Intelligent Environments 2025: Combined Workshop Proceedings
R. Aquino Santos and S. Faquiri (Eds.)

doi:10.3233/AISE250006

Depth Cameras for Frailty Assessment: A Dataset for Automatic Balance Tests Evaluation

Mario QUINDE [a1], Bruno RIOJA [a], Oscar BELMONTE-FERNANDEZ [c], Gerardo CASTILLO [b] and Miguel CASTRO [a]

[a] *Faculty of Engineering, Universidad de Piura, Peru*
[b] *Faculty of Human Medicine, Universidad de Piura, Peru*
[c] *Computer Languages and Systems Department, University Jaume I*

Abstract. Information and communication technology (ICT) plays a crucial role in developing tools to enhance efficiency and reduce uncertainty in health-related procedures. The widespread adoption of sensors has facilitated the generation of large amounts of data that can be exploited with advanced data processing techniques. This research studies depth cameras (3D cameras) to assist health professionals in assessing frailty. It proposes using depth cameras to generate a dataset that can be used to train machine learning algorithms to automatically evaluate balance tests (side-by-side stance, semi-tandem stance, and tandem stance) for frailty assessment. Non-frail individuals participated in performing the three balance tests. Virtual reality lenses were used to induce imbalanced behaviours in the participants, generating data representing both balanced and imbalanced behaviours during the tests. The article presents the methodology that underpins the generation of this dataset, including the tools employed to validate the adequate performance of the cameras and support the labelling process. The dataset will be made publicly available upon completion of the labelling.

Keywords. depth camera, frailty assessment, balance test, artificial intelligence, augmented reality

1. Introduction

Frailty assessment can be defined as *"a state of increased vulnerability to poor resolution of homeostasis following a stress, which increases the risk of adverse outcomes including falls, delirium and disability"* [1]. Frailty is a strong predictor of falls among the elderly [2] and it has also been associated with the future presence of other conditions like Alzheimer's, cognitive deterioration and dependency [3], being more prevalent among women [4]. Diagnosing frailty soon is crucial as it can be potentially revert [5].

Balance assessment is one of the evaluations used to diagnose a person's physical frailty as part of the Short Physical Battery Test [6]. It consists of three test in which the person is asked to remain standing with their arms wide open and their feet in a side-by-side position for ten seconds (test 1), their feet in a semi-tandem position for ten seconds (test 2), and their feet in a full tandem position for fifteen seconds (test 3).

If they exhibit imbalance behaviour during that time, a point-based system is used to evaluate their performance for these tests. The sooner they exhibit imbalanced behaviour, the worse their assessment is.

The current balance assessment method involves a health professional providing instructions to the person being assessed and manually measuring the time from the start of each test to the moment the person exhibits imbalanced behaviour. The time is measured with a stopwatch. This procedure depends on the health professional's subjective assessment of what imbalanced behaviour means and their reaction time to record the moment (with the stopwatch) when the person exhibits imbalanced behaviour. Both human-dependent actions might result in inconsistencies in the person's balance assessment.

This research proposes using depth cameras (3D cameras) for the digitalisation of balance test evaluation for frailty assessment. Depth cameras are non-intrusive and privacy-preserving sensors that detect and monitor the position of a person's strategic joints while they perform a specific activity. They have been used for posture detection [7,8] and other tests related to functional decline [9]. For balance test evaluation, a depth camera is used to monitor the participant's joints while they perform the three tests. The long-term objective is to develop algorithms capable of detecting imbalanced behaviour automatically and record the time it took the person to exhibit it.

This paper reports on generating the dataset that will be used to develop (or train, in case machine learning techniques are used) the algorithms that will automatically detect imbalanced behaviour. Non-frail individuals performed the balance tests as participants without exhibiting imbalanced behaviour. They were then asked to perform the balance tests using virtual reality lenses immersing them in a virtual scenario that made them exhibit imbalanced behaviour. These experiments were done while a depth camera monitored and recorded their joints' positions.

