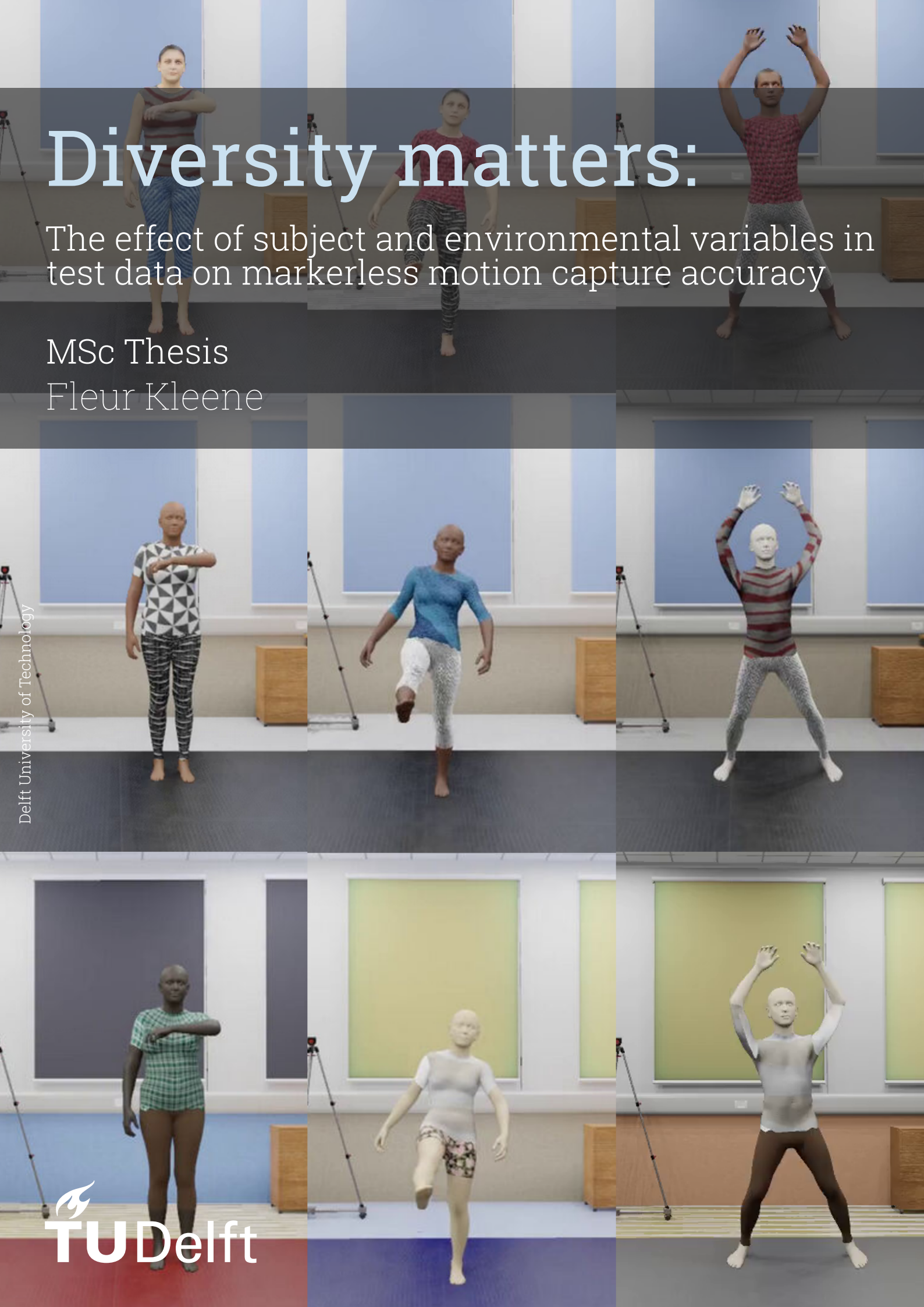

# Diversity matters:

## The effect of subject and environmental variables in test data on markerless motion capture accuracy

MSc Thesis

Fleur Kleene

Delft University of Technology

TU Delft

# Diversity matters:

## The effect of subject and environmental variables in test data on markerless motion capture accuracy

by

Fleur Kleene

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thursday February 20, 2025 at 13:00.

