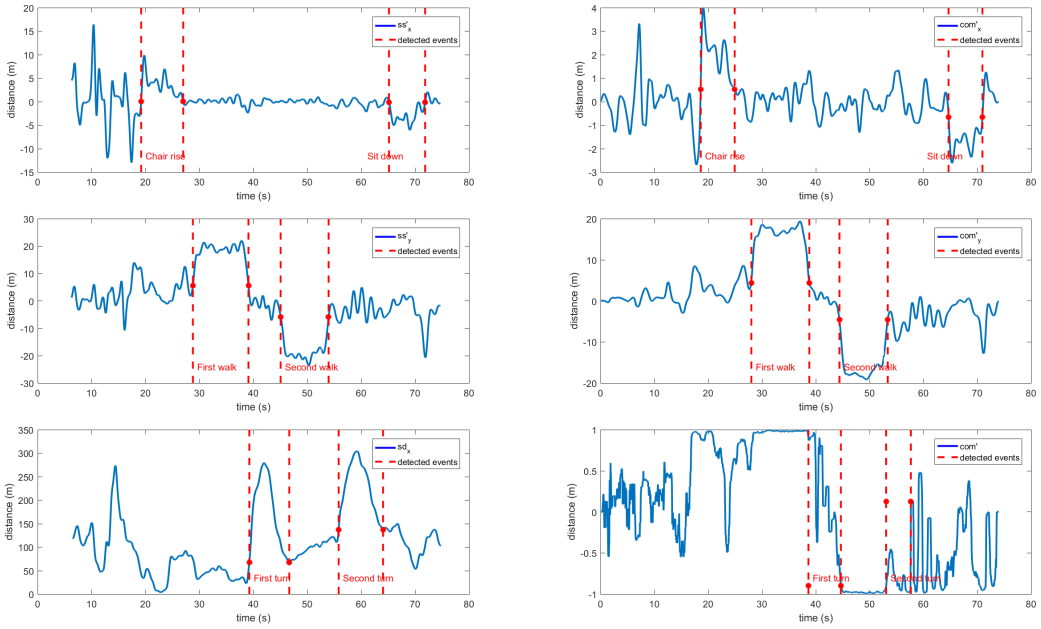


Figure 3.18: TUG analysis from depth data

Figure 3.19: Detection of 6 TUG phases based on skeleton data. The left side shows the occurrence of detected TUG phases within the signal of a whole TUG sequence. The right side shows enlarged snippets of detected TUG phases with the corresponding TUG movements.



(a) TUG analysis from skeleton data

(b) TUG analysis from depth data

Figure 3.20: Detection of chair rise and sit down events (upper images), walking events (center images) and turn events (bottom images) during the performance of a TUG test.

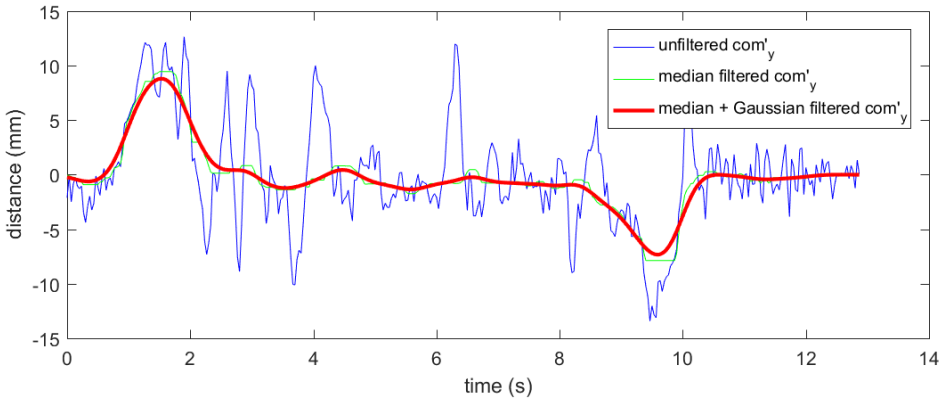


Figure 3.21: Applying a median filter and a Gaussian filter on com'_y; *blue:* unfiltered, *green:* median filter, *red:* median filter and Gaussian filter.

Depth data For depth data, the center of mass *com* of the point cloud of the subject is used instead of the trajectory of the joint *ss*. A median filter with a windows size of 1 second or 30 frames and a Gaussian filter with $\sigma = 5.0$ is applied on com'. A median filter with a large window size is used since com' is less stable and more noisy than ss' due to potential errors in the silhouette extraction. Figure 3.21 shows the effect of the median filter and the Gaussian filter on com'_y.

For the detection of turns the movement direction during the first walk $\vec{d_W}$ and the normalized ground projected movement direction from frame to frame $\vec{d}(k)$ are computed.

$$\vec{d}(k) = \|com_{xy}(k+l) - com_{xy}(k)\| \tag{3.12}$$

$com_{xy}(k)$ denotes the ground-projected position of the center of mass *com* and l is a constant with $l = 2$ being used. Then the function $\Delta(k) \in [-1, 1]$ is computed, which represents the similarity of the current walking direction with the walking direction of the first walk.

$$\Delta(k) = \vec{d_W} \cdot \vec{d}(k) \tag{3.13}$$

$\Delta(k) = 1$ denotes the same walking direction as the first walk and $\Delta(k) = -1$ denotes the opposite walking direction. A median filter with window size 30 is applied on Δ before the extraction of turn events. Based on the center of mass trajectory *com* and the function for indicating turns Δ, the following criteria are used for the detection of TUG events:

- *walking:* The detection of walking events is done analogue to the skeleton data. The only difference is that com'_x is used instead of ss'_x.

- *chair rise and sit down:* Like walking, chair rise and sit down events are detected analogue to the skeleton data. Besides using com'_y instead of ss'_y a higher threshold is used to better cope with noise. Chair rise is detected as
  $C^1 = \mathcal{G}(com'_y > T_1, com_{\text{start}}, \frac{W^1_{\text{start}}+W^1_{\text{end}}}{2})$ with $T_1 = \frac{1}{2}\sigma_{com'_y}$.
  Sitting down is detected as
  $C^2 = \mathcal{G}(com'_y < T_2, \frac{W^2_{\text{start}}+W^2_{\text{end}}}{2}, com_{\text{end}})$ with $T_2 = -\frac{1}{2}\sigma_{com'_y}$.
- *turning:* Detecting turning events from a single trajectory is an ill-posed problem, since there is no or little movement when a person turns while standing on the same spot. In order to detect the first turn, the transition from the walking direction of the first walk to the walking direction of the second walk has to be found. This is achieved by $(-, T^1_{\text{start}}) = \mathcal{G}(\Delta > T_1, \Delta_{\text{start}}, \Delta_{\text{end}})$ with $T_1 = 0.9$ and $(T^1_{\text{end}}, T^2_{\text{start}}) = \mathcal{G}(\Delta < T_2, \Delta_{\text{start}}, \Delta_{\text{end}})$ with $T_2 = -0.9$. The end of the first turn is detected when the current walking direction reaches the walking direction of the second turn ($\Delta < -0.9$). The beginning of the second turn is found when the walking direction differs again from the walking direction of the second walk, which marks the end of the ($\Delta < -0.9$) condition. The end of the second turn is found when Δ reaches zero after the event T^2_{start} or the end of Δ is reached.

Examples for detected TUG events using the approach based on the depth data are shown in Figure 3.20b.

CHAPTER 4

Results

This chapter describes the three acquired datasets and presents the obtained results for the proposed methodology. The chapter is divided in three parts as follows. The first part of the chapter describes the two acquired gait datasets for gait of both healthy and elderly adults and the acquired TUG test dataset of elderly adults. The second part presents the results for the gait analysis approach on both gait datasets and discusses the most important findings. The third part covers the results of the proposed TUG analysis approaches and compares them with a previously proposed method by Lohmann et al. [164]. Figure 4.1 shows an overview how each dataset is evaluated. The first row shows the three dataset, the second row shows the data type used for each dataset, then the extracted time series per data type, the extracted parameters for each time series and the bottom row shows the evaluation metrics applied on the parameters.

4.1 Data acquisition

A review of existing datasets is made and since no suitable public datasets are found, three datasets have been acquired as a part of this work.

4.1.1 Review of public data sets

Cai [41] discusses 46 RGBD datasets for object detection and tracking, human activity analysis, object and scene recognition, simultaneous localization and mapping and hand gesture analysis. Only three datasets are reviewed for human activity analysis and none of the reviewed datasets are relevant for this work. Firman [82] reviews 96 RGBD datasets for various applications and two can be considered relevant for this work, Kinect 3D Active and TST TUG.

Kinect 3D Active [155] includes recordings of standardized medical assessments including the TUG test. Participants are between 18 and 81 years old. TST TUG [54] contains

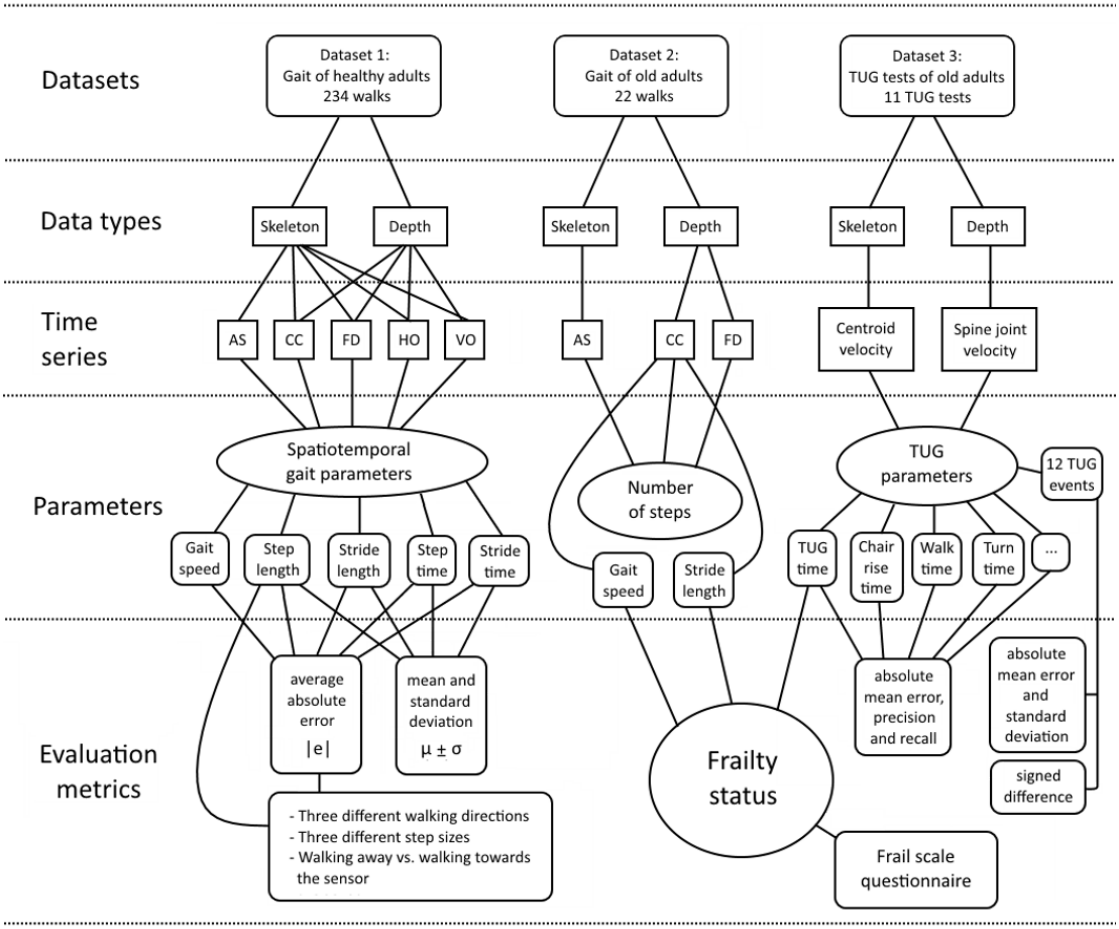


Figure 4.1: Overview of the evaluation of the acquired datasets and the proposed methods. The following abbreviations have been used: Ankle speed (AS), correlation coefficient (CC), feet distance (FD), horizontal oscillation (HO), vertical oscillation (VO), mean μ, standard deviation σ and mean absolute error |e|.

a TUG test performed by 20 healthy participants and is available online [1]. While the authors use this data to extract several TUG parameters which are also relevant for this work, e.g. time duration of the sit-to-stand phase, step length, cadence and TUG time, they do not evaluate the accuracy of their results. Instead, only mean and standard deviation of each parameter is provided in the paper. Moreover, their dataset does not include subjects with gait pathologies, making it unsuitable to assess the eligibility in real-world usage.

Kinect 3D Active and TST TUG do not have a labeling for TUG events and the accuracy of TUG event extraction is not yet assessed on these datasets. Since no comparable

[1] http://www.tlc.dii.univpm.it/blog/databases4kinect#IDTUG, Accessed 2017-07-12

results exist for these datasets and manual labeling would be necessary, these datasets are not used for the evaluation of the TUG automatization approach presented in section 3.2. For RGBD gait datasets, existing public datasets, e.g. TUM Gait [118] or DGait Database [35], are intended for gait recognition. While they include a large number of walking sequences, they do not have a labeling for the accuracy of spatiotemporal gait parameters.

4.1.2 Dataset 1: Gait of healthy adults

The gait dataset is used to evaluate the accuracy of extracted spatiotemporal parameters based on the described and implemented approach. In order to evaluate the accuracy under different conditions, multiple viewing angles, sensor heights and different step widths are used. Choices and details regarding participants, sensor, sensor positioning, walking path specifications and ground truth acquisition are discussed in this section.

Participants A total of 13 healthy subjects (11 male and 2 female) aged between 24 and 77 years participate in the gait recording.

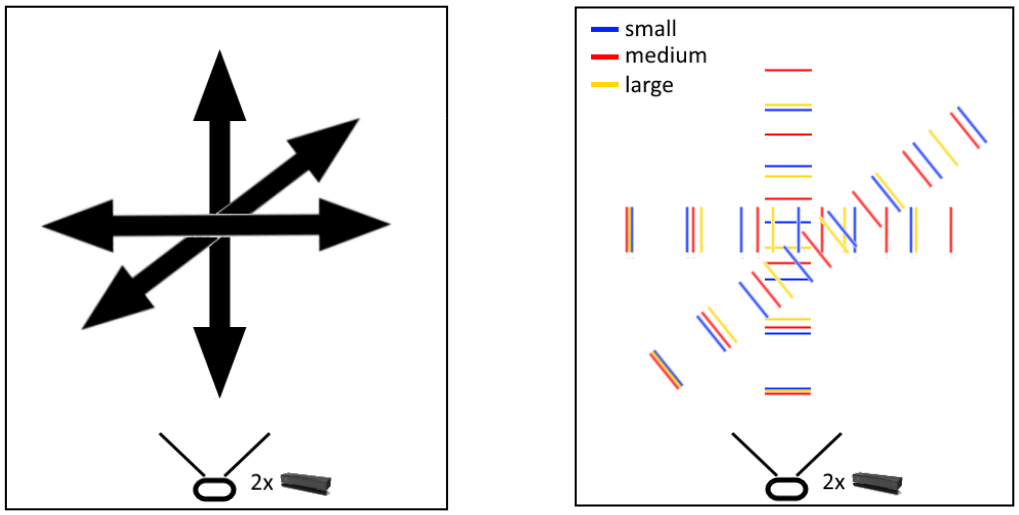
Sensor choice Kinect v2 is chosen as a sensor for several reasons: First, the superior field of view, resolution and accuracy discussed in section 2.2.2. Second, the observation that the amount of accuracy experiments done with Kinect v2 is significantly less than Kinect v1 and third, because it does not have the interference problems caused by structured light of Kinect v1.

Walking path specifications Three straight walking paths in two directions each are used to evaluate the effect of viewing angle on parameter accuracy. The walking paths are facing away and towards the sensor, fronto parallel to the sensor and diagonal to the sensor. Figure 4.2a illustrates the position of the walking paths compared to the sensor.

Recommendations for the assessment of gait variabilities suggest, that the distance for assessing gait variability should be at least 20 m [192] or 12 continuous steps [166]. Considering that the walk should be recorded from different viewing angles and Kinect v2 only supports a maximum range of 4 meters, multiple walks back and forth are necessary. Therefore each walking path is taken three times back and forth in order to reach the suggested 20 meters walking distance.

Sensor and Sensor positioning Two Kinect v2 are placed on top of each other in different heights using tripods. Kinect v2 #1 is placed at 1.2 meter height tilted down around 10° and Kinect v2 #2 close to the ceiling at 2.4 meters with a downward tilt of around 25°. The experimental setup is shown in Figure 4.5a. In order for a point on the ground to be within the field of view of the sensor the minimum distance from the sensor position is given by

$$d = h \ \tan(\frac{\pi}{2} - \frac{\alpha}{2} + \gamma) \tag{4.1}$$



(a) Three walking paths with walking two directions each.

(b) Markers with different step lengths placed in the field of view of both Kinects.

Figure 4.2: Specifications of the walking paths (left image) and placement of the ground markers (right image).

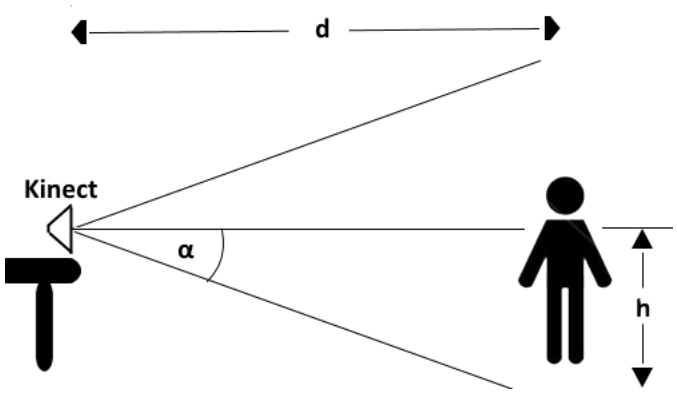
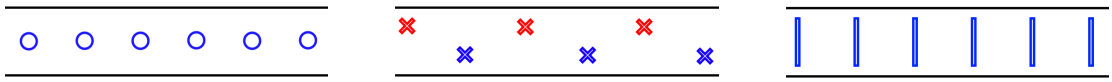


Figure 4.3: Required distance d for the subjects' feet to be within the vertical field of view α from the Kinect placed in height h.

where d represents the distance from the walking subject, h denotes the sensor height, γ is the tilt of the Kinect sensor (negative in case of a downward tilt), h depicts the height of the sensor and α is the vertical field of view of the camera. Equation 4.1 is illustrated in Figure 4.3. For Kinect #1 the minimum distance between the subject and the sensor is 1.43 meters and for Kinect #2 the minimum distance is 1.68 meters.



(a) (b) (c)

Figure 4.4: Left image: Point shaped markers placed in the center. Center image: X-shaped markers with lateral spacing. Right image: Line-shaped markers perpendicular to the walking direction.

Ground truth acquisition

Spatial ground truth Ground markers are used to determine the ground truth for step length and stride length during the evaluation of the walking sequences. In order to pick a suitable ground marker, three different types of markers are tested in a small gait experiment with four participants. Subjects are asked to step on top of the markers that are placed in a 4.5 meter long hallway. Two persons are below average height (152cm and 160cm height) and two persons are above average height (182cm and 185cm). All markers are placed on the ground using stick tape.

The first type of tested markers, simple point markers (see Figure 4.4a), indicates the position of either foot in the center of the walking path. These markers make the individuals step inward with each foot in order to reach the marker and they made gait less natural. The second type of tested markers is x-shaped (see Figure 4.4b) and indicates the position of left and right foot separately with lateral spacing. All four participants report that these markers require additional focus in order to correctly reach the markers with the heel strike. The third type of tested markers is line-shaped (around 20cm long) and is illustrated in Figure 4.4c. They can be stepped on by either foot and do not require the subject to step inward due to their length. All four subjects report that they prefer the line-shaped markers, since they require the least amount of additional focus and therefore facilitate the most natural walk among the three tested markers. The chosen line-shaped markers distributed on the three walking paths are shown schematically in Figure 4.2b and in the actual experimental setup in Figure 4.5b. The same line-shaped stick tape markers are also used by Staranowicz et al. [253] in their gait experiment.