The paper's main contribution is the explanation of the process followed to generate the dataset, which will then be used to develop algorithms to detect imbalanced behaviour automatically. The paper provides enough details for others to replicate the experiments and generate their datasets, adapting them to their context. The scientific community can also use the dataset to develop and train algorithms for automatically detecting imbalanced behaviour. The research also describes the validation process of the gathered data using Tableau Desktop to visualise the joints' positions during the experiments (frame by frame). This validation process is crucial because it allows data to be discarded from experiments where the depth camera did not capture data consistently. The graphic tool facilitated the validation of the camera readings.

The rest of the paper is divided as follows. Section 2 describes the methodology followed for the dataset generation and validation. Section 3 presents the results of the experiments that were done to generate and validate the dataset. Section 4 critically analyses the research's findings and Section 5 presents the conclusions of the research work.

2. Methodology

The research reported in this paper aims to generate a dataset representing balanced and imbalanced behaviour exhibited by people performing balance tests to assess physical frailty. This section describes the methodology that guided the experiments for generating the dataset, focusing on explaining the equipment used, the individuals included as participants, the experiment setup, and the tools used to validate the dataset.

2.1. Equipment

The equipment used for the data generation is the following:

Azure Kinect Development Kit A depth camera (3D camera) that differs from 2D cameras primarily in its ability to measure the distance from the tracked object to the camera lens (depth). This measurement is achieved using an infrared sensor with Time-of-Flight technology, which measures the time it takes for the infrared light emitted by the camera to reach the tracked object and return. This time interval is used to calculate the distance to the tracked object. The camera operates at a rate of 15 frames per second (fps) and features body tracking, which allows it to monitor 32 joints of an individual.

Laptop Gigabyte Aero The Azure Kinect DK is connected to this laptop, which uses the Azure Kinect DK libraries to connect to the camera. The software developed by the research team to read the data during the experiments was also installed in this equipment. The laptop specifications are: an Intel Core i7 CPU, 32 GB of RAM, 1 TB of SSD storage, and a graphic card GeForce RTX 3070.

Meta Quest 2 A virtual reality headset that works independently (without being connected to another device). This equipment is used to immerse participants in a virtual environment that makes them exhibit imbalanced behaviour during the experiments.

2.2. Participants

The 20 individuals who took part in the experiments signed an Informed Consent. The experiments were previously approved by the Ethics Committee of Universidad de Piura. A classification questionnaire was applied to get biostatistical data.. The results of the participants' answers are summarised below. No personal data was gathered during the experiments.

- Eleven (55%) participants were men and nine (45%) were women.
- Five participants (25%) had some previous experience with virtual reality tools and fifteen (75%) not.
- The average height of the participants was 170 cm. (Min. 150 - Max. 190).
- The average height of the male participants was 175 cm. (Min. 166 - Max. 190).
- The average height of the female participants was 162 cm. (Min. 150 - Max 170).

2.3. Experiments

The participants received an explanation of the experiments and specific instructions on the balance tests to perform, including a quick demonstration. The following description explains the experiments each participant went through. The Azure Kinect DK was located 2 meters from the participant, who was asked to place their feet on the marks that were previously drawn on the floor. These marks indicated the desired feet position for each test: side-by-side stance, semi-tandem stance, and tandem stance. These positions are shown in Figure 1.

Firstly, the participant was asked to maintain their balance while the depth camera tracked their joints for the three tests. The participant was asked to maintain balance for 12 seconds for the side-by-side stance, 12 seconds for the semi-tandem stance, and 17 seconds for the tandem stance. These times are two seconds longer than the actual

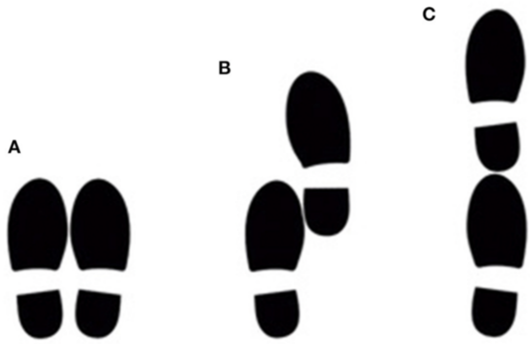

Figure 1. Balance tests positions: Side-by-side stance (A), Semi-Tandem stance (B), Tandem stance (C) [10]

requirement, so that the experiment can be recorded adequately. None of them exhibit imbalanced behaviour during these tests.