| | | |
|---|---|---|
| Student number: | 5637104 | |
| Project duration: | November, 2023 – February, 2025 | |
| Thesis committee: | Dr. A. Seth, | TU Delft, supervisor |
| | Dr. E. van der Kruk | TU Delft |
| | H.K.H.W. Osman | TU Delft |

*This thesis is confidential and cannot be made public until August 20, 2025.*



TU Delft

# Abstract

3D human motion capture can provide unique information on joint angles and position, and body segment lengths. It has many potential medical applications, including diagnostics, disease monitoring, and clinical decision making. Markerless motion capture makes motion capture accessible and has the potential to objectify motion analysis, and possibly reduce bias. However, neural networks for markerless motion capture can become biased by using unrepresentative training data. Mainstream datasets used for training do not provide detailed information about their subjects characteristics, like skin tone and clothing types per video, making it difficult to study their diversity. This study uses the OpenSim Driven Animated Human (ODAH) pipeline to create three datasets with increasingly diverse subjects and environments and assess the bias and generalization of a biomechanics-aware neural network. The Mean Per Joint Absolute Error (MPJAE), Procrustes Aligned Mean Per Joint Position Error (PA-MPJPE), and absolute relative scale error are determined per video and statistically analyzed together with the video's variables. Significant differences were found in the MPJAE and PA-MPJPE for videos with different motion types, subjects, skin tones, and blinds colors, and significant correlations were found with weight and BMI. The scale error differed significantly based on the subject, skin tone, environmental colors, weight, and height. The model performed significantly better on seen skin tones and environments than unseen skin tones and variables. This study has shown that it cannot be assumed that a neural network's performance will generalize to different skin tones and different environmental colors, because this will result in greater errors. In future research, the created test data should be expanded to retrain the network and determine the effect of increasing training data diversity.

# Contents

# 1

# Introduction

Human motion capture is the acquisition of kinematic movement data from various types of measurements, including videos, marker trajectories, and inertial measurements. This kinematic data provides an opportunity for extensive kinematic analysis of human movement. Applications include medical diagnostics, animation, sports analysis, and more. This report will focus on motion capture in a medical setting. Motion capture provides quantitative information, including joint angles and position, and bone length, that would otherwise not be available. This information can be used to aid in diagnostics, disease monitoring, and clinical decision-making (Salisu et al., 2023; Scott et al., 2022). Motion capture research spans many injuries and disorders, including Parkinson's, stroke, cerebral palsy, Multiple Sclerosis, and other neurological and musculoskeletal disorders (Scott et al., 2022). However, for motion capture to be a useful tool, it must provide accurate information.

Although research is being conducted on many motion capture measurement methods, the systems currently used in medical settings are mainly marker-based (Scott et al., 2022). This type of motion capture is performed in a laboratory setting with the use of multiple high-quality cameras and markers. These markers must be placed on key points of the body, which requires extensively trained personnel, and still leaves a risk of human error and variation (Colyer et al., 2018; Gorton et al., 2009; Sinclair et al., 2014). These characteristics combined make marker-based motion capture a time-consuming and expensive method (Cappozzo et al., 1996). As a result, kinematic analysis is available to only a select group of patients.

Markerless motion capture is a promising approach to making motion analysis more accessible to a larger patient population. With the use of few cameras, a model created by researchers, and minimal technical expertise, a physician would be able to perform kinematic motion analysis in the consultation room. This increases the ease of motion capture and decreases the variation between motion capture sessions (R. M. Kanko et al., 2021). In addition, this method is physically and psychologically more comfortable for the patient and allows them to move more naturally (Colyer et al., 2018; Scott et al., 2022).

Instead of basing the subjects' poses on marker positions, markerless motion capture systems use a trained neural network to extract motion data from video. The network learns based on training data, consisting of videos and matching pose data (ground truth) (R. Kanko et al., 2021). Using such a quantitative model could improve equality in medicine, as research has shown that healthcare workers have implicit biases toward different people, based on race, gender, weight, etc. (FitzGerald & Hurst, 2017), and markerless motion capture models aid with objective motion analysis. However, in order for markerless motion capture to improve medical equality, the motion capture model itself has to be accurate and bias-free.