Determining step lengths The goal is to use close to average step lengths in order to facilitate natural walkings. Montero-Odasso et al. [182] report an average stride length of 1.4 meters under fast walking pace and 1.27 meters under usual walking pace among healthy individuals. Based on this data the ground markers are put in a distance of 63 cm for normal step length and 70 cm for large step length. An additional small step length of 55 cm is added to the previous two.

Determining temporal ground truth In order to obtain the temporal ground truth for each walking sequence, all walks are manually inspected to label the frame number of

(a) Sensor setup: Two Kinect v2 placed in different heights using tripods.

(b) Walking path setup: Differently colored line markers indicate different step length: *blue:* short steps. *red:* normal steps. *yellow:* large steps.

Figure 4.5: Left image: Experimental setup. Right image: Line shaped markers have been used to indicate three different step lengths in three directions.

each heel strike and toe-off event. Recorded videos from an additional RGB camera are used for the manual ground truth annotation.

Figure 4.6 shows color images and corresponding depth images of the acquired gait dataset from both Kinect #1 and Kinect #2. Sample frames for Kinect #2 are shown on the left side and sample frames for Kinect #1 are shown on the right side.

4.1.3 Dataset 2 and dataset 3: Elderly gait and TUG tests

The purpose of the second and third dataset is to evaluate the TUG automatization method presented in section 3.2 and to evaluate the extraction of gait parameters with the gait of older adults. Residents of a nursing facility in Grieskirchen, Austria are asked to participate in performing the TUG test. Additionally they are asked to walk along a three meter walking path back and forth once. A total of eleven participants, 7 female and 4 male, volunteer to participate. Participants are between 85 and 95 years old with a mean age of $\mu_{\text{age}} = 89.3$ years and a standard deviation of $\sigma_{\text{age}} = 3.6$ years. All participants are asked to perform the TUG test once. Failed TUG tests, e.g. due to unclear instructions, are repeated until the test was performed smoothly. 9 out of 11 participants prefer to use a walking aid for walking and during the performance of the TUG test. No participant requires physical assistance in order to complete the TUG test and the two walking sequences. The experimental setup used for the recording and several sample images are shown in Figure 4.7. A Kinect v2 sensor was placed fronto

Figure 4.6: Sample images and corresponding depth frames of recorded walking sequences from two different heights: Kinect #2 (left images) and Kinect #1 (right images).

parallel to the TUG walking path on a table with 77 cm height around 3.5 meters away from the path.

In order to further evaluate the physical condition of the participants they are asked five questions from a frailty questionnaire, FRAIL scale [275]. FRAIL scale questionnaire consists of the following five questions:

- "How much of the time during the past 4 weeks did you feel tired?" Responding "All of the time" or "Most of the time" corresponds to 1 point.
- "By yourself and not using aids, do you have any difficulty walking up 10 steps without resting?" Responding "Yes" corresponds to 1 point.
- "By yourself and not using aids, do you have any difficulty walking several hundred yards?" Responding "Yes" corresponds to 1 point.
- For 11 illnesses, participants are asked, "Did a doctor ever tell you that you have [illness]?". Responding "Yes" for 5 or more illnesses corresponds to 1 point. The illnesses include hypertension, diabetes, cancer, chronic lung disease, heart attack, congestive heart failure, angina, asthma, arthritis, stroke, and kidney disease.
- "How much do you weigh with your clothes on but without shoes? [current weight]". "One year ago in (MO, YR), how much did you weigh without your shoes and with your clothes on? [weight 1 year ago]". 1 point is scored if (([weight 1 year ago] - [current weight])/[weight 1 year ago]) > 0.05.

(a) Picture of the TUG setup. *red:* x-axis, *green:* z-axis, *blue:* y-axis

(b) Elderly participants during their performance of the TUG test.

Figure 4.7: TUG recording setup (left image) and samples during recording (right image). Red stick tape was used to mark the line 3 meters away from the armchair.

A total score of 0 represents robust, 1-2 represents pre-frail and 3-5 represents frail individuals. FRAIL scale is validated in a clinical context for identifying frail persons and persons at risk of frailty [184, 165, 295]. Since the original FRAIL questionnaire is only available in English, the questionnaire is translated to make it suitable for the German speaking population in the nursing facility. A translated version of the questionnaire is depicted in Appendix A.

4.2 Gait evaluation

The gait accuracy evaluation focuses on the evaluation of gait for the healthy gait dataset and the elderly gait dataset.

Figure 4.8 shows an example of extracted footsteps for all 18 walking sequences of one participant. The location of the extracted footsteps is extracted based on the ankle speed signal from skeleton data. In this example the absolute step length error is 2.14 cm per step.

Figure 4.8 shows an example of extracted footsteps using four gait signals based on depth data. The spatial location of each detected footstep is visualized beside the signal and the first and last footstep are connected with a line to indicate their correspondence. It can be seen that the signals of horizontal oscillation and vertical oscillation are less regular compared to the signals of correlation coefficient and feet distance.

Results for healthy gait dataset

The results on the healthy gait dataset are based on the Kinect #1, which is placed 1.2 m. The following spatiotemporal gait parameters are extracted from each gait signal: Gait speed, step length, step time, stride length and stride time. Mean and standard deviation as well as mean absolute error are used as evaluation metrics. Additionally, for the step length parameter the mean absolute error is examined under different conditions.

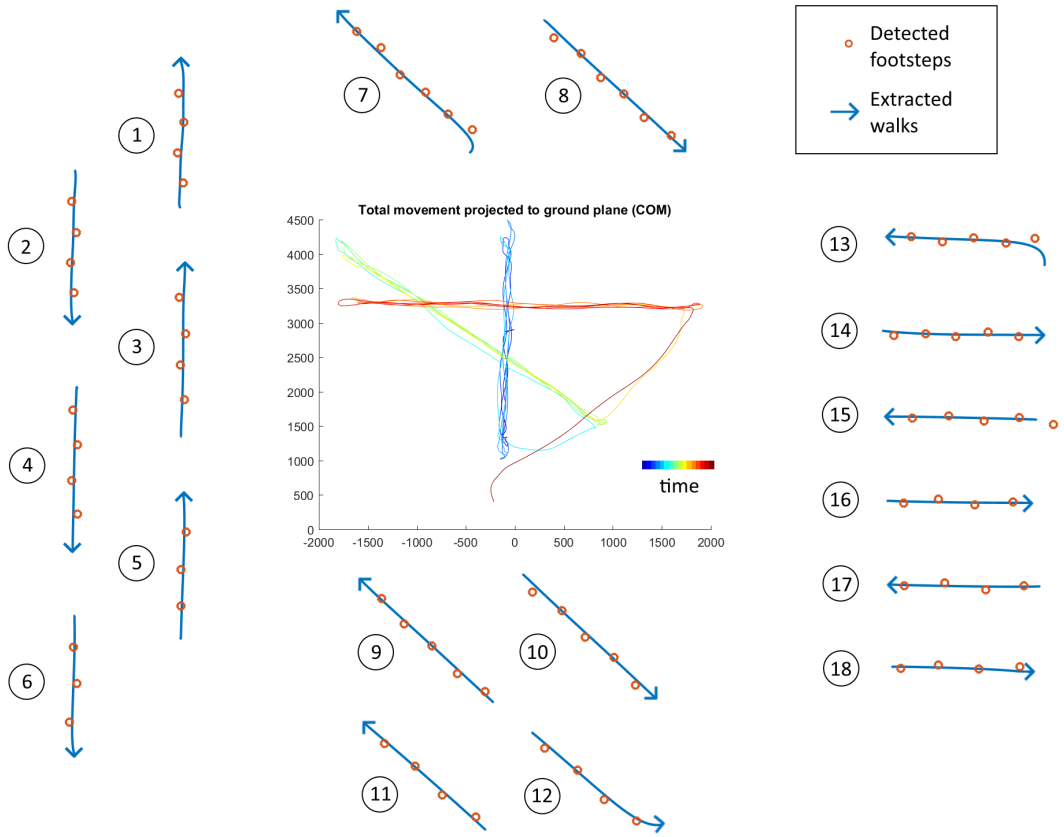


Figure 4.8: Detected footsteps from the extracted 18 walks of a whole walking sequence. The first two walks of a particular walking direction are small steps, followed by two walks of medium steps and two walks of large steps. The mean absolute step length error over all 18 walks on this particular walking sequence is 2.14 cm.

The conditions are three different walking paths, three different step sizes and if the subject walked away or towards the sensor.

Four gait signals are implemented for depth data and five gait signals for skeleton data: Correlation coefficient, feet distance (FD), horizontal oscillation (HO), vertical oscillation (VO) and ankle speed (AS). Correlation coefficient is evaluated twice, once as described in the methodology (here referred as CCTS) and once exactly as stated in the paper of Stone and Skubic [255] (here referred as CC). The main difference between the two signals is that the authors of the paper used moving average filters on the signal, the center of mass and the time series.

Table 4.1 shows the mean and standard deviation for four extracted gait parameters over all 234 walking sequences. For depth data, the correlation coefficient method of Stone and Skubic [255] and the feet distance method perform the best with the smallest

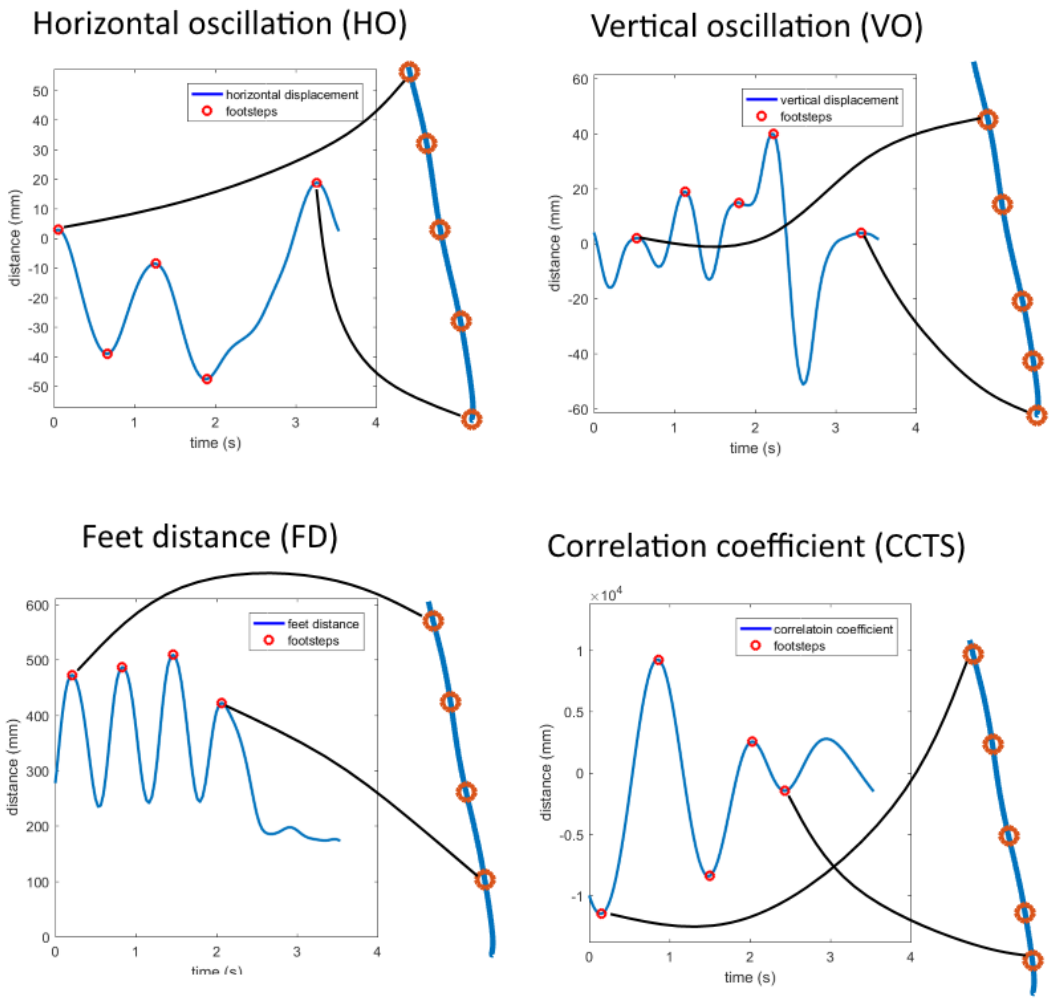


Figure 4.9: Detected footsteps for four gait signals (HO, VO, FD and CCTS) on one walking sequence. For each gait signal, the corresponding spatial locations for the detected footsteps are visualized.

standard deviation increase compared to the ground truth. For skeleton data, the ankle speed method shows similar performance as the top two methods for depth data, while all other skeleton-based methods show significantly higher standard deviations. VO and HO show poor results for both depth data and skeleton data with the results for skeleton data being slightly superior. The reason for the high error rates of VO and HO is that the vertical and horizontal oscillation during gait is not significant enough to robustly and accurately detect all footsteps. Signals obtained from VO and HO are irregular and not sufficiently smooth to be well-suited for gait analysis.

The absolute mean error and the relative mean error for five parameters over all walking

Mean and standard deviation per parameter

	Gait signal	Step length (cm)	Step time (ms)	Stride length (cm)	Stride time (ms)
Ground truth		62.7 ± 6.3	679 ± 59	125.3 ± 12.6	1352 ± 91
Depth	CCTS	63.0 ± 10.0	679 ± 98	125.8 ± 15.8	1323 ± 142
	CC [255]	62.1 ± 8.5	665 ± 76	124.6 ± 13.6	1321 ± 115
	Feet distance	62.8 ± 8.4	657 ± 77	124.6 ± 14.9	1305 ± 130
	VO	64.2 ± 19.5	706 ± 217	128.6 ± 24.5	1399 ± 287
	HO	66.6 ± 37.0	737 ± 396	125.8 ± 43.4	1392 ± 484
Skeleton	Ankle speed	61.8 ± 8.4	662 ± 126	123.2 ± 13.9	1313 ± 162
	CCTS	65.8 ± 21.4	734 ± 232	126.7 ± 20.9	1394 ± 236
	CC [255]	63.7 ± 17.1	702 ± 184	125.2 ± 18.6	1360 ± 204
	Feet distance	68.0 ± 20.6	734 ± 225	133.9 ± 29.4	1441 ± 335
	VO	63.2 ± 13.8	686 ± 157	126.7 ± 20.0	1370 ± 235
	HO	64.7 ± 29.3	712 ± 318	126.3 ± 34.6	1380 ± 397

Table 4.1: Mean and standard deviation of extracted spatiotemporal gait parameters for both ground truth and various gait signals.

sequences are shown in Table 4.2. The absolute mean error is computed as the averaged absolute difference from the ground truth. Similar to the previous result, the smallest error is found for the correlation coefficient approach of Stone and Skubic [255] and the feet distance approach for the depth data and the ankle speed approach for the skeleton data.

From all extracted parameters, gait speed shows the lowest error for all gait signals. This is expected, since gait speed is calculated as distance traveled divided by time elapsed and therefore it is not affected by wrong footstep detections within a walking sequence. Moreover, this is consistent with previous observations, e.g. the work of Dubois and Charpillet [76]. They report that gait speed is the most accurate of three estimated gait parameters (step length, step time and gait speed).

Table 4.3 shows a the absolute mean step length error based on three different walking paths. For depth data and correlation coefficient, the walking path has almost no effect on the accuracy. For the feet distance signal, best performance is obtained for fronto parallel walking and diagonal walking shows the worst performance. Gait signals based on skeleton data are generally more affected by the direction of the walking path. Ankle speed and correlation coefficient perform worst for diagonal walking, however, the effect is not as noticeable for the ankle speed data. Considering only the best performing gait signals, depth CC, depth FD and skeleton AS, the walking path does not have a significant effect on the step length error. This is expected, as both depth and skeleton approaches are model-based and therefore considered invariant to viewing angle.

Mean absolute error per parameter

Data Source	CCTS	CC [255]	FD	HO	VO	AnkSp
Step length error (mm)						
Depth	48 (7.6%)	45 (7.2%)	41 (6.5%)	254 (40.6%)	142 (22.6%)	–
Skeleton	84 (13.4%)	73 (11.6%)	104 (16.3%)	191 (30.5%)	79 (12.6%)	38 (6.1%)
Stride length error (mm)						
Depth	75 (6.0%)	62 (5.1%)	61 (5.0%)	317 (25.8%)	152 (12.3%)	-
Skeleton	93 (7.5%)	81 (6.6%)	156 (12.6%)	226 (18.1%)	102 (8.2%)	51 (4.1%)
Step time error (s)						
Depth	0.06 (8.8%)	0.04 (6.6%)	0.05 (8.0%)	0.28 (41.8%)	0.16 (24.3%)	-
Skeleton	0.10 (15.0%)	0.08 (12.1%)	0.12 (17.9%)	0.21 (31.9%)	0.10 (14.7%)	0.08 (12.4%)
Stride time error (s)						
Depth	0.07 (5.7%)	0.05 (3.7%)	0.07 (5.3%)	0.33 (25.3%)	0.18 (13.3%)	-
Skeleton	0.09 (7.1%)	0.09 (6.6%)	0.18 (13.8%)	0.25 (18.6%)	0.12 (9.4%)	0.08 (6.3%)
Gait speed error (mm/s)						
Depth	16.3 (1.70%)	10.7 (1.12%)	26.4 (2.75%)	8.99 (0.94%)	28.11 (2.94%)	-
Skeleton	34.7 (3.62%)	27.7 (2.89%)	18.8 (1.96%)	29.6 (3.09%)	16.3 (1.70%)	15.2 (1.58%)

Table 4.2: Mean absolute error over all walking sequences of several spatiotemporal gait parameters. The following additional abbreviations were used: Feet distance (FD), horizontal oscillation (HO), vertical oscillation (VO) and ankle speed (AnkSp).

Step length: Mean absolute error (cm) per walking direction

Data	Gait signal	All views	Frontal view	Diagonal view	Fronto parallel view
Depth	CC [255]	45	45	42	44
	CCTS	48	48	45	47
	Feet distance	41	58	37	29
	Horizontal oscillation	254	212	266	284
	Vertical oscillation	142	154	228	51
Skeleton	Ankle speed	38	37	43	34
	CC [255]	73	62	103	54
	CCTS	84	66	135	58
	Feet distance	104	136	125	50
	Horizontal oscillation	191	196	178	209
	Vertical oscillation	79	50	72	112

Table 4.3: Comparison of mean absolute step length error in mm based on three different walking directions and different gait signals.

A comparison of step length mean and average step length error based on three different step lengths is shown in Table 4.4. Larger step lengths show a slightly increased step length error, but this increase is relative to the increase in step length. After normalizing the step length error based on the size of the corresponding step, the different step sizes do not show a significant effect on the step length error for all gait signals. Therefore the proposed gait analysis approach is considered independent of the step size.

Table 4.5 compares the mean absolute step length error between walking towards the Kinect sensor and walking away from the sensor. HO and VO are not considered further as they both showed poor performance in the previous results. While almost no difference is found for the depth data, the difference for the skeleton data is significant. For skeleton AS, the mean absolute error increases from 2.7 cm to 4.6 cm when walking away along the z-axis instead of walking towards the sensor. For skeleton CC and skeleton FD the increase is even larger: 10.1 cm instead of 3.1 cm for skeleton CC and 20.9 instead of 6.3 for skeleton FD. The step length error is approximately three times as large for walking away compared to walking towards the sensor. Therefore the poor overall results of skeleton CC and skeleton FD are caused by poor accuracy when walking away from the sensor.