Secondly, the participant was asked to repeat the abovementioned tests using the Meta Quest 2. This device was loaded with a video (virtual environment) designed to elicit imbalanced behaviour during the tests. The video was a compilation of the most intense moments from various roller coaster videos shown from a first-person perspective. The research team edited this video to provoke imbalanced behaviour.

Twenty participants replied to the call for the experiments. The first six participants repeated the three tests (side-by-side, semi-tandem and tandem) twenty times (ten demonstrating balanced behavior and ten demonstrating imbalanced behavior). The average time for the sixty repetitions took approximately 70 minutes. The research team decided to reduce the number of repetitions to make the experiments less physically demanding and stressful for the participants. Hence, the other fourteen participants repeated the three tests ten times (five demonstrating balance behaviour and five demonstrating imbalanced behaviour). Doing this made the experiment noticeably less demanding for the participants and took approximately 25 minutes.

The depth camera tracked the participant's joints during the previously described experiments. The software that recorded these data classified them based on the test being recorded (side-by-side, semi-tandem, and tandem) and whether the data gathered aimed at representing balanced or imbalanced behaviour.

2.4. Data validation

Data validation is crucial for discarding inaccurate readings obtained during the experiments. As with any other sensor, depth cameras' performance may be affected by factors such as light intensity or the color of the participants' clothes. Before the experiments that included the participants, the depth camera was tested and calibrated to perform well in the environment where the experiments occurred. However, this calibration was done with the research team members as participants during an initial experiment setup validation. Because of this, it was relevant to validate the data obtained with the participants.

Using a graphic tool to analyse the participants' behaviour during the experiments was essential because analysing the raw data gathered by the depth camera (X-Y-Z coordinates for 32 joints at a rate of 15 fps) was too challenging. Tableau Desktop was

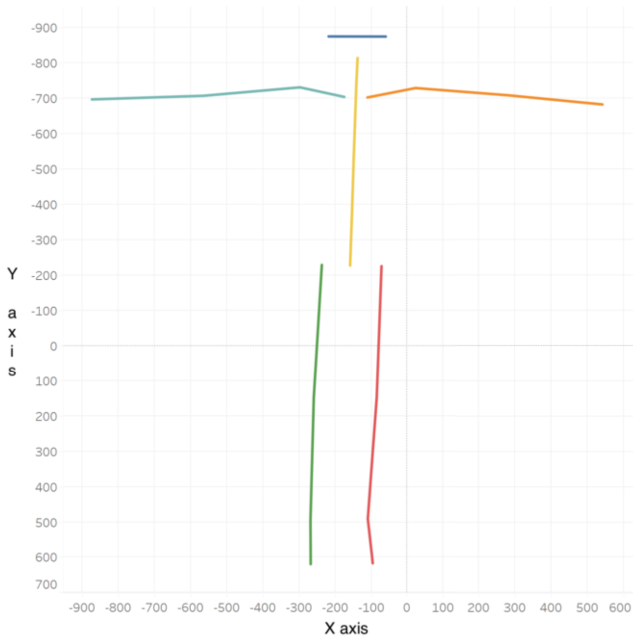

Figure 2. Example of the graphic aid use for data validation.

used to generate a representation of the joints' readings. The 32 joints were not used to generate the visual aid. The more representative ones were included to generate an easier-to-analyse image of the person, without affecting the assessment of the behaviour exhibited by the participant. Tableau Desktop's Pages feature was used to review the exhibited behaviour frame by frame with a dynamic representation of it. See Figure 2 as an example (X-Y perspective) of the visualisation used for the data validation. This visualisation was combined with another one representing the Y-Z axis perspective to have a more comprehensive representation of the participants' movements. By doing this, the behaviour of each participant during the experiments was validated.