Bias in machine learning can arise when a model is trained on a dataset that is not representative of the data that the final model will be used on (Géron, 2023). In the case of markerless motion capture, this means that the people and environments in the videos used for training the model should be representative of the full spectrum of human skin tones, body types, abilities, and movement patterns, and

of all possibly used environments. Unfortunately, the most commonly used datasets for training motion capture models (Deepcap (Habermann et al., 2020), Human3.6M (Ionescu et al., 2014), HumanEva (Sigal et al., 2010), MoVi (Ghorbani et al., 2021)) do not provide detailed information about the subjects within their datasets. Generally, dataset creators provide information about their subjects' sex, age, height, and weight, but data on their skin tones and body types are only available through visual inspection of all videos in the dataset. This makes it very difficult to judge the diversity and representativeness within a dataset. It is also important to analyze the test results of a model more extensively when the variation within the test- and train-sets correspond. This is because there might still be variables that negatively affect the model performance. It is important to be aware of these possible biases within pose estimation models in order to be able to take it into account or to find possible solutions. However, no studies could be found that study the differences in model performance as an effect of changing subject variables like skin tone and body type.

One important variable to study closely is skin tone. Research has shown that facial recognition models, widely used applications of computer vision, often have racial bias (Yucer et al., 2024). This shows that racial bias is an existing problem within computer vision that should be tested in markerless motion capture.

Moreover, videos within a motion capture dataset are often recorded in the same lab. This means that models trained on a dataset like this have only seen a single background in videos. However, users of the resulting model will have different locations for recording motion capture videos. This results in a discrepancy between the training data and the data used when the model is deployed, which could lead to overfitting. Therefore, it is important to look at model performance on videos with different backgrounds. In the same way, clothing changes between different real users. Therefore, it is also valuable to study the effect of different clothing types and colors on model performance.

More insight can be gained into these possible biases by performing tests with diverse datasets, but this requires diverse data with clear variable annotation. Research has shown that training and testing a model on synthetic video data provide results that are representative of the results on real videos (Black et al., 2023; Lin et al., 2024). Aside from speed and cost, an advantage of using synthetic data is the ease with which it is possible to change visual parameters. This creates the opportunity to test multiple variables without needing additional subjects. These characteristics make synthetic data a useful tool for testing the effects of different variables on markerless motion capture model performance.

This study aims to show which subject and environmental variables significantly affect markerless motion capture accuracy and to quantify their influence. This information will be used to give advice for the creation of future datasets and video-based markerless motion capture models. Combined, this will be an important step towards equality in motion capture.

# 2

# Methods

## 2.1. Overview

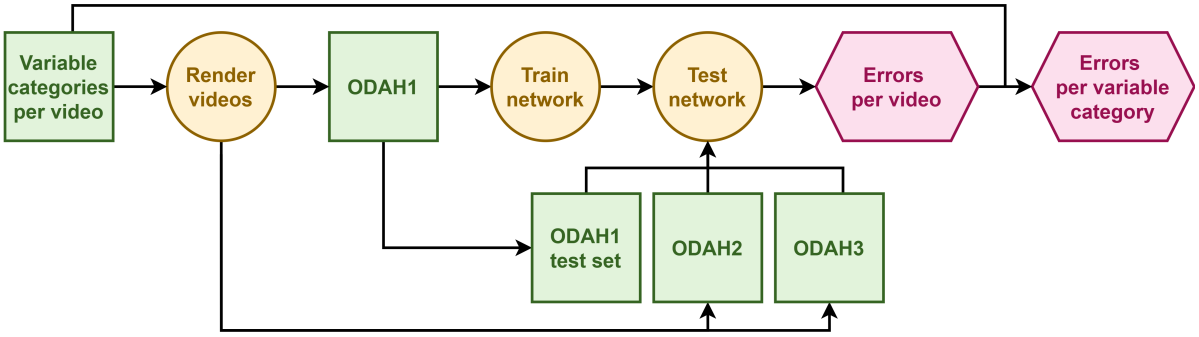


**Figure 2.1:** OpenSim Driven Animated Human (ODAH) datasets are created with increasingly diverse subjects and environments. Variables are put into the video rendering pipeline, resulting in ODAH1, ODAH2, and ODAH3. ODAH2 and ODAH3 use the same subjects and motions as the test set of ODAH1. The train set of ODAH1 is used to train a biomechanics aware neural network, which is then tested on the test set of ODAH1, and ODAH2 and ODAH3. The errors of the network's prediction are determined, and are then analyzed with the corresponding variables for each video, to get the mean errors for every variable category.