The reason for the higher errors is that the Kinect pose estimation is trained for subjects

Step length: Mean and mean absolute error in cm per step size

Data	Gait signal	Small steps		Medium steps		large steps	
		mean	error	mean	error	mean	error
Ground truth		55	0	63	0	70	0
Depth	CC [255]	56.2	3.9	62.7	4.8	69.5	4.9
	CCTS	57.2	4.7	63.9	5.0	69.5	4.7
	Feet distance	57.4	4.4	63.8	4.3	69.3	5.0
	Horizontal oscillation	61.7	22.4	68.5	26.0	73.2	30.3
	Vertical oscillation	61.6	13.9	64.3	14.7	67.3	13.8
Skeleton	Ankle speed	55.4	3.5	63.2	4.0	69.1	4.1
	CC [255]	58.2	6.4	64.7	7.2	69.7	8.5
	CCTS	60.5	8.1	66.1	8.2	72.4	9.2
	Feet distance	61.9	9.8	68.6	9.7	75.5	12.2
	Horizontal oscillation	59.9	15.1	65.2	19.1	70.3	23.9
	Vertical oscillation	58.6	9.7	63.3	6.6	69.2	7.2

Table 4.4: Comparison of mean step length and mean absolute step length error in cm based on three different step sizes.

facing the Kinect sensor.Moreover, when walking towards the sensor, there is no occlusion at the time of the heel-strike for the foot swinging forward. When walking away from the sensor, the foot swinging forward is likely to be occluded from the standing foot that supports the body weight at the time of the heel-strike. Figure 4.10 shows ground-projected ankle joints from two walking sequences, one walking towards the sensor and one walking away from the sensor. Left and right ankle joints are connected with a line to illustrate their temporal correspondence. Both original and filtered joints are shown. In the case of walking towards the sensor, it can be seen that the lateral distance remains roughly equal during the whole walking sequence. Moreover, footsteps (stationary ankle positions) are clearly visible as clusters. However, when walking away from the sensor, the ankle joint positions are highly inaccurate for the left footsteps, which represents the occluded foot. In this case it is likely that false positives occur for the footsteps of the occluded foot. The joint coordinate remains stationary during the time of occlusion, which may cause a peak or trough in the gait signal and therefore a false footstep may be detected at this location.

When comparing the results of individual participants, it is noticeable that one participant wearing a floor-length skirt has the highest error rates among all participants. For example

Step length: Mean absolute error (cm) for walking towards/away from the sensor

Data	Gait signal	Towards the sensor		Away from the sensor			
		frontal	diagonal	frontal	diagonal		
Ground truth		55	0	63	0	70	0
Depth	CC [255]	4.3	3.9	4.8	4.5		
	Feet distance	6.8	3.3	4.7	4.2		
Skeleton	Ankle speed	2.7	3.9	4.6	4.7		
	CC [255]	3.1	4.3	10.1	19.6		
	Feet distance	6.3	5.4	20.9	21.9		

Table 4.5: Mean absolute step length error in cm based on walking towards the sensor or walking away from the sensor. Frontal refers to walking parallel to the z-axis and diagonal refers to walking in a 45° angle from the z-axis towards and away from the sensor.

the step length errors for this participant is 13.8 cm for the depth FD signal, 6.0 cm for the depth CC signal [255] and 9.2 for the skeleton AS signal. This indicates that the correlation coefficient signal is the most robust towards occlusion of the feet among the tested signals.

4.2.1 Evaluation of elderly gait

The best three performing gait signals from the accuracy evaluation are further evaluated on the walking sequences of the older adults. This is a particularly challenging dataset for several reasons. The first reason is that the majority of older adults in the dataset use extremely small steps while barely lifting their feet from the ground. Second reason is that 9 out of 11 elderly adults use a walker for additional balance and stability support. The occlusions of the leg and feet caused by the walking aid leads to a noticable decrease in skeleton joint accuracy. In case of the depth data, the walking aid is arbitrarily considered foreground or background, depending whether the silhouettes of the subject and the walking aid are connected or not. While this problem could possibly be solved with an adapted silhouette extraction method, one purpose of this evaluation is to see how the evaluated approaches behave with the additional noise.

An evaluation based on the number of steps for each walking sequence is conducted. For all 11 elderly participants, both walking sequences are considered and the steps for each walking sequence are counted. As shown in Table 4.6, none of the top 3 original gait signals show adequate accuracy for the older adults' gait. The reason for the poor performance on the gait of older adults is the window size in equation 3.11. It decreases for faster gait and increases for slower gait, however, the frequency of the gait signal is determined by the cadence and not by gait speed. The equation is based on the

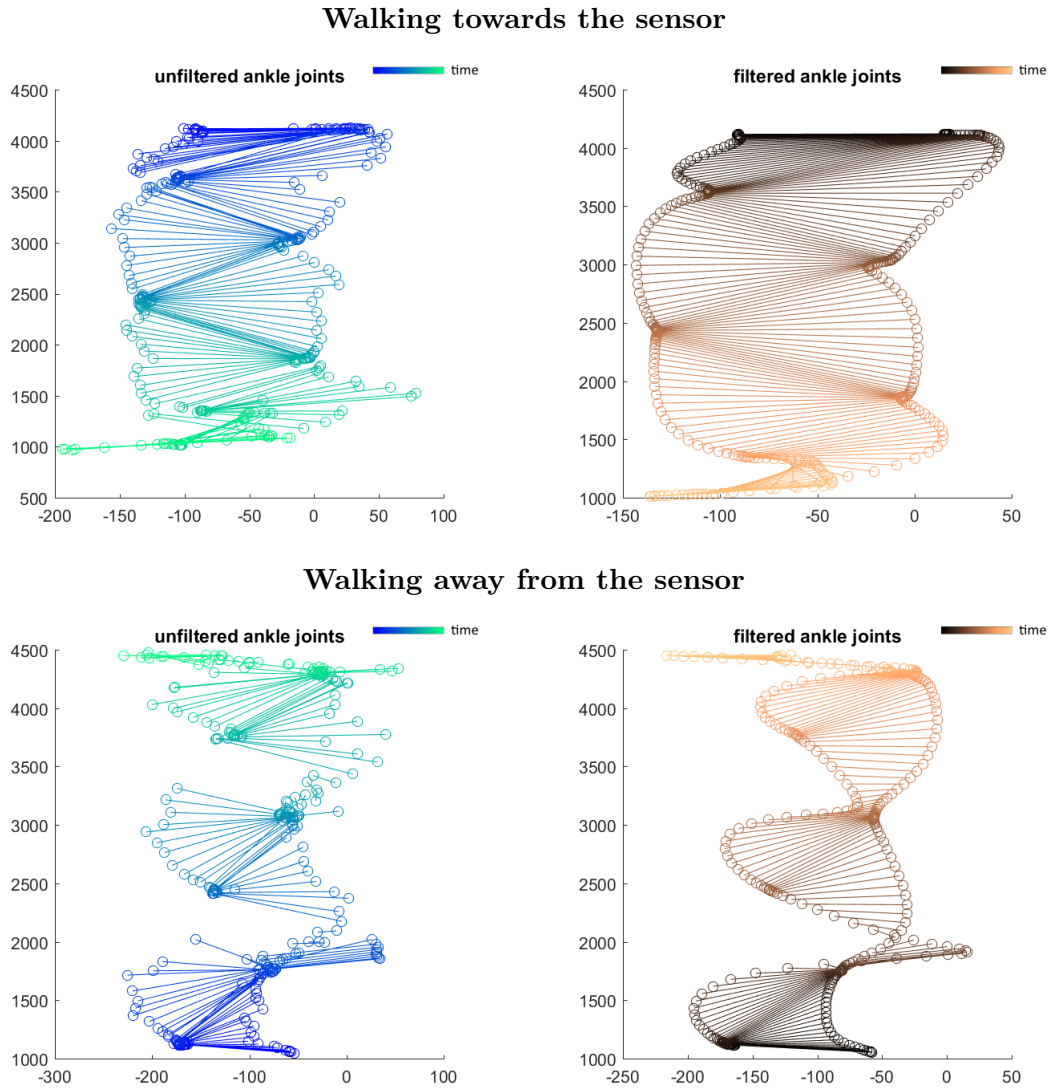


Figure 4.10: Ground-projected corresponding ankle joints of the skeleton model for a walking sequence towards the sensor (top row) and away from the sensor (bottom row). Left column shows unfiltered corresponding ankle joints and right column row shows filtered ankle joints.

assumption that a faster gait is caused by a larger cadence and should therefore be filtered with a smaller filter size compared to a slower gait. However, this assumption is wrong, since the natural reduction in walking speed among older people is caused by a reduction in stride length instead of a decrease in cadence [292]. Normalizing the filter size based on gait speed instead of cadence like in equation 3.11 leads to very large window sizes for very slow walkers and subsequently over-filtering of the signal. Therefore,

		Original			Adjusted		
		Depth		Skeleton	Depth		Skeleton
P#	Ground truth	CC [255]	Feet distance	Ankle speed	CC [255]	Feet distance	Ankle speed
#1	5 5	5 5	7 6	5 6	5 5	6 5	5 6
#2	7 7	7 7	6 7	9 7	7 7	7 7	9 7
#3	9 8	9 8	10 8	14 9	9 8	10 9	10 9
#4	9 7	8 5	7 3	9 7	9 7	10 10	11 11
#5	12 15	6 8	5 8	0* 11	13 15	13 17	0* 15
#6	6 9	6 8	5 8	6 9	6 9	6 8	6 9
#7	10 10	10 9	9 7	12 10	10 10	10 11	14 11
#8	9 10	8 9	7 5	13 9	9 10	12 12	9 11
#9	6 7	6 5	6 6	0* 10	6 7	6 9	0* 11
#10	10 9	7 7	4 4	8 9	10 10	13 11	10 9
#11	8 5	6 4	3 2	7 5	8 5	9 5	9 6
# of correct items		9	3	8	**21**	7	8

Table 4.6: Evaluation for the number of steps for both described and adjusted gait signals. The following abbreviation was used: P# for the participant number. The asterisk denotes that no skeleton tracking was available for this walking sequence.

the following adjusted window size is applied to solve this problem

$$w = f * k \tag{4.2}$$

Again, the frame rate f and the height constant k are used for determining the filter size,

but gait speed is replaced with cadence and cadence is estimated as a constant 1 stride per second. This is feasible as cadence changes only slightly between young, old and frail gait and similar to height, the possible range of cadence between different subjects is small. However, the value for cadence can also be learned over time for long-term gait assessment.

Figure 4.11 shows a comparison of CC filtered with the original window size (see equation 3.11) and the adjusted window size of equation 4.2 for the walk of three participants. The CC signal filtered with the original window size is shown in Figure 4.11a and the CC signal filtered with the adjusted window size is shown in Figure 4.11b. It can be seen that the filter size of the original window size is too large and the peaks and troughs of the unfiltered signal (illustrated as black signal) disappear after applying the median filter. In total, the problem of over-filtering for very slow walkers occurred for 5 out of 11 participants.It should be noted that even though this window size is applied in several papers [258, 255, 121], none of them reported this problem.

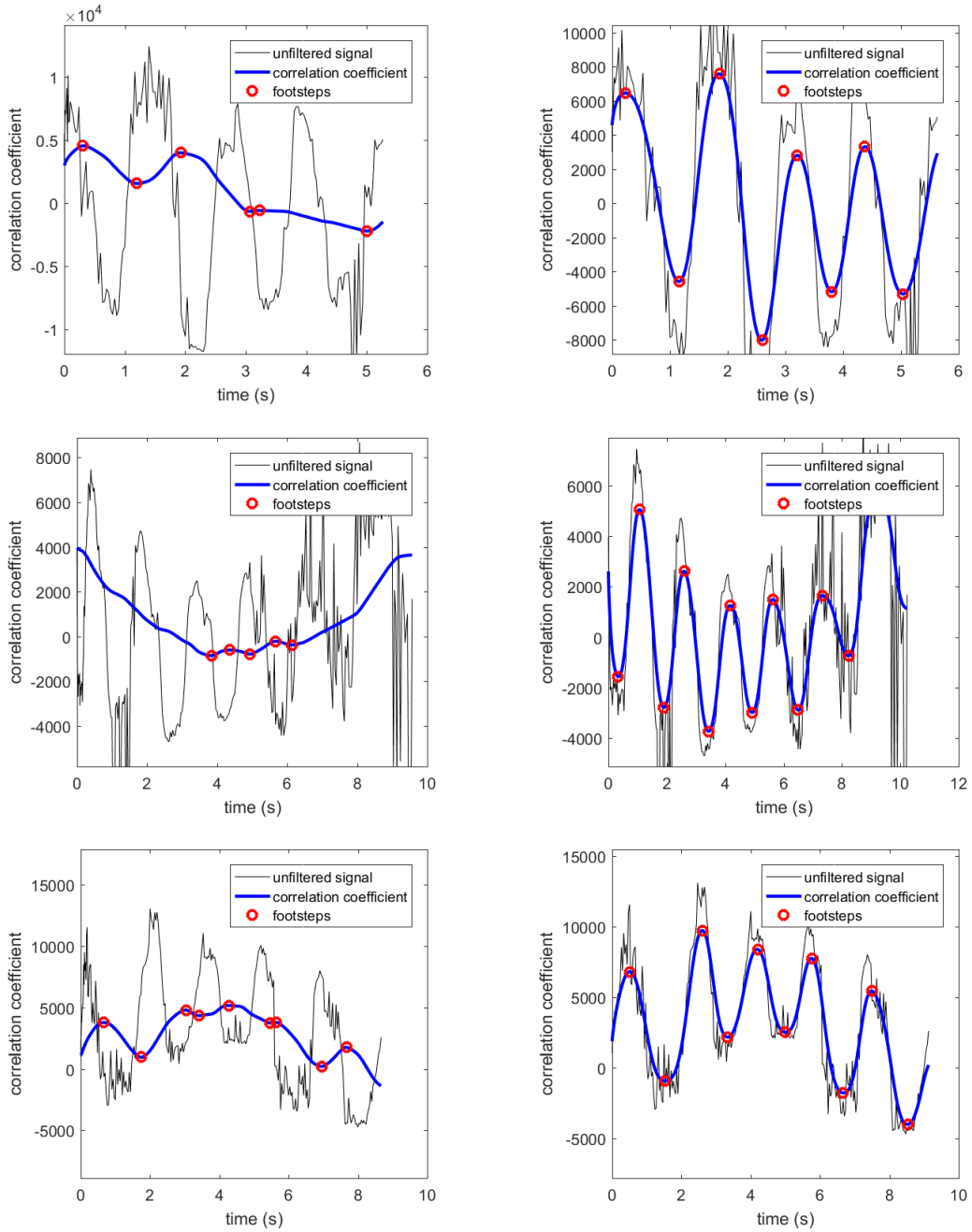
From the results of the adjusted window size in Table 4.6, only adjusted correlation coefficient performs well. The presumed reason is that both feet distance and ankle velocity signals are strongly affected by the additional noise of the dataset and correcting the filter size does not solve this problem. While the correlation coefficient time series describes two steps per wavelength, both the feet distance and ankle velocity signals only describe one step per wavelength, doubling the frequency. This exposes an important observation regarding the advantage of the correlation coefficient signal over the other gait signals: In order for a gait signal to be maximally robust towards noise, one wavelength should represent one whole gait cycle.

4.2.2 Gait results summary

Considering both the gait accuracy evaluation and the evaluation of the best-performing gait signals on the gait of older adults, the adjusted correlation coefficient based on depth data shows the best performance. While the results are similar for the first gait evaluation, only the adjusted correlation coefficient shows satisfying results on the elder people's gait dataset (see Table 4.6). Moreover, since it is based on depth data, it does not have the range limitation of the skeleton tracking and the additional inaccuracy when subjects are walking away from the sensor. Furthermore, skeleton data is not fully reliable for the older adults gait dataset as no skeleton tracking is available for 2 out of 22 walking sequences.

4.2.3 Frailty comparison: Gait, TUG and questionnaire

The frailty status of the elderly participants is evaluated based on four sources: Gait speed determined from adjusted correlation coefficient, stride length determined from adjusted correlation coefficient, TUG score and the result of the FRAIL scale questionnaire. The following cutoff values are used based on Table 1.1, the TUG score cutoff of the Edmonton Frail Scale [228] and the suggested evaluation of the FRAIL scale [184]



(a) Original filter window size proposed by Stone and Skubic [255] is too large for slow walkers.

(b) Adjusted window size with movespeed replaced by cadence and cadence set to a constant 1 stride per second.

Figure 4.11: Correlation coefficient filtered with original window size used by [255, 121] (left image) and adjusted window size (right image). The black line illustrates the unfiltered signal.

P#	Gait speed (m/s)	frailty status	Stride length (m)	frailty status	TUG time	frailty status	FRAIL scale	frailty status
#1	1.04	robust	1.06	pre-frail	9.8	robust	0/5	robust
#2	0.78	pre-frail	0.85	frail	19.9	pre-frail	3/5	frail
#3	0.57	frail	0.68	frail	22.0	frail	3/5	frail
#4	0.35	frail	0.51	frail	26.2	frail	4/5	frail
#5	0.27	frail	0.25	frail	53.0	frail	3/5	frail
#6	0.57	frail	0.63	frail	16.9	pre-frail	0/5	robust
#7	0.41	frail	0.57	frail	25.2	frail	3/5	frail
#8	0.34	frail	0.62	frail	27.7	frail	3/5	frail
#9	0.45	frail	0.76	frail	21.4	frail	2/5	pre-frail
#10	0.43	frail	0.32	frail	47.1	frail	3/5	frail
#11	0.40	frail	0.58	frail	27.7	frail	3/5	frail

Table 4.7: Comparison of detected frailty status between gait speed, stride length, TUG test and FRAIL scale among the 11 participants.

- Gait speed: Robust $>$ 1.0 m/s. Frail $<$ 0.85 m/s.
- Stride length: Robust $>$ 1.2 m. Frail $<$ 1.0 m.
- TUG score: Robust $<$ 10 s. Frail $>$ 20 s.
- FRAIL scale questionnaire: Robust $= 0$. Frail ≥ 3.

Table 4.7 shows the assigned frailty status for each of the frailty measures and each participant. 9 participants are assigned frail based on gait speed, 10 based on stride length and 8 based on TUG score and FRAIL scale questionnaire. Participant #6 is especially interesting. He is assigned at the same time to be robust, pre-frail and frail based on different frailty measures. This participant is noticeably jolly during the interview and confident in his abilities to perform activities of daily life independently. However, despite his positive nature and not using a walking aid, his low gait speed and stride length suggest possible mobility impairments. His data match the values that are common for physical frailty.

4.3 TUG automatization results

The proposed TUG automatization approaches based on depth and skeleton data are evaluated on the acquired elderly TUG test dataset. For the purpose of evaluation, the start and end events of six TUG phases are labeled using manual video inspection.

TUG time is estimated as the elapsed time between the first and last TUG event. Additionally, a stopwatch is used during the recording of the TUG test and manually obtained TUG times are compared with the automatically estimated TUG times.

4.3.1 Skeleton Timed Up and Go

For the purpose of evaluation, the TUG automatization method of Lohmann et al. [164], called Skeleton Timed Up and Go (sTUG), is implemented in addition to the described methods. Their approach detects a total of ten TUG events during the duration of the TUG test. Nine TUG events detected in their approach are also detected in the approaches proposed in this work. sTUG uses thresholds for the acceleration of joint trajectories to detect the events *start moving* M_s, *start walking* W_s, *end uprising* U_e, *start lowering* D_s and *end moving* M_e. The events *start rotating* R_s and *end rotating* R_e are detected based on the distance of the two shoulder joints. The events *start accelerating* A_s and *end accelerating* A_e, which represent the end of the first walk and the beginning of the second walk, are detected based on thresholding the velocity in walking direction. In order to evaluate both walk times, an additional event *end walking* W_e is detected based on their approach. W_e is detected as the end of the second walk using the same method they used for detecting A_s and A_e.