The data validation process aimed to identify strange readings from the camera. For instance, the data obtained from one participant was discarded because the visualisation aid showed the participant doing a substantially different activity (i.e. walking backwards consistently during the whole experiment) while trying to maintain their balance. This issue was interpreted as a critical camera misreading. The validation described in this section did not aim to assess the accuracy of the camera distance readings. Data from other two individuals who participated in the experiments were discarded because they did not use the Meta Quest 2 due to discomfort with the device, which made them feel dizzy. They exhibited imbalanced behaviour by pretending it instead of using the VR device. Hence, the data from seventeen participants were included in the final dataset.

2.5. *Data labelling*

Data labelling is critical to generate a dataset for applying supervised learning techniques. The software that obtained the depth camera readings also classifies each record into the type of test it represents (side-by-side, semi-tandem and tandem) and the desired type of behaviour to represent (balanced or imbalanced repetitions). However, even if the participants use the virtual reality device during the tests, they do not exhibit imbalanced behaviour during all the time they use it. Thus, it is necessary to identify the moments in which they exhibit imbalanced behaviour and add this classification to the dataset.

Video recordings of the participants performing the experiments would be used to obtain the ground truth for each frame. However, given the restriction on not collecting personal data from the participants, Tableau Desktop visual aids were also used to identify the frames in which the participants genuinely exhibit imbalanced behaviour during the experiments. In this case, data labelling is very demanding and time-consuming, but the visualisation aids generated with Tableau Desktop make this process feasible. The identification of imbalanced behaviour for the data labelling is based on the definition of disequilibrium as a *"state of nonvertiginous altered static (eg, standing) or dynamic (eg, walking) postural balance"* [11].

3. Results

This section presents the results of the dataset generation, including a description of the experiment's outcomes, the validation of the data gathered by the experiments, and the corresponding data exclusion criteria. It also presents the advances of the data labelling process, which is still a work in progress.

Table 1 summarises the number of balance tests repetitions recorded with the depth camera during the experiments per participant, type of test (Side-by-Side, Semi-Tandem and Tandem) and the desired behaviour to exhibit during the experiments (Balance and Imbalance). Some necessary clarifications about the experiments: (a) For Participant 1, no data of the Semi-Tandem test exhibiting balanced behaviour was recorded; (b) For Participant 4, only nine repetitions of the Tandem test exhibiting balanced behaviour were recorded; (c) For Participant 8, only four repetitions of the Semi-Tandem test exhibiting imbalanced behaviour were recorded. These mistakes were human errors during the experiments.

The dataset structure is presented in Table 2, which describes the variables that will be included. A single repetition of a test was recorded in a CSV file. The 648 CSV files (total of repetitions, see Table 1) generated from the experiments were integrated and ordered using the software Tableau Prep Builder. A record of the dataset represents the X-Y-Z coordinates of a joint of a participant (*participant_id*) performing a specific test (*test*), in which they try to exhibit balanced or imbalanced behaviour (*test_objective*) for all the repetitions that were made (*test_rep*, use Table 1 as reference of the number of repetitions for each test) and for a frame (*frame*) recorded by the camera. The dataset can be changed if a long (more rows) or wide (more columns) structure is needed. The current version of the dataset has a long structure because it is more convenient for generating the abovementioned graphical aids in Tableau Desktop.

Labeling the dataset is critical to ensure the data ground truth. The time-demanding nature of this process requires people to review the readings and label the actual behaviour (balanced or imbalanced) exhibited by the participants. This is still a work in progress, whose advance is currently at 27% labelling of the dataset. These labels will allow the use of supervised learning techniques.