For assessing the effects of diversity on markerless motion capture performance, a pre-trained neural network is used. The biomechanics aware network has been trained on synthetic video data and is tested on the test set of this original dataset and on two others. The two new datasets consist of the same subjects and motions as the original, but skin tones, clothing textures, and environmental colors have been edited. The model's accuracy on the different datasets and on the different variables is then analyzed (Fig.2.1).

## 2.2. Synthetic video generation pipeline

Synthetic videos were created using the OpenSim Driven Animated Human (ODAH) pipeline by Lyu (2023). In this pipeline, an SMPL-X model (Pavlakos et al., 2019) is rigged against a full-body OpenSim skeletal model (Delp et al., 2007; Seth et al., 2018). The OpenSim model used is a male model that consists of 22 rigid bodies and 37 degrees of freedom and was designed for gait simulation (Rajagopal et al., 2016). A set of markers is added to the model to help with human skin mesh generation (Lyu, 2023). The SMPL-X model is a 3D human mesh model, representing the full surface of the human body. The mesh is defined by its shape and its pose (Pavlakos et al., 2019). The mesh is driven by an internal armature that does not correspond to an anatomically accurate skeleton. Lyu (2023) trained a joint regressor to relate the joint key points of the OpenSim model to the SMPL-X mesh vertices. This resulted in a fixed offset between OpenSim joints and their corresponding SMPL-X joint. Next, a personalized body mesh is created. Using MoSh++, an initial guess was made. This is a method

that converts data from marker-based motion capture into SMPL-X mesh sequences (Mahmood et al., 2019). The resulting shape and pose are then optimized to create the final mesh (Lyu, 2023).

## 2.3. Neural Network

The pre-trained biomechanics-aware neural network created by Lin et al. (2024) is used for pose estimation. Two views of a subject performing a motion are used as input to the network. First, a stacked hourglass network encodes each frame into a frame feature. Next, a spatio-temporal feature refinement uses spatial and temporal information for optimization of the features across multiple frames. The output of the network is the joint parameters of an OpenSim model. The network has been trained on videos from ODAH1, with their corresponding OpenSim models as ground truth. The loss function used for training consists of terms for joint angles, body segment scales, biomechanical joint angle constraints, and keypoint positions. The network has been shown to outperform previous state-of-the-art markerless motion capture methods.

## 2.4. Experiments

### 2.4.1. Video rendering

The described ODAH pipeline was used to create three datasets (Lyu, 2023). The OpenSim models and motions and their corresponding SMPL-X meshes were derived for 56 subjects in the AMASS BMLmovi subset (Mahmood et al., 2019). Seven of these subjects were included in the test sets. All subjects perform 21 motions, including walking, jogging, running in place, side gallop, crawling, vertical jumping, jumping jacks, kicking, stretching, crossing arms, sitting down on a chair, crossing legs while sitting, pointing, clapping hands, scratching one's head, throwing and catching, waving, pretending to take a picture, pretending to talk on the phone, and a self-chosen motion (freestyle).

The SMPL-X meshes created in the ODAH pipeline were rendered in Blender. The human body mesh was colored to resemble human skin and skin tight clothing. The shirts and pants had 4 different sleeve and pants lengths, respectively, that were randomized for every rendered video. The subjects were rendered in a small office-like environment containing a floormat, room partitions, a desk with a chair, and blinds (Fig. 2.2). Two static cameras were places at a height of $1.1 \pm 0.1$ meters. They were placed so that one camera captured the frontal view, while the other one captured the sagittal view. The positions of the cameras were randomly perturbed within a small range. The cameras had a fixed focal length of 33 mm. The sensor fit of the cameras was set to horizontal with the sensor width set to 36 mm (Lyu, 2023).