The paper of Lohmann et al. [164] has inconsistencies and not all details are specified in their paper. Therefore the following assumptions have been made for the implementation:

- The usage of their coordinate system is inconsistent: While they originally define y-axis as walking direction and z-axis as upward vector, walking direction is later referred to as z-axis and upward direction is referred to as y-axis during their definition of gait events. Therefore y-axis and z-axis are interchanged, as it would not make sense otherwise. This concerns the detection of the events *start moving* M_s, *end uprising* D_s, *start walking* W_s, *start accelerating* A_s, *end accelerating* A_e, *end moving* M_e and start lowering D_s. The usage of x-axis is consistent with their original coordinate system definition.

- For the events *start rotating* and *end rotating* minimum and maximum are interchanged. This is assumed to be a mistake in the paper, since the derivation of a trough is trough followed by a peak, not the other way around.

- Both *end uprising* and *start lowering* are referred to as D_s. This is assumed to be a mistake in the paper and U_e is estimated based on the definition given in the description for the *end uprising* event.

- The definition of *end accelerating* A_e is equivalent with the definition of *start accelerating* A_s, *end accelerating* is even referred to as A_s in the paper. This is assumed to be a mistake in the paper and the definition for *end accelerating* is assumed to be $A_e = tmin(sc_y, T_m, S_e, m)$.

- No signals filters are specified, therefore the same filters have been applied as in the proposed method.
- Thresholds l_1, l_2 and l_3 are not specified and set to $\frac{3}{4}\sigma_{ss}$ with the exception of the threshold for the *end uprising* event U_e, which was set to $\frac{3}{20}\sigma_{ss}$. The only thresholds that is specified was the mean walking speed used for the detection of *start accelerating* A_s and *end accelerating* A_e.
- Shoulder center trajectory sc is replaced with spine shoulder trajectory ss. This is necessary since the authors used Kinect v1 and the skeletal model differs slightly compared to Kinect v2.
- The equivalent events for the start and end of turn #2 T_s^2 and T_e^2 are not defined in their paper so these events are not implemented.

4.3.2 TUG analysis results

The proposed approach based on skeleton data is referred to as Skeleton TUG and the proposed approach based on depth data as Depth TUG.

Figure 4.12 shows the result of Skeleton TUG, Depth TUG and sTUG for participant #1. Detected events are illustrated in the top left for Depth TUG, in the top right for Skeleton TUG and in the bottom for sTUG. TUG time is computed as the elapsed time between start of chair rise C_s^1 (start moving M_s for sTUG) and the end of sitting down C_e^2 (end moving M_e for sTUG). It can be seen that Depth TUG, Skeleton TUG and sTUG detect both the start and end event within a ± 1 second deviation from the ground truth for this participant with Skeleton TUG detecting these events the most accurately. For the remaining events, sTUG also detects them within a ± 1 second deviation. For Depth TUG, the end of the second walk W_e^2 and the detection of the start and end of the second turn are detected too late. For Skeleton TUG, start sitting down C_s^2 is detected significantly too early.

Figure 4.13 shows the results of the obtained TUG times, manually labeled ground truth and stopwatch in seconds for all 11 participants. The most notable outliers are participants #4, #5 and #7 for the sTUG method. The average error for TUG time is 0.294 s for Depth TUG, 0.227 s for Skeleton TUG, 0.536 s for the stopwatch and 2.549 s for sTUG.

Table 4.8 shows the mean absolute error for TUG time and the 6 TUG phases. Precision and recall for total TUG time and TUG phases are computed based on the overlap of the groundtruth interval and the detected interval. The resulting F1 scores are 0.824 for Depth TUG, 0.857 for Skeleton TUG and 0.695 for sTUG.

The mean absolute error of single TUG events and its standard deviation are shown in Table 4.9. The same data is also illustrated in Figure 4.14. From the 10 TUG events extracted from the three approaches, only the end of the first walking sequence W_e^1 has a lower error for sTUG compared to skeleton TUG and depth TUG. For all other events

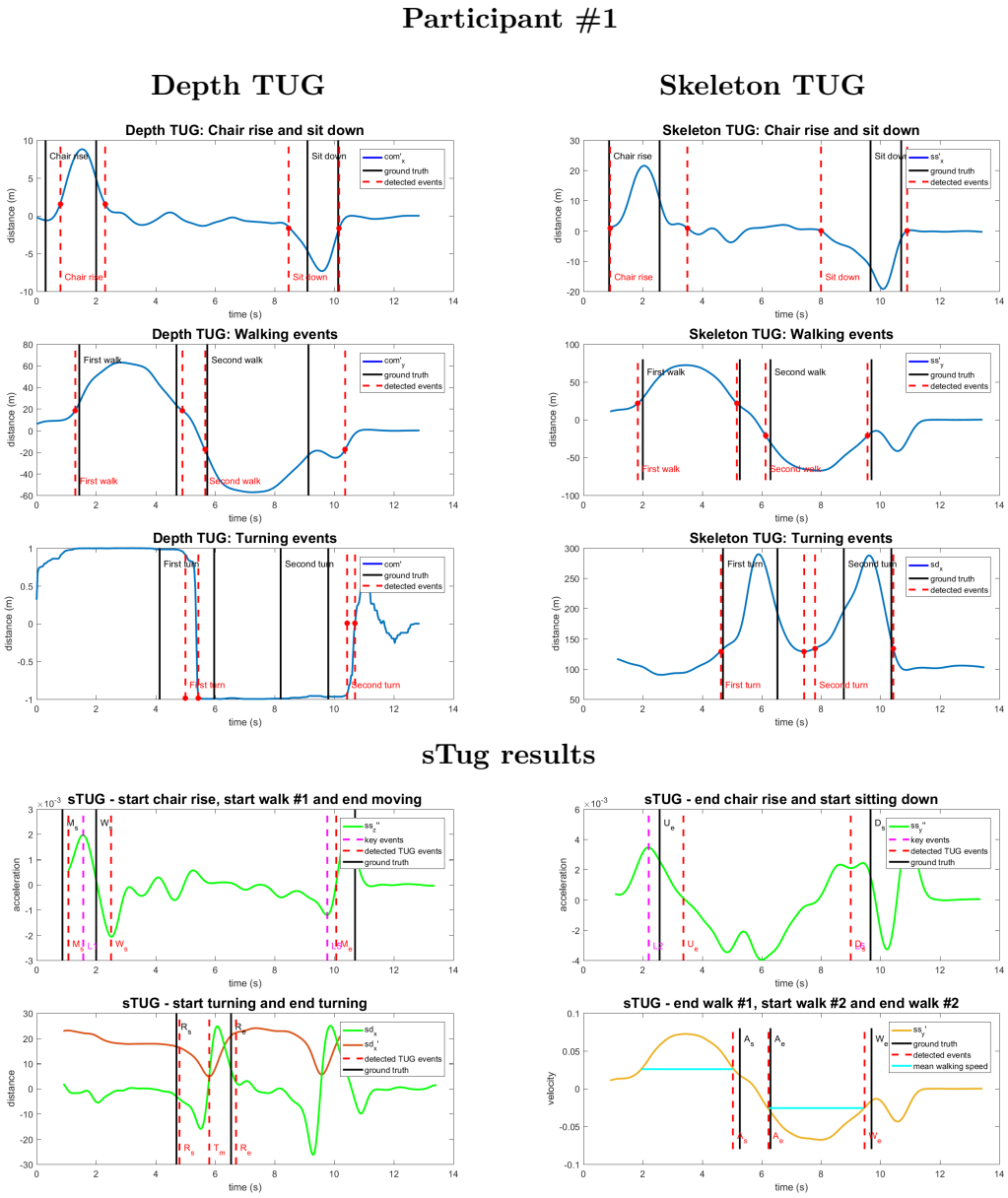


Figure 4.12: Detection of TUG events for participant #1 based on Depth TUG, Skeleton TUG and sTUG. Top left: Depth TUG results. Top right: Skeleton TUG results. Bottom: sTUG results.

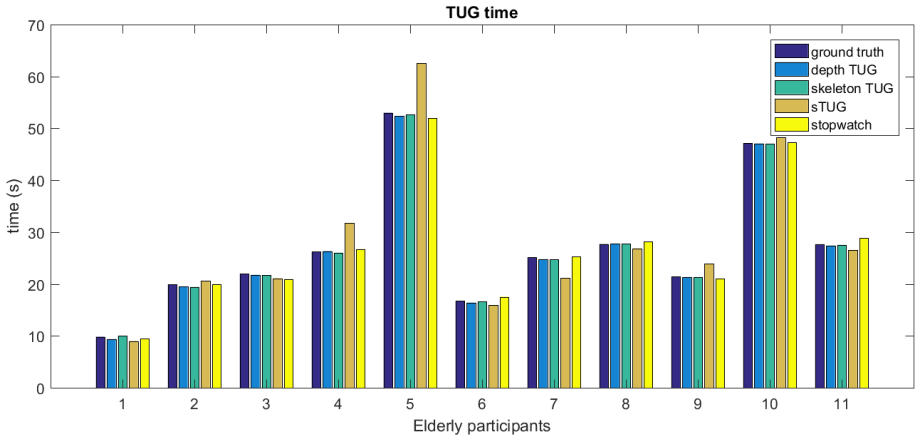


Figure 4.13: Comparison of TUG times obtained from depth TUG, skeleton TUG and sTUG with ground truth and stopwatch results for all eleven participants.

Mean absolute error, precision and recall of TUG phases

		TUG time	Chair rise time	Walk #1 time	Turn #1 time	Walk #2 time	Turn #2 time	Sit down time
Depth TUG	mean error (s)	0.294	0.182	1.324	1.227	2.364	2.222	0.730
	Precision	0.998	0.874	0.900	0.929	0.729	0.467	0.783
	Recall	0.986	0.877	0.938	0.65	0.995	0.473	0.943
Skeleton TUG	mean error (s)	0.227	1.024	0.903	1.061	1.224	2.182	1.570
	Precision	0.997	0.647	0.961	0.793	0.831	0.832	0.593
	Recall	0.990	0.928	0.906	0.871	0.983	0.759	0.952
sTUG	mean error (s)	2.503	2.003	-	1.342	-	-	2.097
	Precision	0.941	0.294	0.809	0.769	0.719	-	0.438
	Recall	0.947	0.574	0.961	0.772	0.790	-	0.345

Table 4.8: Mean absolute error, precision and recall for TUG time and the six TUG phases obtained from depth TUG, skeleton TUG and sTUG [164].

both the mean error and standard deviation are similar or higher than Depth TUG or Skeleton TUG.

Table 4.10 shows the average difference between detected TUG events and the corresponding ground truth. It shows that the majority of TUG events are detected too early for sTUG. For Depth TUG, the only event which has a clear bias is the end of walk #2 W_e^2, which is consistently detected too late. For Skeleton TUG, the end of chair rise C_e^1 and the end of walk #2 W_e^2 are detected too late by 1 second on average. The start of the sit down event C_s^2 is detected too early by around 1.5 seconds.

Absolute mean error (s) and its standard deviation for 12 TUG events

	Depth TUG	Skeleton TUG	sTUG
Chair rise start	0.22 ± 0.16	0.16 ± 0.13	2.05 ± 3.80
Chair rise end	0.26 ± 0.18	1.10 ± 0.65	3.58 ± 4.42
Walk #1 start	0.65 ± 1.13	0.49 ± 0.89	2.945.52
Walk #1 end	0.89 ± 1.01	0.68 ± 1.03	0.46 ± 0.39
Turn #1 start	0.79 ± 0.40	0.61 ± 0.72	1.31 ± 2.87
Turn #1 end	0.53 ± 0.39	0.82 ± 0.62	0.79 ± 0.86
Walk #2 start	0.35 ± 0.31	0.37 ± 0.25	0.77 ± 0.83
Walk #2 end	2.08 ± 1.57	1.04 ± 0.80	2.01 ± 3.27
Turn #2 start	1.64 ± 1.46	1.33 ± 1.38	-
Turn #2 end	3.02 ± 1.87	1.13 ± 1.52	-
Sit down start	0.70 ± 0.64	1.61 ± 1.15	1.27 ± 1.07
Sit down end	0.094 ± 0.089	0.13 ± 0.08	1.51 ± 1.07

Table 4.9: Absolute mean error and standard deviation in seconds for all detected TUG events

Average signed difference (s) of 10 detected TUG events

	C_s^1	C_e^1	W_s^1	W_e^1	T_s^1	T_e^1	W_s^2	W_e^2	C_s^2	C_e^2
Depth TUG	0.21	0.23	0.04	0.22	0.52	-0.08	-0.28	2.08	-0.45	-0.05
Skeleton TUG	0.1	1.1	0.43	-0.19	-0.08	0.47	-0.17	1.01	-1.57	-0.10
sTUG [164]	-1.81	-0.84	-2.83	0.18	-1.02	-0.53	-0.52	-1.12	0.12	-0.84

Table 4.10: Average signed difference of detected TUG event times and ground truth in seconds. The following abbreviations have been used: Chair rise start C_s^1 and end C_e^1, walk #1 start W_s^1 and end W_e^1, turn #1 start T_s^1 and end T_e^1 and sit down start C_s^2 and end C_e^2.

4.3.3 TUG analysis discussion

The average time for sit-to-stand is 1.7 s, for walk #1 3.3 s, for turn #1 1.8 s, for walk #2 3.4 s, for turn #2 1.6 s and for sitting down 1.0 s. This observation is consistent with the study by Manckoundia et al. [170], who show that sit-to-stand takes longer than stand-to-sit for elderly subjects.

One reason for the poor performance of sTUG on this dataset is that their approach works on the assumption, that the subject does not move before and after the TUG. However, this assumption is not necessarily true when working with elderly subjects, who often need multiple attempts to perform a certain movement or additional instruction

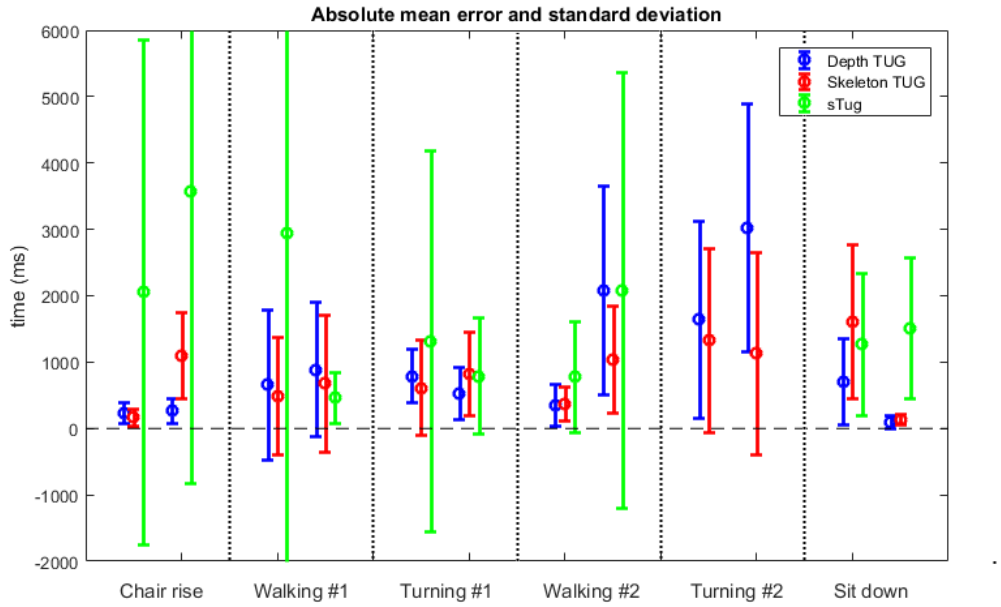


Figure 4.14: Absolute mean error and standard deviation in ms for 12 TUG events. The dotted line separates the five gait phases. For each phase, the first three values represent the error of the start event of the phase and the last three values represent the error of the end event of the phase.

before rising from the chair. Moreover, Kinect skeleton tracking is noisy and not reliable when the subject is sitting. This commonly causes sTUG to detect the *start moving* event too early. Two examples for this are the movement of participant #4 and participant #5, as shown in Figure 4.15 and Figure 4.16. The detection of the start chair rise events M_s for sTUG is depicted in the top left from the bottom part of the figures. It can be seen that there are a lot of acceleration spikes at the beginning, but the participant only starts to rise from the chair at approximately 12 seconds for participant #4 and approximately 20 seconds for participant #5. Therefore joint acceleration as a sole feature as used by STUG is not robust enough for the detection of TUG events.

Moreover, sTUG detects the beginning of TUG as the point where the subject starts to lean forward to move his center of gravity over his feet, followed by the subject lifting the body from the chair. However, this approach does not consider that the subject might move his upper body back and forth prior to starting the TUG. Therefore it is necessary to verify that the subject stands up from the chair shortly after leaning forward. An approach to achieve this is the cubic space approach used by Kitsunezaki et al. [145]. They count the number of joints inside a virtual cubic space around the chair in order to find the transition from sitting on the chair to standing up and walking away from the chair.

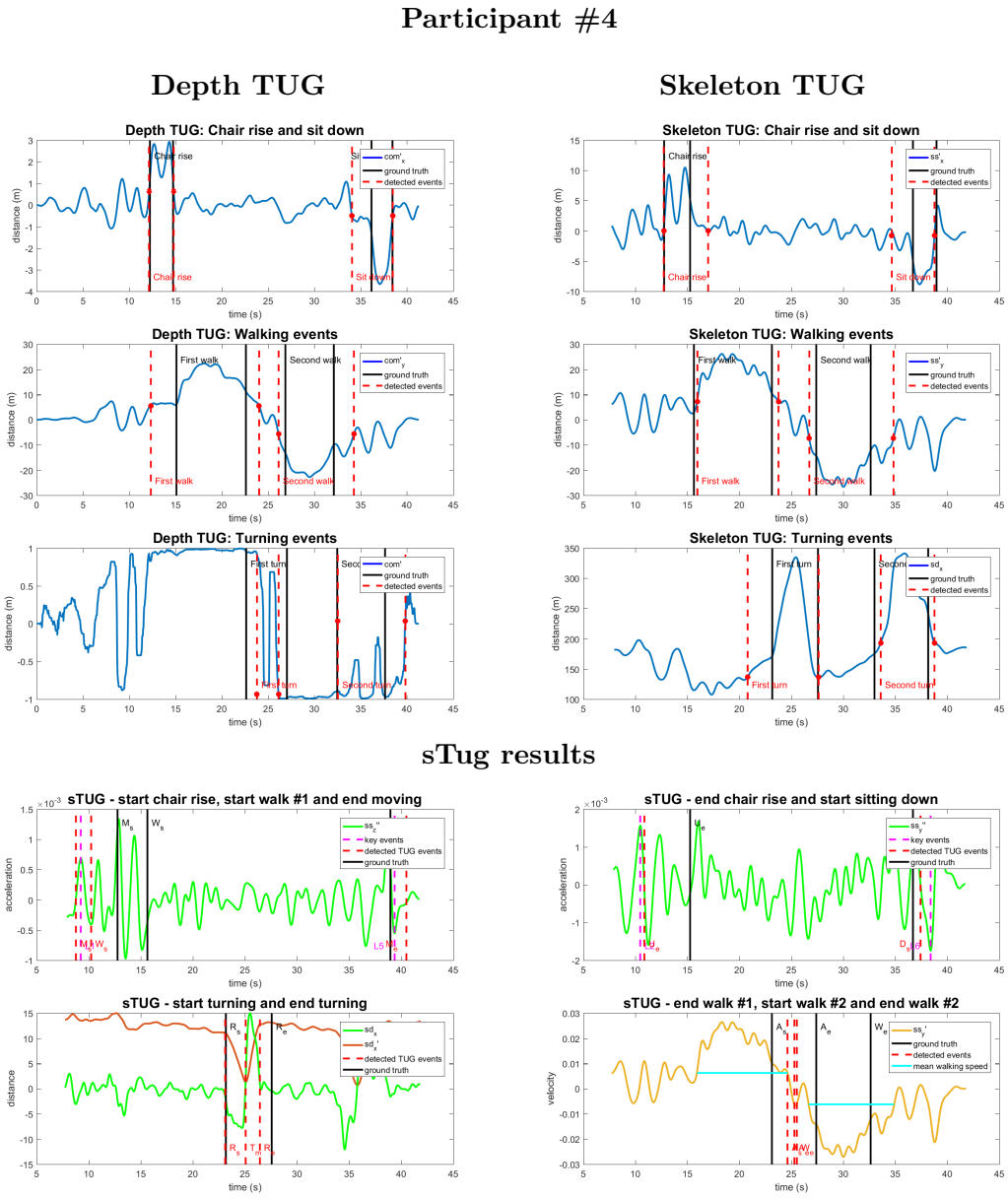


Figure 4.15: Detection of TUG events for participant #4 based on Depth TUG, Skeleton TUG and sTUG. Top left: Depth TUG results. Top right: Skeleton TUG results. Bottom: sTUG results.