Table 1. Number of tests represented in the dataset

Participant	Side-by-Side		Semi-Tandem		Tandem		Total
	Balance	Imbalance	Balance	Imbalance	Balance	Imbalance	
1	10	10	–	10	10	10	50
2	10	10	10	10	10	10	60
3	10	10	10	10	10	10	60
4	10	10	10	10	10	10	60
5	10	10	10	10	9	10	59
6	5	5	5	5	5	5	30
7	5	5	5	5	5	5	30
8	5	5	5	5	5	5	29
9	5	5	5	4	5	5	29
10	5	5	5	5	5	5	30
11	5	5	5	5	5	5	30
12	5	5	5	5	5	5	30
13	5	5	5	5	5	5	30
14	5	5	5	5	5	5	30
15	5	5	5	5	5	5	30
16	5	5	5	5	5	5	30
17	5	5	5	5	5	5	30
Total	110	110	100	109	109	110	648

Table 2. Dataset: variables description

Variable name	Description	Data type
participant_id	Identifier of the participant	Integer
test	Detail of the balance test conducted	{"Side-by-Side", "Semi-Tandem", "Tandem"}
test_objective	Behaviour to exhibit during the test	{"Balance", "Imbalance"}
test_rep	Number of the repetition for each test	Integer
frame	Frame number for the test and repetition	Integer
joint	Joint whose position is recorded	String
x	X-coordinate of the joint position	Real
y	Y-coordinate of the joint position	Real
z	Z-coordinate of the joint position	Real
label	Label identified as ground truth	{"Balance", "Imbalance"}

4. Discussion

Balance tests are critical for assessing physical frailty. However, a technologically precise way of evaluating these tests has not been widely adopted yet. Using machine learning techniques to identify imbalanced behaviour during balance tests presents an opportunity to improve digitalisation in these tests. To train these algorithms, datasets representing balanced and imbalanced behaviours are needed.

Previous research has shown that depth cameras have a massive potential for being used in physical frailty assessment digitalisation [12]. Their body-tracking, non-intrusive and privacy-preserving features are also relevant for their application in medical health-care scenarios. Their use does not require installing sensors in the environment where the tests are done, and they avoid the use of wearables' sensors (e.g. Bluetooth Low Energy Beacons [13]) whose quality strongly depends on the wearable manufacturers.

This research presents the creation of a dataset representing real depth camera readings of people exhibiting balanced and imbalanced behaviours while performing the three balance tests. The research contributes to sharing the methodology with the scientific community, so that they can replicate it and generate their datasets. The scientific community will also benefit from the dataset because they can use it to create predictive models able to identify balanced and imbalanced behaviours automatically.

The use of virtual reality technology to provoke participants to exhibit imbalanced behaviour is also one of the research's highlights. This technology can immerse individuals in scenarios that can aid in generating datasets representing actual life behaviours. The main limitation evidenced during the use of this technology is the discomfort it may cause in people who tend to present vertigo, dizziness, etc. while using it. People using lenses may also present discomfort with this technology.

Labelling is the most critical process for creating a dataset in which supervised learning techniques can be used. These techniques are especially relevant in medical domains because they tend to produce more accurate models, which is a must in this application field where people's lives are at stake. The people in charge of labelling must correctly interpret the definitions of balanced and imbalanced behaviour to create accurate and consistent labels. In this research, labelling was done by the same two people who reviewed all the records together. They received an explanation of balanced and imbalanced behaviours from medical staff.

The dataset creation is currently in the labelling process. Although human bias can be present in the dataset, it is expected to be minimal and consistent because of the followed labelling strategy. An interesting point to highlight at this project stage is that, so far, not many large timespans of records generated when imbalanced behaviour was induced have been labelled as "Balance". Although some of these timespans (time windows) are of approximately fifty frames, most of them are of less than twenty frames. This means that the virtual reality tool works well and, at the end of the labelling process, it will be necessary to define a minimum threshold to classify a timespan as "Balance".

The dataset promotes future work and has great potential to become a relevant asset in research on human activity recognition. Researchers will use it to develop machine learning algorithms that automatically detect balanced and imbalanced behaviour, which is a relevant estimator for other medical issues like fall risk assessment and rehabilitation monitoring tasks. Tools automatically detecting balanced and imbalanced behaviour are also useful for physical deterioration diagnosis, as well as for automating tasks related to rehabilitation follow-up and rehabilitation treatment adjustment.

The paper presents the first stage of an ongoing research. The data obtained in this first stage is from non-frail individuals, and assessing how well the data fits with frail individuals is still pending for the next stage. The individuals who participated in the experiments are not a representative sample in terms of diversity, age, body type, etc. Future work must also include gathering more data from a representative sample and studying the dataset risk to overfit and its capacity to produce models that can be generalised.