Lyu (2023) rendered ODAH1. In this dataset, one skin tone was used for all female subjects and a different skin tone was used for all male subjects. The shirt and pants had six and five different textures, respectively, that were independently randomized for every rendered video. The objects in the environment and their colors were the same in every video. The test set of ODAH1 consisted of seven test subjects. Their age, weight, height, and BMI are shown in table 2.2.

| | **ODAH1** | **ODAH2** | **ODAH3** |
|---|---|---|---|
| Motion types | 21 motion types per subject | | |
| Number of subjects | 4f, 3m | | |
| Body types | fixed per subject | | |
| Shirt types | 4 randomized sleeve lengths | | |
| Pants types | 4 randomized pants lengths | | |
| Shirt materials | 6 randomized patterns | | 9 randomized patterns |
| Pants materials | 5 randomized patterns | | 8 randomized patterns |
| Background colors | fixed | | randomized |
| Skin tones | 1 per sex | 10 randomized MST colors | |

**Table 2.1:** Overview of variables used in ODAH1, ODAH2, and ODAH3

Two new test sets, called ODAH2 and ODAH3 were created. These datasets are based on the same OpenSim models and meshes as the test set of ODAH1, and thus consist of the same subjects and motions, but they were created with additional visual variations in the rendering phase of the pipeline

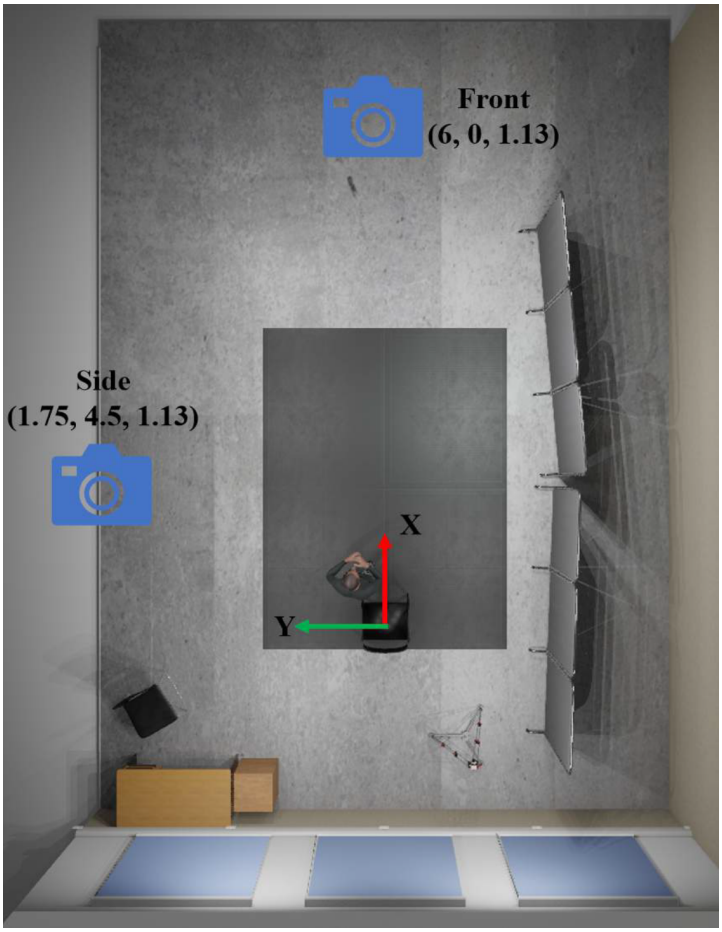


**Figure 2.2:** Rendering scene and camera placement from top view (image by Lyu (2023)). The room contains a floormat, desk, desk chair, partitions, blinds, and a tripod.

| subject number | sex | age | height (cm) | weight (kg) | BMI (m/kg$^2$) |
|---|---|---|---|---|---|
| Subject 4 | m | 19 | 178 | 68 | 21.46 |
| Subject 12 | f | 26 | 178 | 80 | 25.25 |
| Subject 30 | f | 32 | 156 | 53 | 21.5 |
| Subject 38 | f | 18 | 144 | 68 | 32.79 |
| Subject 43 | m | 26 | 162 | 68 | 25.91 |
| Subject 48 | f | 21 | 175 | 80 | 26.12 |
| Subject 72 | m | 18 | 173 | 59 | 19.71 |