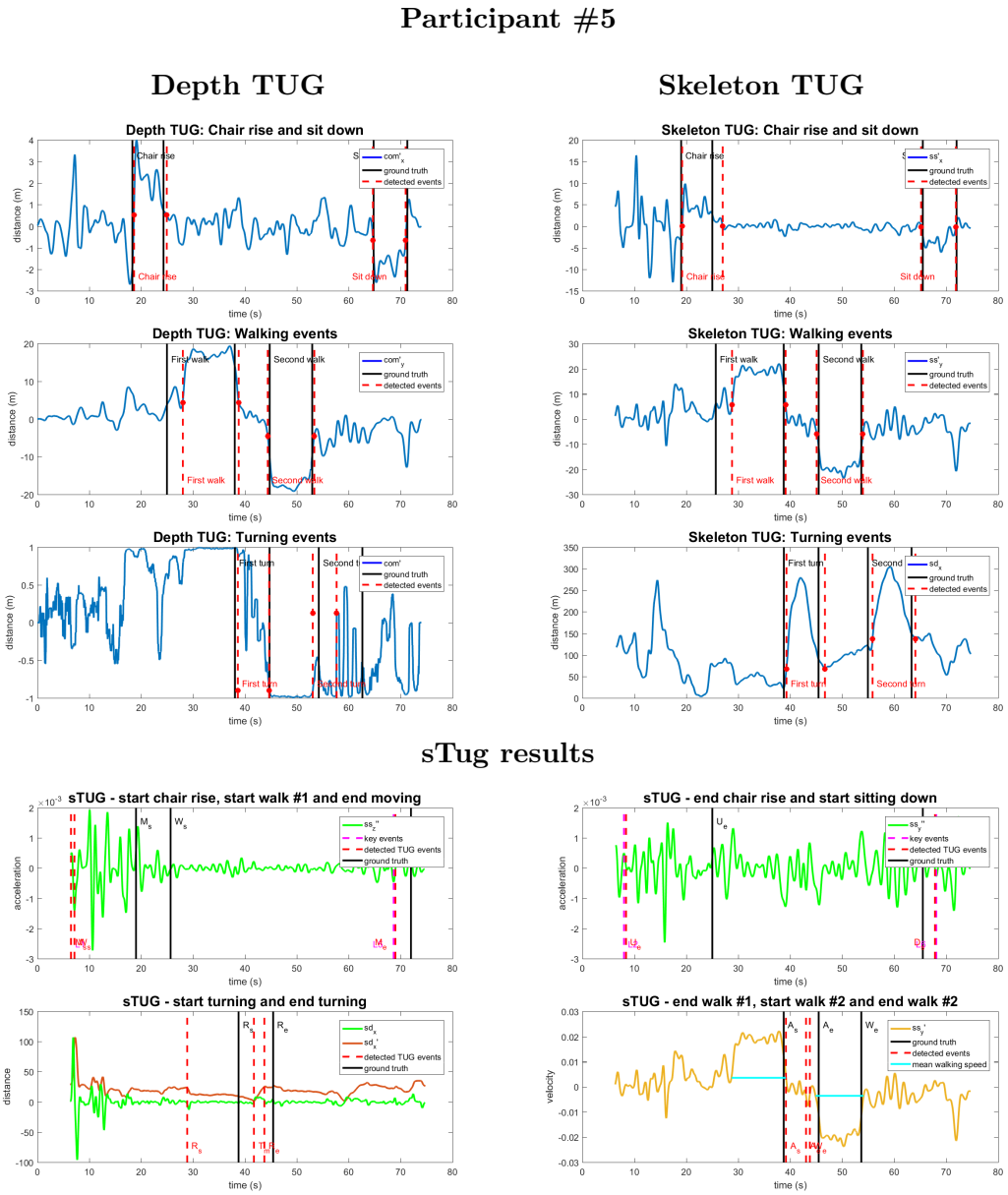
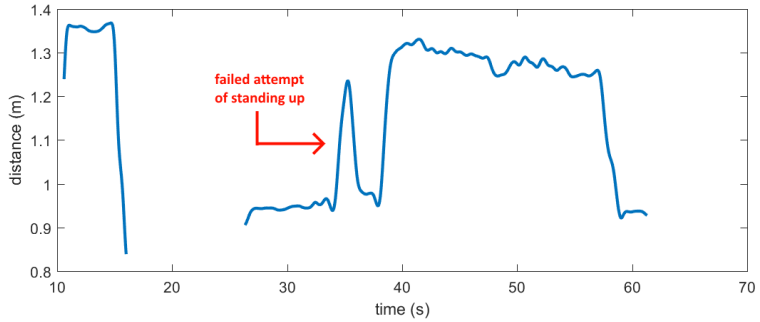
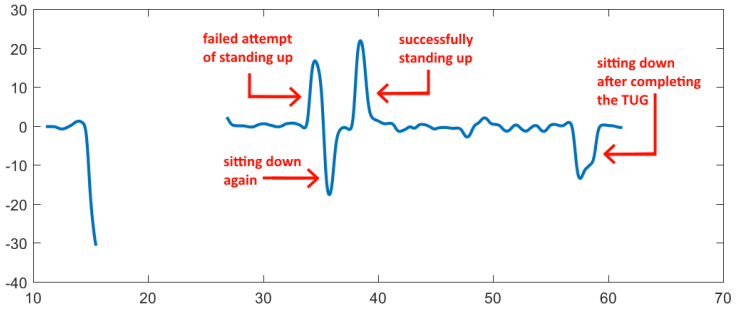


Figure 4.16: Detection of TUG events for participant #5 based on Depth TUG, Skeleton TUG and sTUG. Top left: Depth TUG results. Top right: Skeleton TUG results. Bottom: sTUG results.



(a) *Blue line:* Position of spine shoulder joint in y-direction (height).



(b) *Blue line:* Velocity of spine shoulder joint in y-direction (upwards).

Figure 4.17: An older person needing two attempts of standing up. The failed attempt can be easily identified in the joint height trajectory (top image), but not in its derivative (bottom image).

Another reason why sTUG tends to identify the *start moving* event too early is that an examiner starts the stopwatch only when the patient's hips are no longer touching the chair [145] instead of when the patient is leaning forward. As pointed out by Kitsunezaki et al. [145], it is challenging to detect standing up or sitting down movements due to the very slow movements of old adults. Even the proposed methods are probably not robust enough to perform well on a larger dataset. One option to increase the robustness of the approach is to additionally consider the trajectory of the joint position in addition to the velocity. By looking at the current position of the subject, the current TUG state can be verified and certain state transitions are only allowed if the subject is in the correct position. Moreover, this can help to resolve cases like the one depicted in Figure 4.17. One subject attempts to rise from the chair, but is unsuccessful and sits down again before making a second attempt. When only a derivative of the joint trajectory is considered, the resulting peaks can easily be falsely classified.

In addition to the increased robustness, future methods should also consider the additional challenges present in a real-world setting. It is common during a TUG test that a second person is walking around, e.g. a medical assistant being nearby to assist the patient if

necessary [145]. This creates additional challenges for consistently tracking the subject and a second sensor might be be necessary to consistently resolve this problem. Moreover, it is common for older adults to execute the TUG incorrectly. Future methods should also be able to determine whether the TUG is executed correctly or whether certain movements are missed and the trial should be repeated.

CHAPTER 5

Conclusion

In this thesis, a Kinect v2 sensor was used to assess mobility based on both human gait and automatized Timed Up & Go (TUG) tests. For both types of mobility assessment an approach based on depth data and an approach based on skeleton data was presented. These two approaches have been compared according to the error of the retrieved parameters.

Three Kinect datasets were acquired. The first dataset features gait of healthy adults. Thirteen participants aged 24 to 77 years walked twice over 9 sets of predefined floor markers to accumulate a gait dataset of 234 walking sequences. Toe-off and heel-strike events were manually labeled, which allowed evaluation not just for spatial but also for temporal gait parameters. The second dataset features elderly gait. It consists of 22 walking sequences of 11 elderly subjects, 85 years and older. The number of steps was manually labeled for each walking sequence. The third dataset consists of 11 TUG tests from the same subjects who participated for the elderly gait dataset. The start and end of six TUG phases were manually labeled. These TUG phases are chair rise, walking away from the chair, turning, walking towards the chair, turning before sitting down and sitting down. All three acquired datasets are available online.

In the first part of this work gait analysis approaches based on both Kinect depth and skeleton data were proposed. The gait analysis approaches were based on gait signals, which were obtained from the movement of the lower limbs, e.g. the correlation coefficient time series or the time series of the feet distance. Five spatiotemporal gait parameters were extracted from each gait signal, gait speed, step length, stride length, step time and stride time. Evaluation on both - healthy and elderly - gait datasets showed that the correlation coefficient time series based on depth data performed the best. Moreover, it was the most robust towards external influences, e.g. walking direction, use of a walking aid, wearing a floor-length skirt, etc. From the assessed spatiotemporal parameters, gait speed was the most accurate parameter, followed by stride parameters and step parameters. This was expected and consistent with previous studies.

The experiments in this work showed, that state-of-the-art gait analysis approaches that work well with healthy adults in a laboratory setting may not work well for the gait of older adults. Gait from older adults in a real-world setting is more challenging for a number of reasons, e.g. the potential use of walking aids, small step lengths, slow gait speeds, etc. The acquired elderly gait dataset can help in evaluating future gait analysis approaches to be more robust towards the gait of older adults. However, due to the limited size of the acquired elderly gait dataset, a larger gait dataset of older adults is desirable for future works.

The choice between using the raw depth data and the skeleton model of Kinect should be made based on the complexity of the required parameters. For mobility parameters that depend on several body limbs or joints, Kinect skeleton tracking is preferable over raw depth data as it provides a reliable, efficient and accurate pose estimation. However, based on related literature, frailty is linked to simple gait parameters such as gait speed and step length. Depth data should be favored over skeleton tracking for assessing these simple gait parameters for a number of reasons. First, it is independent of the walking direction of the subject. Second, it has less range limitation, which increases the potential field of view of the sensor. Third, it is independent of the subject's pose. And fourth, it is more robust towards partial occlusions, which may cause the skeleton tracking to fail.

In the second part of this work, approaches for the analysis of the Timed Up & Go (TUG) test were presented and evaluated on the acquired elderly TUG test dataset. Both depth and skeleton data were used to automatically assess TUG time and the start and end of six TUG phases. TUG time and TUG events were assessed by analyzing the movement of the subject based on either the center shoulder joint velocity or the centroid velocity. The approach based on skeleton data performed slightly superior compared to the approach based on depth data regarding the error of detected TUG times and TUG phases. However, both approaches showed significantly lower error rates for the TUG time when compared with the time obtained from a manual stopwatch. Moreover, the two approaches have been further evaluated with a previously presented approach and superior results were obtained not just for the detection of the TUG time, but also for the detection of TUG events.

The experiments in this work showed, that the TUG test can be automatically analyzed by a single Kinect sensor. Moreover, they showed that TUG analysis based on joint acceleration as a sole feature is not able to robustly measure TUG times and detect TUG events. It was shown that the proposed approach worked well on the acquired TUG test dataset. However, it may not be robust enough for future real-world datasets that include failed movement attempts or other atypical behavior. Potential improvements of the proposed approach should consider the position of the subject in relation to the chair instead of solely considering the movement of the subject. This ensures that the movement state of the subject can be continuously tracked and mistakes - caused by the subject's movement being different than expected - can be avoided. Furthermore, a large, manually labeled TUG test dataset is needed for the evaluation of future TUG automatization approaches.

To sum up, it was shown that human mobility analysis based on Kinect has the potential to assess frailty in older adults.

Appendix

A Translated FRAIL scale questionnaire in German

F.R.A.I.L – Fragebogen 17. November 2016

Alter:

Geschlecht: [ ] weiblich [ ] männlich

Wie oft während der letzten 4 Wochen haben Sie sich müde gefühlt?
Mögliche Antworten: Dauernd, oft, manchmal, selten, nie.

[ ] Dauernd [ ] Oft [ ] Manchmal [ ] Selten [ ] Nie

Haben Sie irgendwelche Schwierigkeiten, alleine und ohne Gehhilfe 10 Stufen einer Stiege zu steigen ohne zu pausieren?

[ ] Ja [ ] Nein

Haben Sie irgendwelche Schwierigkeiten, alleine und ohne Gehhilfe einige Hundert Meter zu gehen?

[ ] Ja [ ] Nein

Hat Ihnen Ihr Doktor gesagt, dass Sie eine oder mehrere der folgenden Krankheiten haben?

Bluthochdruck?	[ ] Ja	[ ] Nein
Diabetes oder Zuckerkrankheit?	[ ] Ja	[ ] Nein
Krebs, ausgenommen ein kleiner Hautkrebs?	[ ] Ja	[ ] Nein
Chronische Lungenerkrankung?	[ ] Ja	[ ] Nein
Herzinfarkt?	[ ] Ja	[ ] Nein
Herzinsuffizienz?	[ ] Ja	[ ] Nein
Angina? Bzw. Brust- oder Herzenge?	[ ] Ja	[ ] Nein
Asthma?	[ ] Ja	[ ] Nein
Arthritis oder Gelenkentzündung?	[ ] Ja	[ ] Nein
Schlaganfall?	[ ] Ja	[ ] Nein
Nierenerkrankung?	[ ] Ja	[ ] Nein

Wie viel wiegen Sie inklusive Kleidung aber ohne Schuhe? Gewicht: [ ] kg

Wie viel wogen Sie vor einem Jahr inklusive Kleidung aber ohne Schuhe? Gewicht: [ ] kg

Bibliography

[1] W. Z. W. Z. Abiddin, R. Jailani, A. R. Omar, and I. M. Yassin. Development of matlab kinect skeletal tracking system (mksts) for gait analysis. In *IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, pages 216–220. IEEE, 2016.

[2] J. K. Aggarwal and Q. Cai. Human motion analysis: A review. In *Nonrigid and Articulated Motion Workshop, 1997. Proceedings., IEEE*, pages 90–102. IEEE, 1997.

[3] J. K. Aggarwal, Q. Cai, W. Liao, and B. Sabata. Articulated and elastic non-rigid motion: A review. In *Motion of Non-Rigid and Articulated Objects, 1994., Proceedings of the 1994 IEEE Workshop on*, pages 2–14. IEEE, 1994.

[4] J. K. Aggarwal and L. Xia. Human activity recognition from 3d data: A review. *Pattern Recognition Letters*, 48:70–80, 2014.

[5] N. Agoulmine, M. J. Deen, J.-S. Lee, and M. Meyyappan. U-health smart home. *IEEE Nanotechnology Magazine*, 5(3):6–11, 2011.

[6] F. Ahmed, P. Polash Paul, and M. L. Gavrilova. Kinect-based gait recognition using sequences of the most relevant joint relative angles. *WSCG 2015*, 23(2), 2015.

[7] C. Amon, F. Fuhrmann, and F. Graf. Evaluation of the spatial resolution accuracy of the face tracking system for kinect for windows v1 and v2. In *Proceedings of the 6th Congress of the Alps Adria Acoustics Association*, pages 16–17, 2014.

[8] V. O. Andersson and R. M. de Araújo. Person identification using anthropometric and gait data from kinect sensor. In *AAAI*, pages 425–431, 2015.

[9] E. Auvinet, F. Multon, C.-E. Aubin, J. Meunier, and M. Raison. Detection of gait cycles in treadmill walking using a kinect. *Gait & posture*, 41(2):722–725, 2015.

[10] E. Auvinet, F. Multon, and J. Meunier. Gait analysis with multiple depth cameras. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6265–6268. IEEE, 2011.

[11] E. Auvinet, F. Multon, and J. Meunier. Lower limb movement asymmetry measurement with a depth camera. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6793–6796. IEEE, 2012.

[12] E. Auvinet, F. Multon, and J. Meunier. New lower-limb gait asymmetry indices based on a depth camera. *Sensors*, 15(3):4605–4623, 2015.

[13] C. Bachmann, H. Gerber, and A. Stacoff. Messsysteme, messmethoden und beispiele zur instrumentierten ganganalyse. *Schweizerische Zeitschrift fur Sportmedizin und Sporttraumatologie*, 56(2):29, 2008.

[14] G. Baldewijns, G. Verheyden, B. Vanrumste, and T. Croonenborghs. Validation of the kinect for gait analysis using the gaitrite walkway. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5920–5923. IEEE, 2014.

[15] A. Ball, D. Rye, F. Ramos, and M. Velonaki. Unsupervised clustering of people from'skeleton'data. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 225–226. ACM, 2012.

[16] S. J. M. Bamberg, A. Y. Benbasat, D. M. Scarborough, D. E. Krebs, and J. A. Paradiso. Gait analysis using a shoe-integrated wireless sensor system. *IEEE transactions on information technology in biomedicine*, 12(4):413–423, 2008.

[17] T. Banerjee, J. M. Keller, and M. Skubic. Resident identification using kinect depth image data and fuzzy clustering techniques. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5102–5105. IEEE, 2012.

[18] Y. Barak, R. C. Wagenaar, and K. G. Holt. Gait characteristics of elderly people with a history of falls: a dynamic approach. *Physical therapy*, 86(11):1501–1510, 2006.

[19] K. Bashir, T. Xiang, and S. Gong. Feature selection on gait energy image for human identification. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008).*, pages 985–988. IEEE, 2008.

[20] S. Bauer, A. Seitel, H. Hofmann, T. Blum, J. Wasza, M. Balda, H.-P. Meinzer, N. Navab, J. Hornegger, and L. Maier-Hein. Real-time range imaging in health care: a survey. In *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications*, pages 228–254. Springer, 2013.

[21] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE transactions on pattern analysis and machine intelligence*, 24(4):509–522, 2002.

[22] C. BenAbdelkader, R. Cutler, and L. Davis. Stride and cadence as a biometric in automatic person identification and verification. In *Fifth IEEE International Conference on Automatic Face and Gesture Recognition, 2002. Proceedings.*, pages 372–377. IEEE, 2002.

[23] C. Benedek, B. Gálai, B. Nagy, and Z. Jankó. Lidar-based gait analysis and activity recognition in a 4d surveillance system. *IEEE Transactions on Circuits and Systems for Video Technology*, 2016.

[24] C. Benedek, D. Molnár, and T. Szirányi. A dynamic mrf model for foreground detection on range data sequences of rotating multi-beam lidar. In *Advances in Depth Image Analysis and Applications*, pages 87–96. Springer, 2013.

[25] P. Bera, R. Kar, and A. Konar. Joint pain detection by gait analysis for elderly healthcare. In *2015 IEEE International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, pages 220–224. IEEE, 2015.

[26] Y. Berdnikov and D. Vatolin. Real-time depth map occlusion filling and scene background restoration for projected-pattern based depth cameras. In *Graphic Conf., IETP*, 2011.