5. Conclusions

This research studies the use of depth cameras for physical frailty assessment. It proposes and validates the creation of a dataset representing depth camera readings of people performing the three balance tests of the SPPB. The proposed methodology uses virtual reality technology to make the participants exhibit imbalanced behaviour while they perform the tests. The outcomes present the creation and validation of a less-imbalanced dataset due to the use of virtual reality to provoke exhibiting imbalanced behaviour. The research also presents the use of visualisation aids to validate the collected data and label it for future application of advanced modelling techniques to identify balanced and imbalanced behaviour automatically. The visual aids were proven efficient for data validation and a critical supporting tool for the labelling process, which is still a work in progress. The research outcomes are promising and hold great potential to benefit future work on the automatic assessment of physical frailty.

Acknowledgement

We acknowledge the funding made available by CONCYTEC and the World Bank through PROCIENCIA (PE501087047-2024-PROCIENCIA).

We thank the Korea International Cooperation Agency in Peru (KOICA Peru) for contributing to the laboratory where this research was conducted.

References

[1] Clegg A, Young J, Iliffe S, Rikkert MO, Rockwood K. Frailty in elderly people. The Lancet. 2013;381(9868):752-62.

[2] Kojima G. Frailty as a Predictor of Future Falls Among Community-Dwelling Older People: A Systematic Review and Meta-Analysis. Journal of the American Medical Directors Association. 2015;16(12):1027-33.

[3] Buchman AS, Boyle PA, Wilson RS, Tang Y, Bennett DA. Frailty is Associated With Incident Alzheimer's Disease and Cognitive Decline in the Elderly. Biopsychosocial Science and Medicine. 2007;69(5).

[4] Collard BHSRAOVRC Rose M. Prevalence of Frailty in Community-Dwelling Older Persons: A Systematic Review. Journal of the American Geriatrics Society. 2012;60(8).

[5] Ng TP, Feng L, Nyunt MSZ, Feng L, Niti M, Tan BY, et al. Nutritional, Physical, Cognitive, and Combination Interventions and Frailty Reversal Among Older Adults: A Randomized Controlled Trial. The American Journal of Medicine. 2015;128(11):1225-36.e1.

[6] Río X, Guerra-Balic M, González-Pérez A, Larrinaga-Undabarrena A, Coca A. Valores de referencia del SPPB en personas mayores de 60 años en el País Vasco. Atención Primaria. 2021;53(8):102075.

[7] Ding W, Hu B, Liu H, Wang X, Huang X. Human posture recognition based on multiple features and rule learning. International Journal of Machine Learning and Cybernetics. 2020;11(11):2529-40.

[8] Vonstad EK, Su X, Vereijken B, Bach K, Nilsen JH. Comparison of a Deep Learning-Based Pose Estimation System to Marker-Based and Kinect Systems in Exergaming for Balance Training. Sensors. 2020;20(23).

[9] Kampel M, Doppelbauer S, Planinc R. Automated Timed Up & Go Test for functional decline assessment of older adults. PervasiveHealth '18. New York, NY, USA: Association for Computing Machinery; 2018. p. 208–216.

[10] Warmerdam E, Schumacher M, Beyer T, Nerdal PT, Schebesta L, Stürner KH, et al. Postural Sway in Parkinson's Disease and Multiple Sclerosis Patients During Tasks With Different Complexity. Frontiers in Neurology. 2022;13.

[11] Lanska DJ. Disequilibrium; 2024. Available from: https://www.medlink.com/articles/disequilibrium.
[12] Boronat P, Perez-Francisco M, Gasco-Compte A, Pardo-Navarro M, Belmonte-Fernández O. GPS Suitability for Physical Frailty Assessment. Sensors. 2024;24(14).
[13] Sansano-Sansano E, Montoliu R, Belmonte-Fernandez O, Aranda FJ, Alvarez FJ. Continuous Non-Invasive Assessment of Gait Speed Through Bluetooth Low Energy. IEEE Sensors Journal. 2022;22(8).