**Table 2.2:** Overview of subject characteristics

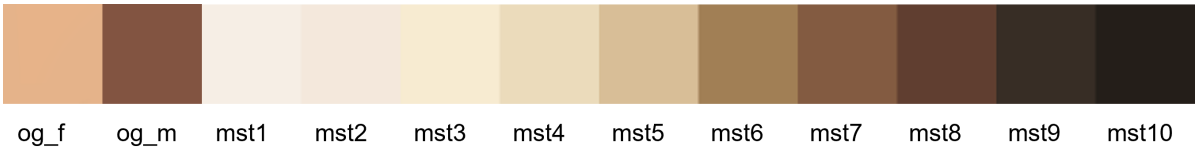


**Figure 2.3:** Skin tones included in the datasets. ODAH1 uses og f for all female subjects and og m for all male subjects. ODAH2 and ODAH3 use the 10 skin tones of the Monk Skin Tone (MST) scale (Monk, 2019). In ODAH2 and ODAH3 the skin tones are randomized.

(table 2.1). ODAH2 and ODAH3 used a randomly assigned skin tone in every video. The randomized colors were the 10 colors in the Monk Skin Tone (MST) scale (Fig. 2.3) (Monk, 2019). This scale was designed to reflect a wide range of possible skin tones and aid in building more inclusive computer

vision systems.

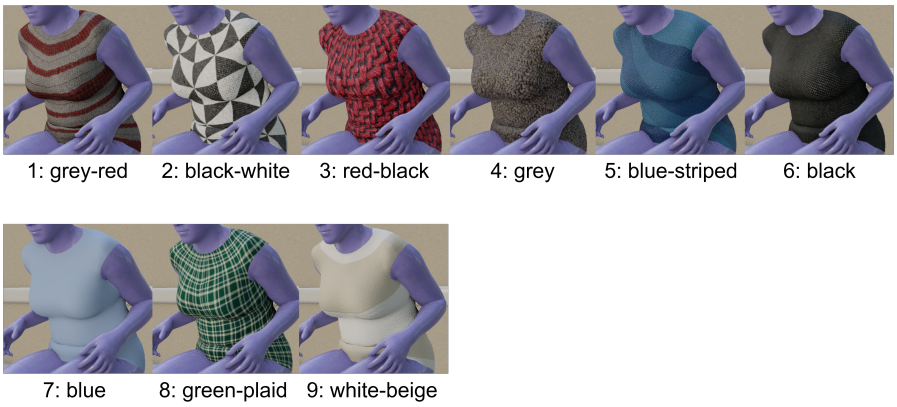


**Figure 2.4:** Shirt materials included in ODAH1-3. ODAH1 and ODAH2 use material 1-6, ODAH3 uses material 1-9. The materials are randomized in every video, independently from the other variables.

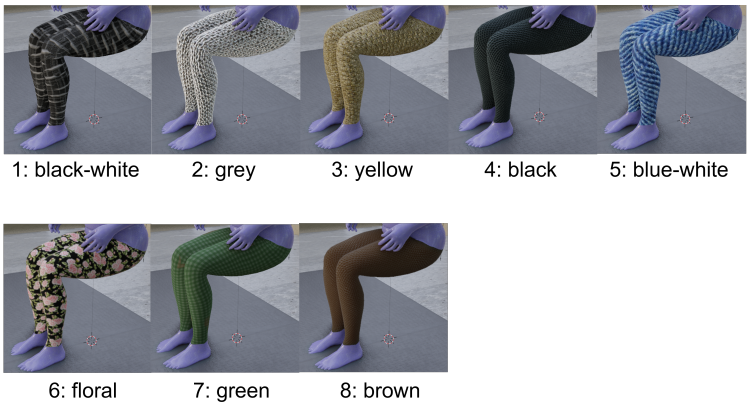


**Figure 2.5:** Pants materials included in ODAH1-3. ODAH1 and ODAH2 use material 1-5, ODAH3 uses material 1-8. The materials are randomized in every video, independently from the other variables.