[27] K. O. Berg, B. E. Maki, J. I. Williams, P. J. Holliday, and S. L. Wood-Dauphinee. Clinical and laboratory measures of postural balance in an elderly population. *Archives of physical medicine and rehabilitation*, 73(11):1073–1080, 1992.

[28] K. Berger, S. Meister, R. Nair, and D. Kondermann. A state of the art report on kinect sensor setups in computer vision. In *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications*, pages 257–272. Springer, 2013.

[29] D. Berrada, M. Romero, G. Abowd, M. Blount, and J. Davis. Automatic administration of the get up and go test. In *Proceedings of the 1st ACM SIGMOBILE international workshop on Systems and networking support for healthcare and assisted living environments*, pages 73–75. ACM, 2007.

[30] R. Best and R. Begg. Overview of movement analysis and gait features. *Computational intelligence for movement sciences: neural networks and other emerging techniques*, 1:1–69, 2006.

[31] B. Bilney, M. Morris, and K. Webster. Concurrent related validity of the gaitrite® walkway system for quantification of the spatial and temporal parameters of gait. *Gait & posture*, 17(1):68–74, 2003.

[32] A. F. Bobick and A. Y. Johnson. Gait recognition using static, activity-specific parameters. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. CVPR 2001.*, volume 1, pages I–I. IEEE, 2001.

[33] B. Bonnechère, B. Jansen, P. Salvia, H. Bouzahouene, L. Omelina, J. Cornelis, M. Rooze, and S. V. S. Jan. Can the kinect^{TM} sensors be used for motion analysis? *Transaction on Electrical and Electronic Circuits and Systems*, 4(1), 2013.

[34] B. Bonnechere, B. Jansen, P. Salvia, H. Bouzahouene, L. Omelina, J. Cornelis, M. Rooze, and S. Van Sint Jan. What are the current limits of the kinect sensor. In *Proc 9th Intl Conf. Disability, Virutal Reality & Associated Technologies, Laval, France*, pages 287–294, 2012.

[35] R. Borràs, À. Lapedriza, and L. Igual. Depth information in human gait analysis: an experimental study on gender recognition. In *International Conference Image Analysis and Recognition*, pages 98–105. Springer, 2012.

[36] A. Bottino and A. Laurentini. A silhouette based technique for the reconstruction of human movement. *Computer Vision and Image Understanding*, 83(1):79–95, 2001.

[37] K. Bouillon, M. Kivimaki, M. Hamer, S. Sabia, E. I. Fransson, A. Singh-Manoux, C. R. Gale, and G. D. Batty. Measures of frailty in population-based studies: an overview. *BMC geriatrics*, 13(1):64, 2013.

[38] M. Brand. Shadow puppetry. In *The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999.*, volume 2, pages 1237–1244. IEEE, 1999.

[39] C. Bregler. Motion capture technology for entertainment [in the spotlight]. *IEEE Signal Processing Magazine*, 24(6):160–158, 2007.

[40] F. Buckinx, Y. Rolland, J.-Y. Reginster, C. Ricour, J. Petermans, and O. Bruyère. Burden of frailty in the elderly population: perspectives for a public health challenge. *Archives of Public Health*, 73(1):19, 2015.

[41] Z. Cai, J. Han, L. Liu, and L. Shao. Rgb-d datasets using microsoft kinect or similar sensors: a survey. *Multimedia Tools and Applications*, 76(3):4313–4355, 2017.

[42] F. Caillette and T. Howard. Real-time markerless human body tracking with multi-view 3-d voxel reconstruction. In *Proc. BMVC*, volume 2, pages 597–606, 2004.

[43] A. J. Campbell and D. M. Buchner. Unstable disability and the fluctuations of frailty. *Age and ageing*, 26(4):315–318, 1997.

[44] M. Camplani and L. Salgado. Efficient spatio-temporal hole filling strategy for kinect depth maps. In *IS&T/SPIE Electronic Imaging*, pages 82900E–82900E. International Society for Optics and Photonics, 2012.

[45] J. Carranza, C. Theobalt, M. A. Magnor, and H.-P. Seidel. Free-viewpoint video of human actors. In *ACM SIGGRAPH 2003 Papers*, SIGGRAPH '03, pages 569–577, New York, NY, USA, 2003. ACM.

[46] C. Cedras and M. Shah. Motion-based recognition a survey. *Image and Vision Computing*, 13(2):129–155, 1995.

[47] M. Cesari, G. Gambassi, G. Abellan van Kan, and B. Vellas. The frailty phenotype and the frailty index: different instruments for different purposes. *Age and ageing*, 43(1):10–12, 2013.

[48] M. Cesari, S. B. Kritchevsky, B. W. Penninx, B. J. Nicklas, E. M. Simonsick, A. B. Newman, F. A. Tylavsky, J. S. Brach, S. Satterfield, D. C. Bauer, et al. Prognostic value of usual gait speed in well-functioning older people—results from the health, aging and body composition study. *Journal of the American Geriatrics Society*, 53(10):1675–1680, 2005.

[49] A. A. Chaaraoui, J. R. Padilla-López, and F. Flórez-Revuelta. Abnormal gait detection with rgb-d devices using joint motion history features. In *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*, volume 7, pages 1–6. IEEE, 2015.

[50] C.-Y. Chang, B. Lange, M. Zhang, S. Koenig, P. Requejo, N. Somboon, A. A. Sawchuk, and A. A. Rizzo. Towards pervasive physical rehabilitation using microsoft kinect. In *2012 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, pages 159–162. IEEE, 2012.

[51] R. Chellappa, A. Veeraraghavan, and N. Ramanathan. Gait biometrics, overview. *Encyclopedia of Biometrics*, pages 783–789, 2015.

[52] L. Chen, H. Wei, and J. Ferryman. A survey of human motion analysis using depth imagery. *Pattern Recognition Letters*, 34(15):1995–2006, 2013.

[53] K. Cheung, S. Baker, and T. Kanade. Shape-from-silhouette of articulated objects and its use for human body kinematics estimation and motion capture. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, volume 1, pages I–I. IEEE, 2003.

[54] E. Cippitelli, S. Gasparrini, E. Gambi, S. Spinsante, J. Wåhslény, I. Orhany, and T. Lindhy. Time synchronization and data fusion for rgb-depth cameras and inertial sensors in aal applications. In *2015 IEEE International Conference on Communication Workshop (ICCW)*, pages 265–270. IEEE, 2015.

[55] E. Cippitelli, S. Gasparrini, S. Spinsante, and E. Gambi. Kinect as a tool for gait analysis: validation of a real-time joint extraction algorithm working in side view. *Sensors*, 15(1):1417–1434, 2015.

[56] R. A. Clark, K. J. Bower, B. F. Mentiplay, K. Paterson, and Y.-H. Pua. Concurrent validity of the microsoft kinect for assessment of spatiotemporal gait variables. *Journal of biomechanics*, 46(15):2722–2725, 2013.

[57] R. A. Clark, Y.-H. Pua, K. Fortin, C. Ritchie, K. E. Webster, L. Denehy, and A. L. Bryant. Validity of the microsoft kinect for assessment of postural control. *Gait & posture*, 36(3):372–377, 2012.

[58] R. A. Clark, S. Vernon, B. F. Mentiplay, K. J. Miller, J. L. McGinley, Y. H. Pua, K. Paterson, and K. J. Bower. Instrumenting gait assessment using the kinect in people living with stroke: reliability and association with balance tests. *Journal of neuroengineering and rehabilitation*, 12(1):1, 2015.

[59] A. Clegg, J. Young, S. Iliffe, M. O. Rikkert, and K. Rockwood. Frailty in elderly people. *The Lancet*, 381(9868):752–762, 2013.

[60] M. J. Cree, L. V. Streeter, R. M. Conroy, and A. A. Dorrington. Analysis of the softkinetic depthsense for range imaging. In *International Conference Image Analysis and Recognition*, pages 668–675. Springer, 2013.

[61] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati. Detecting moving objects, ghosts, and shadows in video streams. *IEEE transactions on pattern analysis and machine intelligence*, 25(10):1337–1342, 2003.

[62] D. Cunado, M. Nixon, and J. Carter. Using gait as a biometric, via phase-weighted magnitude spectra. In *Audio-and Video-based Biometric Person Authentication*, pages 93–102. Springer, 1997.

[63] D. Cunado, M. S. Nixon, and J. N. Carter. Automatic gait recognition via model-based evidence gathering. Organisation: IEEE and AIM, October 1999.

[64] S. L. A. da Silva, J. U. Viana, V. G. Da Silva, J. M. D. Dias, L. S. M. Pereira, and R. C. Dias. Influence of frailty and falls on functional capacity and gait in community-dwelling elderly individuals. *Topics in Geriatric Rehabilitation*, 28(2):128–134, 2012.

[65] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893. IEEE, 2005.

[66] N.-L. Dao, Y. Zhang, J. Zheng, and J. Cai. Kinect-based non-intrusive human gait analysis and visualization. In *Multimedia Signal Processing (MMSP), 2015 IEEE 17th International Workshop on*, pages 1–6. IEEE, 2015.

[67] N. De Vries, J. Staal, C. Van Ravensberg, J. Hobbelen, M. O. Rikkert, and M. Nijhuis-Van der Sanden. Outcome instruments to measure frailty: a systematic review. *Ageing research reviews*, 10(1):104–114, 2011.

[68] J. A. DeLisa. *Gait analysis in the science of rehabilitation*, volume 2. Diane Publishing, 2000.

[69] G. Demiris, D. P. Oliver, J. Giger, M. Skubic, and M. Rantz. Older adults' privacy considerations for vision based recognition methods of eldercare applications. *Technology and Health Care*, 17(1):41, 2009.

[70] G. Demiris, M. J. Rantz, M. A. Aud, K. D. Marek, H. W. Tyrer, M. Skubic, and A. A. Hussam. Older adults' attitudes towards and perceptions of smart home technologies: a pilot study. *Medical informatics and the Internet in medicine*, 29(2):87–94, 2004.

[71] M. O. Derawi, C. Nickel, P. Bours, and C. Busch. Unobtrusive user-authentication on mobile phones using biometric gait recognition. In *2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, pages 306–311. IEEE, 2010.

[72] B. Dikovski, G. Madjarov, and D. Gjorgjevikj. Evaluation of different feature sets for gait recognition using skeletal data from kinect. In *37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 1304–1308. IEEE, 2014.

[73] E. Dolatabadi, B. Taati, and A. Mihailidis. Automated classification of pathological gait after stroke using ubiquitous sensing technology. In *IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC)*, pages 6150–6153. IEEE, 2016.

[74] E. Dolatabadi, B. Taati, and A. Mihailidis. Concurrent validity of the microsoft kinect for windows v2 for measuring spatiotemporal gait parameters. *Medical Engineering & Physics*, 38(9):952–958, 2016.

[75] E. Dolatabadi, B. Taati, G. S. Parra-Dominguez, and A. Mihailidis. A markerless motion tracking approach to understand changes in gait and balance: A case study. In *Proceedings of the Rehabilitation Engineering and Assistive Technology Society of North America Annual Conference*, pages 20–24, 2013.

[76] A. Dubois and F. Charpillet. A gait analysis method based on a depth camera for fall prevention. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4515–4518. IEEE, 2014.

[77] T. Dutta. Evaluation of the kinect$^{\text{TM}}$ sensor for 3-d kinematic measurement in the workplace. *Applied ergonomics*, 43(4):645–649, 2012.

[78] R. Elble, S. S. Thomas, C. Higgins, and J. Colliver. Stride-dependent changes in gait of older people. *Journal of neurology*, 238(1):1–5, 1991.

[79] M. Eltoukhy, J. Oh, C. Kuenze, and J. Signorile. Improved kinect-based spatiotemporal and kinematic treadmill gait assessment. *Gait & Posture*, 51:77–83, 2017.

[80] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly. Vision-based hand pose estimation: A review. *Computer Vision and Image Understanding*, 108(1):52–73, 2007.

[81] A. Fern'ndez-Baena, A. Susín, and X. Lligadas. Biomechanical validation of upper-body and lower-body joint movements of kinect motion capture data for rehabilitation treatments. In *4th International Conference on Intelligent Networking and Collaborative Systems (INCoS)*, pages 656–661. IEEE, 2012.

[82] M. Firman. Rgbd datasets: Past, present and future. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 19–31, 2016.

[83] D. A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall Professional Technical Reference, 2002.

[84] T. Frenken. *Technischer Ansatz zur unaufdringlichen Mobilitätsanalyse im Rahmen geriatrischer Assessments*. PhD thesis, University of Oldenburg, 2013.

[85] T. Frenken, B. Vester, M. Brell, and A. Hein. atug: Fully-automated timed up and go assessment using ambient sensor technologies. In *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, pages 55–62. IEEE, 2011.

[86] L. P. Fried, C. M. Tangen, J. Walston, A. B. Newman, C. Hirsch, J. Gottdiener, T. Seeman, R. Tracy, W. J. Kop, G. Burke, et al. Frailty in older adults evidence for a phenotype. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 56(3):M146–M157, 2001.

[87] H. Funaya, T. Shibata, Y. Wada, and T. Yamanaka. Accuracy assessment of kinect body tracker in instant posturography for balance disorders. In *Medical Information and Communication Technology (ISMICT), 2013 7th International Symposium on*, pages 213–217. IEEE, 2013.

[88] M. Gabel, R. Gilad-Bachrach, E. Renshaw, and A. Schuster. Full body gait analysis with kinect. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 1964–1967. IEEE, 2012.

[89] D. Gafurov, E. Snekkenes, and P. Bours. Gait authentication and identification using wearable accelerometer sensor. In *2007 IEEE Workshop on Automatic Identification Advanced Technologies*, pages 220–225. IEEE, 2007.

[90] B. Gálai and C. Benedek. Gait recognition with compact lidar sensors. In *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP)*, February 2017.

[91] D. M. Gavrila. The visual analysis of human movement: A survey. *Computer vision and image understanding*, 73(1):82–98, 1999.

[92] D. M. Gavrila and L. S. Davis. 3-d model-based tracking of humans in action: a multi-view approach. In *1996 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96*, pages 73–80. IEEE, 1996.

[93] D. J. Geerse, B. H. Coolen, and M. Roerdink. Kinematic validation of a multi-kinect v2 instrumented 10-meter walkway for quantitative gait assessments. *PloS one*, 10(10):e0139913, 2015.

[94] J. Geng. Structured-light 3d surface imaging: a tutorial. *Advances in Optics and Photonics*, 3(2):128–160, 2011.

[95] E. Gianaria, N. Balossino, M. Grangetto, and M. Lucenteforte. Gait characterization using dynamic skeleton acquisition. In *Multimedia Signal Processing (MMSP), 2013 IEEE 15th International Workshop on*, pages 440–445. IEEE, 2013.

[96] E. Gianaria, M. Grangetto, M. Roppolo, A. Mulasso, and E. Rabaglietti. Kinect-based gait analysis for automatic frailty syndrome assessment. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 1314–1318. IEEE, 2016.

[97] S. N. Gieser, V. Metsis, and F. Makedon. Quantitative evaluation of the kinect skeleton tracker for physical rehabilitation exercises. In *Proceedings of the 7th International Conference on Pervasive Technologies Related to Assistive Environments*, page 48. ACM, 2014.

[98] T. M. Gill, J. M. McGloin, E. A. Gahbauer, D. M. Shepard, and L. M. Bianco. Two recruitment strategies for a clinical trial of physically frail community-living older persons. *Journal of the American Geriatrics Society*, 49(8):1039–1045, 2001.

[99] M. Gleicher and N. Ferrier. Evaluating video-based motion capture. In *Computer Animation, 2002. Proceedings of*, pages 75–80. IEEE, 2002.

[100] R. Gobbens, K. Luijkx, M. T. Wijnen-Sponselee, and J. Schols. Towards an integral conceptual model of frailty. *The journal of nutrition, health & aging*, 14(3):175–181, 2010.

[101] R. J. Gobbens, K. G. Luijkx, M. T. Wijnen-Sponselee, and J. M. Schols. In search of an integral conceptual definition of frailty: opinions of experts. *Journal of the American Medical Directors Association*, 11(5):338–343, 2010.

[102] R. J. Gobbens, M. A. van Assen, K. G. Luijkx, and J. M. Schols. The predictive validity of the tilburg frailty indicator: disability, health care utilization, and quality of life in a population at risk. *The Gerontologist*, 52(5):619–631, 2012.

[103] R. J. Gobbens, M. A. van Assen, K. G. Luijkx, M. T. Wijnen-Sponselee, and J. M. Schols. The tilburg frailty indicator: psychometric properties. *Journal of the American Medical Directors Association*, 11(5):344–355, 2010.

[104] A. N. M. Gomatam and S. Sasi. Multimodal gait recognition based on stereo vision and 3d template matching. In H. R. Arabnia, editor, *CISST*, pages 405–410. CSREA Press, 2004.

[105] I. González, I. H. López-Nava, J. Fontecha, A. Muñoz-Meléndez, A. I. Pérez-SanPablo, and I. Quiñones-Urióstegui. Comparison between passive vision-based system and a wearable inertial-based system for estimating temporal gait parameters related to the gaitrite electronic walkway. *Journal of Biomedical Informatics*, 62:210–223, 2016.

[106] H. Gonzalez-Jorge, B. Riveiro, E. Vazquez-Fernandez, J. Martínez-Sánchez, and P. Arias. Metrological evaluation of microsoft kinect and asus xtion sensors. *Measurement*, 46(6):1800–1806, 2013.

[107] H. Gonzalez-Jorge, P. Rodríguez-Gonzálvez, J. Martínez-Sánchez, D. González-Aguilera, P. Arias, M. Gesto, and L. Díaz-Vilariño. Metrological comparison between kinect i and kinect ii sensors. *Measurement*, 70:21–26, 2015.

[108] J. M. Guralnik, L. Ferrucci, C. F. Pieper, S. G. Leveille, K. S. Markides, G. V. Ostir, S. Studenski, L. F. Berkman, and R. B. Wallace. Lower extremity function and subsequent disability: consistency across studies, predictive models, and value of gait speed alone compared with the short physical performance battery. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 55(4):M221–M231, 2000.

[109] J. M. Guralnik, E. M. Simonsick, L. Ferrucci, R. J. Glynn, L. F. Berkman, D. G. Blazer, P. A. Scherr, and R. B. Wallace. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *Journal of gerontology*, 49(2):M85–M94, 1994.

[110] J. Hamm, A. G. Money, A. Atwal, and I. Paraskevopoulos. Fall prevention intervention technologies: A conceptual framework and survey of the state of the art. *Journal of biomedical informatics*, 59:319–345, 2016.

[111] J. Han, L. Shao, D. Xu, and J. Shotton. Enhanced computer vision with microsoft kinect sensor: A review. *IEEE transactions on cybernetics*, 43(5):1318–1334, 2013.

[112] I. Haritaoglu, D. Harwood, and L. S. Davis. *W4S: A real-time system for detecting and tracking people in 2 1/2D*, pages 877–892. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998.

[113] A. Hassani, A. Kubicki, V. Brost, and F. Yang. Real-time 3d tug test movement analysis for balance assessment using microsoft kinect. *Aiia2014. Di. Unipi. It*, pages 1–10, 2014.

[114] J. M. Hausdorff, D. A. Rios, and H. K. Edelberg. Gait variability and fall risk in community-living older adults: a 1-year prospective study. *Archives of physical medicine and rehabilitation*, 82(8):1050–1056, 2001.

[115] W. He, D. Goodkind, and P. R. Kowal. *An aging world: 2015*. United States Census Bureau, 2016.

[116] Y. Higashi, K. Yamakoshi, T. Fujimoto, M. Sekine, and T. Tamura. Quantitative evaluation of movement using the timed up-and-go test. *IEEE Engineering in Medicine and Biology Magazine*, 27(4):38–46, 2008.

[117] M. Hofmann, S. Bachmann, and G. Rigoll. 2.5 d gait biometrics using the depth gradient histogram energy image. In *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 399–403. IEEE, 2012.