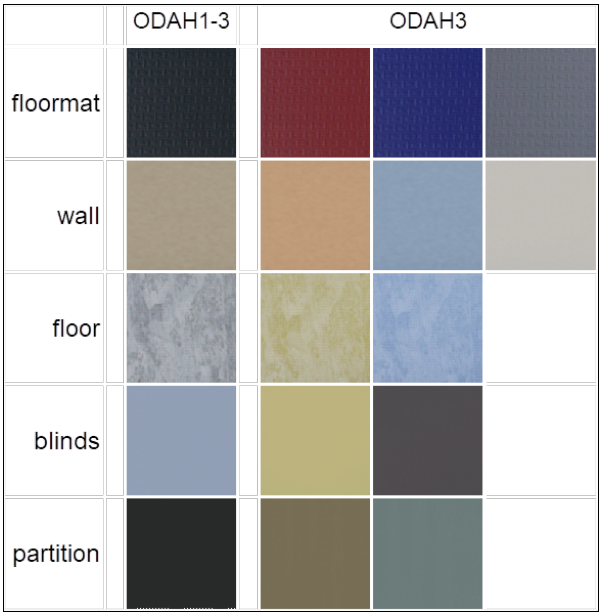


**Figure 2.6:** The environment colors per object included in ODAH1-3. ODAH1 and ODAH2 use the same colors in every video, ODAH3 independently randomizes the shown colors for every video.

ODAH3 additionally added more variation to the subject clothing and environment. Three new textures were added to both the shirt and pants textures (figures 2.4, 2.5). The floor, floormat, wall, blinds, and partitions in the environment were given 2-3 additional colors to be randomized independently (Fig. 2.6).

The videos were rendered using the BLENDER_EVEE engine in Blender 3.4. A video resolution of 1080 by 720 pixels was used with a frame rate of 100 fps. Motion blur effects are disabled. The videos are encoded in H264 formatting with a medium-quality configuration.

### 2.4.2. Pose estimation

The train set of ODAH1, consisting of 42 subjects, was used to train the neural network described in the previous section. The network was then tested on the test set of ODAH1 (hereafter referred to as ODAH1), ODAH2 and ODAH3.

After the model predicts the OpenSim motion data, these predicted values are compared with the ground truth OpenSim data, that was derived from the AMASS BMLMovi dataset in the video generation pipeline, to get the error of the prediction. Three different errors are used to assess the performance of the model.

The Mean per Joint Absolute Error (MPJAE) evaluates the joint error as

$$MPJAE = \frac{1}{T}\sum_{t=1}^{T}\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_t\|_1, \tag{2.1}$$

where $\hat{\boldsymbol{\theta}}_t$ and $\boldsymbol{\theta}_t$ represent the predicted and ground truth joint angles, respectively, for the $t^{\text{th}}$ frame, and $T$ is the total number of frames (Lin et al., 2024).

The Procrustes Aligned Mean Per Joint Position Error (PA-MPJPE) evaluates the joint location error after Procrustes alignment. Procrustes alignment is the optimal alignment of two objects by translation, rotation, and scaling (Gower, 1975). The Procrustes alignment is applied to the predicted pose with respect to the ground truth pose. The acquired joint positions are used to calculate the PA-MPJPE as

$$PA\text{-}MPJPE = \frac{1}{TB}\sum_{t=0}^{T-1}\sum_{i=0}^{B-1}\|\hat{\boldsymbol{p}}_{i,t}^{PA} - \boldsymbol{p}_{i,t}\|_2, \tag{2.2}$$

$$\hat{p}_i^{PA} = R \times \hat{p}_i \times s + t,$$

where $\hat{\boldsymbol{p}}_{i,t}$ and $\boldsymbol{p}_{i,t}$ denote the predicted and ground truth joint location, respectively, for joint $i$ and frame $t$, and $B$ denotes the total number of joint key points (Lin et al., 2024). $\hat{p}_i^{PA}$ is the joint position after translation, rotation, and scaling with rotation matrix $R \in \mathbb{R}^{3\times3}$, translation vector $t \in \mathbb{R}^{3\times1}$ and scaling factor $s$, derived from Procrustes analysis (Lawrence et al., 2019).