[118] M. Hofmann, J. Geiger, S. Bachmann, B. Schuller, and G. Rigoll. The tum gait from audio, image and depth (gaid) database: Multimodal recognition of subjects and traits. *Journal of Visual Communication and Image Representation*, 25(1):195–206, 2014.

[119] D. Hogg. Model-based vision: a program to see a walking person. *Image and Vision computing*, 1(1):5–20, 1983.

[120] J. H. Hollman, E. M. McDade, and R. C. Petersen. Normative spatiotemporal gait parameters in older adults. *Gait & posture*, 34(1):111–118, 2011.

[121] B. Hotrabhavananda, A. K. Mishra, M. Skubic, N. Hotrabhavananda, and C. Abbott. Evaluation of the microsoft kinect skeletal versus depth data analysis for timed-up and go and figure of 8 walk tests. In *IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC)*, pages 2274–2277. IEEE, 2016.

[122] F. Hug and S. Dorel. Electromyographic analysis of pedaling: a review. *Journal of Electromyography and Kinesiology*, 19(2):182–198, 2009.

[123] S. Hussmann, T. Ringbeck, and B. Hagebeuker. A performance review of 3d tof vision systems in comparison to stereo vision systems. In *Stereo vision*. InTech, 2008.

[124] G. Iddan and G. Yahav. 3d imaging in the studio (and elsewhere...). In *Proceedings of the SPIE 4298: Videometrics and Optical Methods for 3D Shape Measurements.*, pages 48–55, 2001.

[125] D. Ioannidis, D. Tzovaras, I. G. Damousis, S. Argyropoulos, and K. Moustakas. Gait recognition using compact feature extraction transforms and depth information. *IEEE Transactions on Information Forensics and security*, 2(3):623–630, 2007.

[126] M. Isken, T. Frenken, M. Frenken, and A. Hein. Towards pervasive mobility assessments in clinical and domestic environments. In *Smart Health*, pages 71–98. Springer, 2015.

[127] R. Jafri and H. R. Arabnia. A survey of face recognition techniques. *Jips*, 5(2):41–68, 2009.

[128] X. Ji and H. Liu. Advances in view-invariant human motion analysis: A review. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(1):13–24, 2010.

[129] S. Jiang, Y. Wang, Y. Zhang, and J. Sun. Real time gait recognition system based on kinect skeleton feature. In *Asian Conference on Computer Vision*, pages 46–57. Springer, 2014.

[130] A. H. Johnston and G. M. Weiss. Smartwatch-based biometric gait recognition. In *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–6. IEEE, 2015.

[131] C. J. Jones, R. E. Rikli, and W. C. Beam. A 30-s chair-stand test as a measure of lower body strength in community-residing older adults. *Research quarterly for exercise and sport*, 70(2):113–119, 1999.

[132] S. X. Ju, M. J. Black, and Y. Yacoob. Cardboard people: A parameterized model of articulated image motion. In *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, 1996.*, pages 38–44. IEEE, 1996.

[133] S. K. Jun, X. Zhou, D. K. Ramsey, and V. N. Krovi. A comparative study of human motion capture and computational analysis tools. In *The 2nd International Digital Human Modeling Symposium, Citeseer*, 2013.

[134] H. Jürges. True health vs response styles: exploring cross-country differences in self-reported health. *Health economics*, 16(2):163–178, 2007.

[135] S. Kaenchan, P. Mongkolnam, B. Watanapa, and S. Sathienpong. Automatic multiple kinect cameras setting for simple walking posture analysis. In *Computer Science and Engineering Conference (ICSEC), 2013 International*, pages 245–249. IEEE, 2013.

[136] A. Kale, A. Sundaresan, A. Rajagopalan, N. P. Cuntoor, A. K. Roy-Chowdhury, V. Kruger, and R. Chellappa. Identification of humans using gait. *IEEE Transactions on image processing*, 13(9):1163–1173, 2004.

[137] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

[138] B. A. H. Kargar, A. Mollahosseini, T. Struemph, W. Pace, R. D. Nielsen, and M. H. Mahoor. Automatic measurement of physical mobility in get-up-and-go test using kinect sensor. In *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3492–3495. IEEE, 2014.

[139] D. Kastaniotis, G. Economou, S. Fotopoulos, G. Kartsakalis, and P. Papathanasopoulos. Using kinect for assesing the state of multiple sclerosis patients. In *2014 EAI 4th International Conference on Wireless Mobile Communication and Healthcare (Mobihealth)*, pages 164–167. IEEE, 2014.

[140] J. Kaye, N. Mattek, H. Dodge, T. Buracchio, D. Austin, S. Hagler, M. Pavel, and T. Hayes. One walk a year to 1000 within a year: Continuous in-home unobtrusive gait assessment of older adults. *Gait & posture*, 35(2):197–202, 2012.

[141] K. Khoshelham and S. O. Elberink. Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors*, 12(2):1437–1454, 2012.

[142] K. G. Kinsella and D. R. Phillips. *Global aging: The challenge of success*, volume 60. Population Reference Bureau Washington, DC, 2005.

[143] A. G. Kirk, J. F. O'Brien, and D. A. Forsyth. Skeletal parameter estimation from optical motion capture data. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 782–788. IEEE, 2005.

[144] A. Kirmani, A. Colaço, F. N. Wong, and V. K. Goyal. Exploiting sparsity in time-of-flight range acquisition using a single time-resolved sensor. *Optics Express*, 19(22):21485–21507, 2011.

[145] N. Kitsunezaki, E. Adachi, T. Masuda, and J.-i. Mizusawa. Kinect applications for the physical rehabilitation. In *2013 IEEE International Symposium on Medical Measurements and Applications Proceedings (MeMeA)*, pages 294–299. IEEE, 2013.

[146] P. Kohli and J. Shotton. Key developments in human pose estimation for kinect. In *Consumer Depth Cameras for Computer Vision*, pages 63–70. Springer, 2013.

[147] A. Kolb, E. Barth, R. Koch, and R. Larsen. Time-of-flight sensors in computer graphics. In *Eurographics (STARs)*, pages 119–134, 2009.

[148] R. W. Kressig, O. Beauchet, et al. Guidelines for clinical applications of spatio-temporal gait analysis in older adults. *Aging clinical and experimental research*, 18(2):174–176, 2006.

[149] R. W. Kressig, R. J. Gregor, A. Oliver, D. Waddell, W. Smith, M. O'Grady, A. T. Curns, M. Kutner, and S. L. Wolf. Temporal and spatial features of gait in older adults transitioning to frailty. *Gait & posture*, 20(1):30–35, 2004.

[150] A. Kurakin, Z. Zhang, and Z. Liu. A real time system for dynamic hand gesture recognition with a depth sensor. In *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, pages 1975–1979. IEEE, 2012.

[151] W. Kusakunniran, Q. Wu, H. Li, and J. Zhang. Multiple views gait recognition using view transformation model based on optimized gait energy image. In *12th International Conference on Computer Vision Workshops (ICCV Workshops), 2009 IEEE*, pages 1058–1064. IEEE, 2009.

[152] C. Lafont, S. Gérard, T. Voisin, M. Pahor, B. Vellas, et al. Reducing "iatrogenic disability" in the hospitalized frail elderly. *The journal of nutrition, health & aging*, 15(8):645–660, 2011.

[153] F. Landi, A. M. Abbatecola, M. Provinciali, A. Corsonello, S. Bustacchini, L. Manigrasso, A. Cherubini, R. Bernabei, and F. Lattanzio. Moving against frailty: does physical activity matter? *Biogerontology*, 11(5):537–545, 2010.

[154] T. K. Lee, M. Belkhatir, and S. Sanei. A comprehensive review of past and present vision-based techniques for gait recognition. *Multimedia tools and applications*, 72(3):2833–2869, 2014.

[155] D. Leightley, M. H. Yap, J. Coulson, Y. Barnouin, and J. S. McPhee. Benchmarking human motion analysis using kinect one: an open source dataset. In *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pages 1–7. IEEE, 2015.

[156] D. Leightley, M. H. Yap, and J. McPhee. Automated analysis and quantification of human mobility using a depth sensor. *IEEE journal of biomedical and health informatics*, 2016.

[157] M. K. Leung and Y.-H. Yang. First sight: A human body outline labeling system. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(4):359–377, 1995.

[158] X. Li and Y. Chen. Gait recognition based on structural gait energy image. *Journal of Computational Information Systems*, 9(1):121–126, 2013.

[159] J. Little and J. Boyd. Recognizing people by their gait: the shape of motion. *Videre: Journal of computer vision research*, 1(2):1–32, 1998.

[160] F. Liu, Y. Wang, Q. Wang, L. Zhang, and W. Zeng. A new gait recognition method using kinect via deterministic learning. In *Intelligent Control and Automation (WCICA), 2016 12th World Congress on*, pages 830–835. IEEE, 2016.

[161] H. Liu, Y. Cao, and Z. Wang. Automatic gait recognition from a distance. In *Control and Decision Conference (CCDC), 2010 Chinese*, pages 2777–2782. IEEE, 2010.

[162] R. Lloréns, E. Noé, V. Naranjo, A. Borrego, J. Latorre, and M. Alcañiz. Tracking systems for virtual rehabilitation: Objective performance vs. subjective experience. a practical scenario. *Sensors*, 15(3):6586–6606, 2015.

[163] B. Lo and S. Velastin. Automatic congestion detection system for underground platforms. In *Intelligent Multimedia, Video and Speech Processing, 2001. Proceedings of 2001 International Symposium on*, pages 158–161. IEEE, 2001.

[164] O. Lohmann, T. Luhmann, and A. Hein. Skeleton timed up and go. In *2012 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1–5. IEEE, 2012.

[165] D. Lopez, L. Flicker, and A. Dobson. Validation of the frail scale in a cohort of older australian women. *Journal of the American Geriatrics Society*, 60(1):171–173, 2012.

[166] S. Lord, T. Howe, J. Greenland, L. Simpson, and L. Rochester. Gait variability in older adults: a structured review of testing protocol and clinimetric properties. *Gait & posture*, 34(4):443–450, 2011.

[167] J. K. Loudon, M. Swift, and S. Bell. *The clinical orthopedic assessment guide*. Human Kinetics, 2008.

[168] B. E. Maki. Gait changes in older adults: predictors of falls or indicators of fear? *Journal of the American geriatrics society*, 45(3):313–320, 1997.

[169] J. Man and B. Bhanu. Individual recognition using gait energy image. *IEEE transactions on pattern analysis and machine intelligence*, 28(2):316–322, 2006.

[170] P. Manckoundia, F. Mourey, P. Pfitzenmeyer, and C. Papaxanthis. Comparison of motor strategies in sit-to-stand and back-to-sit motions between healthy and alzheimer's disease elderly subjects. *Neuroscience*, 137(2):385–392, 2006.

[171] S. Mathias, U. Nayak, and B. Isaacs. Balance in elderly patients: the" get-up and go" test. *Archives of physical medicine and rehabilitation*, 67(6):387–389, 1986.

[172] C. Menier, E. Boyer, and B. Raffin. 3d skeleton-based body pose recovery. In *3rd International Symposium on 3D Data Processing, Visualization and Transmission (DPVT'06)*, pages 389–396. IEEE Computer Society, 2006.

[173] B. F. Mentiplay, L. G. Perraton, K. J. Bower, Y.-H. Pua, R. McGaw, S. Heywood, and R. A. Clark. Gait assessment using the microsoft xbox one kinect: Concurrent validity and inter-day reliability of spatiotemporal and kinematic variables. *Journal of biomechanics*, 48(10):2166–2170, 2015.

[174] R. Merletti and P. A. Parker. *Electromyography: physiology, engineering, and non-invasive applications*, volume 11. John Wiley & Sons, 2004.

[175] I. Mikić, M. Trivedi, E. Hunter, and P. Cosman. Human body model acquisition and tracking using voxel data. *International Journal of Computer Vision*, 53(3):199–223, 2003.

[176] S. Milani and G. Calvagno. Joint denoising and interpolation of depth maps for ms kinect sensors. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 797–800. IEEE, 2012.

[177] N. Millor, P. Lecumberri, M. Gómez, A. Martínez-Ramírez, and M. Izquierdo. An evaluation of the 30-s chair stand test in older adults: frailty detection based on kinematic parameters from a single inertial unit. *J Neuroeng Rehabil*, 10(86):0003–10, 2013.

[178] M. Milovanovic, M. Minovic, and D. Starcevic. Walking in colors: human gait recognition using kinect and cbir. *IEEE MultiMedia*, 20(4):28–36, 2013.

[179] A. B. Mitnitski, J. E. Graham, A. J. Mogilner, and K. Rockwood. Frailty, fitness and late-life mortality in relation to chronological and biological age. *BMC geriatrics*, 2(1):1, 2002.

[180] T. B. Moeslund and E. Granum. A survey of computer vision-based human motion capture. *Computer vision and image understanding*, 81(3):231–268, 2001.

[181] T. B. Moeslund, A. Hilton, and V. Krüger. A survey of advances in vision-based human motion capture and analysis. *Computer vision and image understanding*, 104(2):90–126, 2006.

[182] M. Montero-Odasso, S. W. Muir, M. Hall, T. J. Doherty, M. Kloseck, O. Beauchet, and M. Speechley. Gait variability is associated with frailty in community-dwelling older adults. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 66(5):568–576, 2011.

[183] M. Montero-Odasso, M. Schapira, E. R. Soriano, M. Varela, R. Kaplan, L. A. Camera, and L. M. Mayorga. Gait velocity as a single predictor of adverse events in healthy seniors aged 75 years and older. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 60(10):1304–1309, 2005.

[184] J. E. Morley, T. Malmstrom, and D. Miller. A simple frailty questionnaire (frail) predicts outcomes in middle aged african americans. *The journal of nutrition, health & aging*, 16(7):601–608, 2012.

[185] J. E. Morley, B. Vellas, G. A. van Kan, S. D. Anker, J. M. Bauer, R. Bernabei, M. Cesari, W. Chumlea, W. Doehner, J. Evans, et al. Frailty consensus: a call to action. *Journal of the American Medical Directors Association*, 14(6):392–397, 2013.

[186] H. Mousavi Hondori and M. Khademi. A review on technical and clinical impact of microsoft kinect on physical therapy and rehabilitation. *Journal of medical engineering*, 2014, 2014.

[187] S. D. Mowbray and M. S. Nixon. Automatic gait recognition via fourier descriptors of deformable objects. In *International Conference on Audio-and Video-Based Biometric Person Authentication*, pages 566–573. Springer, 2003.

[188] B. C. Munsell, A. Temlyakov, C. Qu, and S. Wang. Person identification using full-body motion and anthropometric biometrics from kinect videos. In *European Conference on Computer Vision*, pages 91–100. Springer, 2012.

[189] A. Muro-de-la Herran, B. Garcia-Zapirain, and A. Mendez-Zorrilla. Gait analysis methods: an overview of wearable and non-wearable systems, highlighting clinical applications. *Sensors*, 14(2):3362–3394, 2014.

[190] M. P. Murray. Gait as a total pattern of movement: Including a bibliography on gait. *American Journal of Physical Medicine & Rehabilitation*, 46(1):290–333, 1967.

[191] B. Najafi, K. Aminian, A. Paraschiv-Ionescu, F. Loew, C. J. Bula, and P. Robert. Ambulatory system for human motion analysis using a kinematic sensor: monitoring of daily physical activity in the elderly. *IEEE Transactions on biomedical Engineering*, 50(6):711–723, 2003.

[192] B. Najafi, T. Khan, and J. Wrobel. Laboratory in a box: wearable sensors and its advantages for gait analysis. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6507–6510. IEEE, 2011.

[193] D. Ndayikengurukiye and M. Mignotte. High-frequency spectral energy map estimation based gait analysis system using a depth camera for pathology detection. In *International Conference Image Analysis and Recognition*, pages 38–45. Springer, 2016.

[194] H. A. Nguyen, E. Auvinet, M. Mignotte, J. A. de Guise, and J. Meunier. Analyzing gait pathologies using a depth camera. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4835–4838. IEEE, 2012.

[195] S. Nishiguchi, M. Yamada, K. Nagai, S. Mori, Y. Kajiwara, T. Sonoda, K. Yoshimura, H. Yoshitomi, H. Ito, K. Okamoto, et al. Reliability and validity of gait analysis by android-based smartphone. *Telemedicine and e-Health*, 18(4):292–296, 2012.

[196] S. A. Niyogi and E. H. Adelson. Analyzing and recognizing walking figures in xyt. In *1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94.*, pages 469–474. IEEE, 1994.

[197] Š. Obdržálek, G. Kurillo, F. Ofli, R. Bajcsy, E. Seto, H. Jimison, and M. Pavel. Accuracy and robustness of kinect pose estimation in the context of coaching of elderly population. In *Engineering in medicine and biology society (EMBC), 2012 annual international conference of the IEEE*, pages 1188–1193. IEEE, 2012.

[198] T. Oggier, B. Büttgen, F. Lustenberger, G. Becker, B. Rüegg, and A. Hodac. Swissranger sr3000 and first experiences based on miniaturized 3d-tof cameras. *Proc. of the First Range Imaging Research Day at ETH Zurich*, 2005.

[199] T. Pallejà, M. Teixidó, M. Tresanchez, and J. Palacín. Measuring gait using a ground laser range sensor. *Sensors*, 9(11):9133–9146, 2009.

[200] G. Paolini, A. Peruzzi, A. Mirelman, A. Cereatti, S. Gaukrodger, J. M. Hausdorff, and U. Della Croce. Validation of a method for real time foot position and orientation tracking with microsoft kinect technology for use in virtual reality and treadmill based gait training programs. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 22(5):997–1002, 2014.

[201] M. Parajuli, D. Tran, W. Ma, and D. Sharma. Senior health monitoring using kinect. In *2012 Fourth International Conference on Communications and Electronics (ICCE)*, pages 309–312. IEEE, 2012.

[202] J. S. Partridge, D. Harari, and J. K. Dhesi. Frailty in the older surgical patient: a review. *Age and ageing*, 41(2):142–147, 2012.

[203] M. Pavel, T. Hayes, I. Tsay, D. Erdogmus, A. Paul, N. Larimer, H. Jimison, and J. Nutt. Continuous assessment of gait velocity in parkinson's disease from unobtrusive measurements. In *2007 3rd International IEEE/EMBS Conference on Neural Engineering*, pages 700–703. IEEE, 2007.

[204] M. Pavel, T. L. Hayes, A. Adami, H. Jimison, and J. Kaye. Unobtrusive assessment of mobility. In *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, pages 6277–6280. IEEE, 2006.

[205] V. I. Pavlovic, R. Sharma, and T. S. Huang. Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7):677–695, 1997.

[206] X. Perez-Sala, S. Escalera, C. Angulo, and J. Gonzalez. A survey on model based approaches for 2d and 3d visual human pose recovery. *Sensors*, 14(3):4189–4210, 2014.

[207] L. L. Peters, H. Boter, E. Buskens, and J. P. Slaets. Measurement properties of the groningen frailty indicator in home-dwelling and institutionalized elderly people. *Journal of the American Medical Directors Association*, 13(6):546–551, 2012.

[208] M. J. Peterson, C. Giuliani, M. C. Morey, C. F. Pieper, K. R. Evenson, V. Mercer, H. J. Cohen, M. Visser, J. S. Brach, S. B. Kritchevsky, et al. Physical activity as a preventative factor for frailty: the health, aging, and body composition study. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 64(1):61–68, 2009.

[209] A. Pfister, A. M. West, S. Bronner, and J. A. Noah. Comparative abilities of microsoft kinect and vicon 3d motion capture for gait analysis. *Journal of medical engineering & technology*, 38(5):274–280, 2014.

[210] L. J. Phillips, C. B. DeRoche, M. Rantz, G. L. Alexander, M. Skubic, L. Despins, C. Abbott, B. H. Harris, C. Galambos, and R. J. Koopman. Using embedded sensors in independent living to predict gait changes and falls. *Western Journal of Nursing Research*, page 0193945916662027, 2016.

[211] M. Piccardi. Background subtraction techniques: a review. In *2004 IEEE international conference on Systems, man and cybernetics*, volume 4, pages 3099–3104. IEEE, 2004.

[212] C. Plagemann, V. Ganapathi, D. Koller, and S. Thrun. Real-time identification and localization of body parts from depth images. In *2010 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3108–3113. IEEE, 2010.

[213] R. Planinc and M. Kampel. Detecting changes in elderly's mobility using inactivity profiles. In *International Workshop on Ambient Assisted Living*, pages 100–103. Springer, 2013.

[214] R. Plänkers and P. Fua. Tracking and modeling people in video sequences. *Computer Vision and Image Understanding*, 81(3):285–302, 2001.

[215] D. Podsiadlo and S. Richardson. The timed up & go: a test of basic functional mobility for frail elderly persons. *Journal of the American geriatrics Society*, 39(2):142–148, 1991.

[216] R. Poppe. Vision-based human motion analysis: An overview. *Computer vision and image understanding*, 108(1):4–18, 2007.

[217] R. Poppe. A survey on vision-based human action recognition. *Image and vision computing*, 28(6):976–990, 2010.

[218] J. Preis, M. Kessel, M. Werner, and C. Linnhoff-Popien. Gait recognition with kinect. In *1st international workshop on kinect in pervasive computing*, pages P1–P4. New Castle, UK, 2012.

[219] A. Procházka, O. Vyšata, M. Vališ, O. Ťupa, M. Schätz, and V. Mařík. Use of the image and depth sensors of the microsoft kinect for the detection of gait disorders. *Neural Computing and Applications*, 26(7):1621–1629, 2015.

[220] N. Pugeault and R. Bowden. Spelling it out: Real-time asl fingerspelling recognition. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*., pages 1114–1119. IEEE, 2011.

[221] F. Qi, J. Han, P. Wang, G. Shi, and F. Li. Structure guided fusion for depth map inpainting. *Pattern Recognition Letters*, 34(1):70–76, 2013.

[222] D. Regazzoni, G. de Vecchi, and C. Rizzi. Rgb cams vs rgb-d sensors: low cost motion capture technologies performances and limitations. *Journal of Manufacturing Systems*, 33(4):719–728, 2014.

[223] T. N. Robinson, J. I. Wallace, D. S. Wu, A. Wiktor, L. F. Pointer, S. M. Pfister, T. J. Sharp, M. J. Buckley, and M. Moss. Accumulated frailty characteristics predict postoperative discharge institutionalization in the geriatric patient. *Journal of the American College of Surgeons*, 213(1):37–42, 2011.

[224] A. P. Rocha, H. Choupina, J. M. Fernandes, M. J. Rosas, R. Vaz, and J. P. S. Cunha. Parkinson's disease assessment based on gait analysis using an innovative rgb-d camera system. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3126–3129. IEEE, 2014.

[225] K. Rockwood. What would make a definition of frailty successful? *Age and Ageing*, 34(5):432–434, 2005.

[226] K. Rockwood, M. Andrew, and A. Mitnitski. A comparison of two approaches to measuring frailty in elderly people. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 62(7):738–743, 2007.

[227] K. Rockwood, X. Song, C. MacKnight, H. Bergman, D. B. Hogan, I. McDowell, and A. Mitnitski. A global clinical measure of fitness and frailty in elderly people. *Canadian Medical Association Journal*, 173(5):489–495, 2005.

[228] D. B. Rolfson, S. R. Majumdar, R. T. Tsuyuki, A. Tahir, and K. Rockwood. Validity and reliability of the edmonton frail scale. *Age and ageing*, 35(5):526–529, 2006.

[229] R. Rosales and S. Sclaroff. Inferring body pose without tracking body parts. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 721–727. IEEE, 2000.

[230] M. D. Rothman, L. Leo-Summers, and T. M. Gill. Prognostic significance of potential frailty criteria. *Journal of the American Geriatrics Society*, 56(12):2211–2216, 2008.

[231] C. Rougier, E. Auvinet, J. Meunier, M. Mignotte, and J. A. De Guise. Depth energy image for gait symmetry quantification. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5136–5139. IEEE, 2011.

[232] L. Z. Rubenstein, K. R. Josephson, P. R. Trueblood, S. Loy, J. O. Harker, F. M. Pietruszka, and A. S. Robbins. Effects of a group exercise program on strength, mobility, and falls among fall-prone elderly men. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 55(6):M317–M321, 2000.

[233] N. Sarafianos, B. Boteanu, B. Ionescu, and I. A. Kakadiaris. 3d human pose estimation: A review of the literature and analysis of covariates. *Computer Vision and Image Understanding*, 152:1–20, 2016.

[234] H. Sarbolandi, D. Lefloch, and A. Kolb. Kinect range sensing: Structured-light versus time-of-flight kinect. *Computer vision and image understanding*, 139:1–20, 2015.

[235] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer. The humanid gait challenge problem: Data sets, performance, and analysis. *IEEE transactions on pattern analysis and machine intelligence*, 27(2):162–177, 2005.

[236] G. M. Savva, O. A. Donoghue, F. Horgan, C. O'Regan, H. Cronin, and R. A. Kenny. Using timed up-and-go to identify frail members of the older population. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, 68(4):441–446, 2012.

[237] C. N. Scanaill, S. Carew, P. Barralon, N. Noury, D. Lyons, and G. M. Lyons. A review of approaches to mobility telemonitoring of the elderly in their living environment. *Annals of Biomedical Engineering*, 34(4):547–563, 2006.

[238] M. Schmeing and X. Jiang. Color segmentation based depth image filtering. In *Advances in Depth Image Analysis and Applications*, pages 68–77. Springer, 2013.

[239] A. Schmitz, M. Ye, R. Shapiro, R. Yang, and B. Noehren. Accuracy and repeatability of joint angles measured using a single camera markerless motion capture system. *Journal of biomechanics*, 47(2):587–591, 2014.

[240] H. Schuurmans, N. Steverink, S. Lindenberg, N. Frieswijk, and J. P. Slaets. Old or frail: what tells us more? *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 59(9):M962–M965, 2004.

[241] L. A. Schwarz, A. Mkhitaryan, D. Mateus, and N. Navab. Human skeleton tracking from depth data using geodesic distances and optical flow. *Image and Vision Computing*, 30(3):217–226, 2012.

[242] M. Schwenk, C. Howe, A. Saleh, J. Mohler, G. Grewal, D. Armstrong, and B. Najafi. Frailty and technology: a systematic review of gait analysis in those with frailty. *Gerontology*, 60(1):79–89, 2013.

[243] J. Sell and P. O'Connor. The xbox one system on a chip and kinect sensor. *IEEE Micro*, 34(2):44–53, 2014.

[244] S. Shirke, S. Pawar, and K. Shah. Literature review: Model free human gait recognition. In *2014 Fourth International Conference on Communication Systems and Network Technologies (CSNT)*, pages 891–895. IEEE, 2014.

[245] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore. Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, 56(1):116–124, 2013.

[246] A. Shumway-Cook, S. Brauer, and M. Woollacott. Predicting the probability for falls in community-dwelling older adults using the timed up & go test. *Physical therapy*, 80(9):896–903, 2000.

[247] A. Sinha, K. Chakravarty, and B. Bhowmick. Person identification using skeleton information from kinect. In *Proc. Intl. Conf. on Advances in Computer-Human Interactions*, pages 101–108, 2013.

[248] Z. Skrba, B. O'Mullane, B. R. Greene, C. N. Scanaill, C. W. Fan, A. Quigley, and P. Nixon. Objective real-time assessment of walking and turning in elderly adults. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 807–810. IEEE, 2009.

[249] J. Smisek, M. Jancosek, and T. Pajdla. 3d with kinect. In *Consumer depth cameras for computer vision*, pages 3–25. Springer, 2013.

[250] S. Springer and G. Yogev Seligmann. Validity of the kinect for gait assessment: A focused review. *Sensors*, 16(2):194, 2016.

[251] G. Sprint, D. J. Cook, and D. L. Weeks. Toward automating clinical assessments: A survey of the timed up and go. *IEEE reviews in biomedical engineering*, 8:64–77, 2015.

[252] A. Staranowicz, G. R. Brown, and G.-L. Mariottini. Evaluating the accuracy of a mobile kinect-based gait-monitoring system for fall prediction. In *Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments*, page 57. ACM, 2013.

[253] A. N. Staranowicz, C. Ray, and G.-L. Mariottini. Easy-to-use, general, and accurate multi-kinect calibration and its application to gait monitoring for fall prediction. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4994–4998. IEEE, 2015.

[254] N. Steverink, J. Slaets, H. Schuurmans, and M. Van Lis. Measuring frailty: development and testing of the groningen frailty indicator (gfi). *Gerontologist*, 41(1):236, 2001.

[255] E. Stone and M. Skubic. Evaluation of an inexpensive depth camera for in-home gait assessment. *Journal of Ambient Intelligence and Smart Environments*, 3(4):349–361, 2011.

[256] E. Stone, M. Skubic, M. Rantz, C. Abbott, and S. Miller. Average in-home gait speed: Investigation of a new metric for mobility and fall risk assessment of elders. *Gait & posture*, 41(1):57–62, 2015.

[257] E. E. Stone, D. Anderson, M. Skubic, and J. M. Keller. Extracting footfalls from voxel data. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 1119–1122. IEEE, 2010.

[258] E. E. Stone and M. Skubic. Evaluation of an inexpensive depth camera for passive in-home fall risk assessment. In *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, pages 71–77. IEEE, 2011.

[259] E. E. Stone and M. Skubic. Passive in-home measurement of stride-to-stride gait variability comparing vision and kinect sensing. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 6491–6494. IEEE, 2011.

[260] E. E. Stone and M. Skubic. Silhouette classification using pixel and voxel features for improved elder monitoring in dynamic environments. In *2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, pages 655–661. IEEE, 2011.

[261] E. E. Stone, M. Skubic, and J. Back. Automated health alerts from kinect-based in-home gait measurements. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, pages 2961–2964. IEEE, 2014.

[262] T. Stoyanov, A. Louloudi, H. Andreasson, and A. J. Lilienthal. Comparative evaluation of range sensor accuracy in indoor environments. In *5th European Conference on Mobile Robots, ECMR 2011, September 7-9, 2011, Örebro, Sweden*, pages 19–24, 2011.

[263] E. Stoykova, A. Ayd, P. Benzie, N. Grammalidis, S. Malassiotis, J. Ostermann, S. Piekh, V. Sainov, C. Theobalt, T. Thevar, et al. 3-d time-varying scene capture technologies—a survey. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(11):1568–1586, 2007.

[264] X. Suau, J. Ruiz-Hidalgo, and J. R. Casas. Real-time head and hand tracking based on 2.5d data. *Trans. Multi.*, 14(3):575–585, June 2012.

[265] E. Surer and A. Kose. Methods and technologies for gait analysis. In *Computer Analysis of Human Behavior*, pages 105–123. Springer, 2011.

[266] C. Tacconi, S. Mellone, and L. Chiari. Smartphone-based applications for investigating falls and mobility. In *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, pages 258–261. IEEE, 2011.

[267] W. Tao, T. Liu, R. Zheng, and H. Feng. Gait analysis using wearable sensors. *Sensors*, 12(2):2255–2283, 2012.

[268] M. Teixidó, T. Pallejà, M. Tresanchez, M. Nogués, and J. Palacín. Measuring oscillating walking paths with a lidar. *Sensors*, 11(5):5071–5086, 2011.

[269] G. Thrane, R. M. Joakimsen, and E. Thornquist. The association between timed up and go test and history of falls: the tromsø study. *BMC geriatrics*, 7(1):1, 2007.

[270] P. H. Torr and A. Zisserman. Mlesac: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78(1):138–156, 2000.

[271] C. S. Tucker, I. Behoora, H. B. Nembhard, M. Lewis, N. W. Sterling, and X. Huang. Machine learning classification of medication adherence in patients with movement disorders using non-wearable sensors. *Computers in Biology and Medicine*, 66:120–134, 2015.

[272] J. Tung, W. Gage, K. Zabjek, D. Brooks, B. Maki, A. Mihailidis, G. Fernie, and W. McIlroy. iwalker: a 'real-world'mobility assessment tool. *Proceedings of the 30th Canadian Medical & Biological Engineering Society*, 2007.

[273] C. van Campen. *Frail older persons in the Netherlands*. Netherlands Institute for Social Research, 2011.

[274] T. F. Van de Mortel et al. Faking it: social desirability response bias in self-report research. *Australian Journal of Advanced Nursing, The*, 25(4):40, 2008.

[275] G. A. van Kan, Y. M. Rolland, J. E. Morley, and B. Vellas. Frailty: toward a clinical definition. *Journal of the American Medical Directors Association*, 9(2):71–72, 2008.

[276] J. M. VanSwearingen and J. S. Brach. Making geriatric assessment work: selecting useful measures. *Physical Therapy*, 81(6):1233–1252, 2001.

[277] C. L. Vaughan, B. L. Davis, and J. C. O'connor. *Dynamics of human gait*, volume 24. Human Kinetics Publishers Champaign, Illinois, 1992.

[278] J. Verghese, R. Holtzer, R. B. Lipton, and C. Wang. Mobility stress test approach to predicting frailty, disability, and mortality in high-functioning older adults. *Journal of the American Geriatrics Society*, 60(10):1901–1905, 2012.

[279] J. Vermeulen. *Self-monitoring of physical frailty.* PhD thesis, Maastricht University, 2014.

[280] J. Vermeulen, J. C. Neyens, M. D. Spreeuwenberg, E. van Rossum, D. J. Hewson, and L. P. de Witte. Measuring grip strength in older adults: comparing the grip-ball with the jamar dynamometer. *Journal of Geriatric Physical Therapy*, 38(3):148–153, 2015.

[281] S. Vernon, K. Paterson, K. Bower, J. McGinley, K. Miller, Y.-H. Pua, and R. A. Clark. Quantifying individual components of the timed up and go using the kinect in people living with stroke. *Neurorehabilitation and neural repair*, 29(1):48–53, 2015.

[282] F. Wang, E. Stone, M. Skubic, J. M. Keller, C. Abbott, and M. Rantz. Toward a passive low-cost in-home gait assessment system for older adults. *IEEE journal of Biomedical and Health Informatics*, 17(2):346–355, 2013.

[283] J. Wang, M. She, S. Nahavandi, and A. Kouzani. A review of vision-based gait recognition methods for human identification. In *2010 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pages 320–327. IEEE, 2010.

[284] J. J. Wang and S. Singh. Video analysis of human dynamics—a survey. *Real-time imaging*, 9(5):321–346, 2003.

[285] L. Wang, W. Hu, and T. Tan. Recent developments in human motion analysis. *Pattern recognition*, 36(3):585–601, 2003.

[286] L. Wang, T. Tan, W. Hu, and H. Ning. Automatic gait recognition based on statistical shape analysis. *IEEE transactions on image processing*, 12(9):1120–1131, 2003.

[287] W.-H. Wang, P.-C. Chung, G.-L. Yang, C.-W. Lin, Y.-L. Hsu, and M.-C. Pai. An inertial sensor based balance and gait analysis system. In *Circuits and Systems (ISCAS), 2015 IEEE International Symposium on*, pages 2636–2639. IEEE, 2015.

[288] I. Weber, J. Koch, J. Meskemper, K. Friedl, K. Heinrich, and U. Hartmann. Is the ms kinect suitable for motion analysis? *Biomedical Engineering/Biomedizinische Technik*, 57(SI-1 Track-F), 2012.

[289] T. Welsh, A. L. Gordon, and J. Gladman. Comprehensive geriatric assessment–a guide for the non-specialist. *International journal of clinical practice*, 68(3):290–293, 2014.

[290] L. G. Wiedemann, R. Planinc, I. Nemec, and M. Kampel. Performance evaluation of joint angles obtained by the kinect v2. In *IET International Conference on Technologies for Active and Assisted Living (TechAAL)*, pages 1–6, Nov 2015.

[291] E. R. Wilhelm-Leen, Y. N. Hall, M. K. Tamura, and G. M. Chertow. Frailty and chronic kidney disease: the third national health and nutrition evaluation survey. *The American journal of medicine*, 122(7):664–671, 2009.

[292] D. A. Winter, A. E. Patla, J. S. Frank, and S. E. Walt. Biomechanical walking pattern changes in the fit and healthy elderly. *Physical therapy*, 70(6):340–347, 1990.

[293] L. Wolfson, R. Whipple, P. Amerman, and J. N. Tobin. Gait assessment in the elderly: a gait abnormality rating scale and its relation to falls. *Journal of Gerontology*, 45(1):M12–M19, 1990.

[294] J. Woo, S. C. Ho, and A. L. Yu. Walking speed and stride length predicts 36 months dependency, mortality, and institutionalization in chinese aged 70 and older. *Journal of the American Geriatrics Society*, 47(10):1257–1260, 1999.

[295] J. Woo, R. Yu, M. Wong, F. Yeung, M. Wong, and C. Lum. Frailty screening in the community using the frail scale. *Journal of the American Medical Directors Association*, 16(5):412–419, 2015.

[296] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland. Pfinder: Real-time tracking of the human body. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7):780–785, 1997.

[297] D. Xu, Y. Huang, Z. Zeng, and X. Xu. Human gait recognition using patch distribution feature and locality-constrained group sparse representation. *IEEE Transactions on Image Processing*, 21(1):316–326, 2012.

[298] X. Xu, R. W. McGorry, L.-S. Chou, J.-h. Lin, and C.-c. Chang. Accuracy of the microsoft kinect$^{\text{TM}}$ for measuring gait parameters during treadmill walking. *Gait & posture*, 42(2):145–151, 2015.

[299] Q.-L. Xue. The frailty syndrome: definition and natural history. *Clinics in geriatric medicine*, 27(1):1–15, 2011.

[300] Z. Xue, D. Ming, W. Song, B. Wan, and S. Jin. Infrared gait recognition based on wavelet transform and support vector machine. *Pattern recognition*, 43(8):2904–2910, 2010.

[301] G. Yahav, G. J. Iddan, and D. Mandelboum. 3d imaging camera for gaming application. In *International Conference on Consumer Electronics (ICCE). Digest of Technical Papers*, pages 1–2. IEEE, 2007.

[302] C. Yam, M. S. Nixon, and J. N. Carter. Automated person recognition by walking and running via model-based approaches. *Pattern Recognition*, 37(5):1057–1072, 2004.

[303] C.-Y. Yam and M. S. Nixon. *Gait Recognition, Model-Based*, pages 633–639. Springer US, Boston, MA, 2009.

[304] K. Yang, Y. Dou, S. Lv, F. Zhang, and Q. Lv. Relative distance features for gait recognition with kinect. *Journal of Visual Communication and Image Representation*, 2016.

[305] Y. Yang, F. Pu, Y. Li, S. Li, Y. Fan, and D. Li. Reliability and validity of kinect rgb-d sensor for assessing standing balance. *IEEE Sensors Journal*, 14(5):1633–1638, 2014.

[306] M. Ye, C. Yang, V. Stankovic, L. Stankovic, and S. Cheng. Gait phase classification for in-home gait assessment. In *IEEE International Conference on Multimedia and Expo*, pages 1–6, 2017.

[307] J. W. Youdas, J. H. Hollman, M. J. Aalbers, H. N. Ahrenholz, R. A. Aten, and J. J. Cremers. Agreement between the gaitrite walkway system and a stopwatch–footfall count method for measurement of temporal and spatial gait parameters. *Archives of physical medicine and rehabilitation*, 87(12):1648–1652, 2006.

[308] T. Zhang, J. Lu, G. Uswatte, E. Taub, and E. S. Sazonov. Measuring gait symmetry in children with cerebral palsy using the smartshoe. In *Healthcare Innovation Conference (HIC), 2014 IEEE*, pages 48–51. IEEE, 2014.

[309] Z. Zhang. Microsoft kinect sensor and its effect. *IEEE multimedia*, 19(2):4–10, 2012.