The absolute relative scale error (called scale error hereafter) is used to evaluate the scaling differences between the ground truth and the predicted values.

$$\text{absolute scale error} = \frac{1}{B}\sum_{i=0}^{B-1} \|\hat{\boldsymbol{s}}_i - \boldsymbol{s}_i\| \times 100\%, \tag{2.3}$$

where $\hat{\boldsymbol{s}}_i$ and $\boldsymbol{s}_i$ denote the ground truth and predicted three-dimensional vector of scale factors of body segment $i$, respectively. $B$ is the total number of body segments (Lin et al., 2024).

## 2.5. Data Analysis

To determine the effect of the variables on network performance, three statistical metrics are used, each suitable for a different variable type. All statistics are calculated separately for every error metric. The calculations were performed using the SciPy stats module in Python (Virtanen et al., 2020). Results are considered statistically significant if $p \leq 0.05$. In addition, the differences in mean error are compared between variable categories.

Categorical variables with three or more categories were compared using a one-way analysis of variance (ANOVA) test (Ross & Willson, 2017). This test compares the mean values of a dependent variable between multiple groups. It is used to compare the mean errors of all possible categories per variable, regardless of the dataset. The result is the F-statistic, representing the ratio of between-category variation to within-category variation. The p-value is also calculated to determine the statistical significance. If variables are found to have a significant effect, the error sizes are studied to quantify the effect. In addition to the analysis of all datasets combined, ANOVA tests for the skin tone variable are performed for ODAH1 and ODAH2 combined, to analyze the error difference when the environmental variables are static. In the same way, the ANOVA tests are performed for the environmental variables for ODAH2 and ODAH3 combined, to analyze the error differences when the skin tone range is static.

Binary categorical variables are compared using Welch's t-test (Welch, 1947). This test is used to compare the mean values between two groups. Welch's t-test is chosen because it is designed for unequal group variances. It is calculated as:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2}}, \tag{2.4}$$

$$s_{\bar{x}_i} = \frac{s_i}{\sqrt{N_i}}$$

where $\bar{x}_i$ and $s_{\bar{x}_i}$ are the $i^{\text{th}}$ category mean error and its standard error, with corrected sample standard deviation $s_i$ and sample size $N_i$. The only binary categorical variable in this study is the subjects' sex. The t-test is calculated once using all three datasets and once using only ODAH2 and ODAH3. ODAH1 is omitted from one of the tests due to the fact that the sex variable is directly linked to the skin tone variable in ODAH1, and this could affect the analysis of the sex variable alone. The t-test is also used to compare the network's performance of the three datasets one-to-one.

Numerical variables were analyzed using the Pearson correlation coefficient. This statistic measures the linear relationship between two variables, and is defined as:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}, \tag{2.5}$$

where $r$ is the correlation coefficient, $x_i$ and $y_i$ are the $i^{\text{th}}$ value of the variable and error, respectively, and $\bar{x}$ and $\bar{y}$ are the mean values of the variable and error, respectively (Pearson & Galton, 1895). The r-value ranges from -1 to 1, where -1 means there is a perfect negative linear correlation, 0 means there is no correlation, and +1 means there is a perfect positive correlation.

# 3

# Results


**Figure 3.1:** Examples of the same subject, motion and frame in ODAH1 (left), ODAH2 (middle), and ODAH3 (right). a: subject #30, checking watch; b: subject #38, kicking; c: subject #43, crawling; d: subject #71, jumping jacks.

Figure 3.1 shows example images of the same subject, motion, and frame within the three different datasets.

First, the effect of the categorical variables was calculated. Table 3.1 shows the F-statistic and p-value of the ANOVA for all error metrics and Table 3.2 shows the t-statistic and p-value of the t-tests for sex and datasets. The MPJPE and PA-MPJPE were significantly affected by the dataset, movement type, subject, skin tone, and blinds color. The absolute scale error was significantly affected by the dataset, subject, skin tone, shirt type, and floor, floormat, wall, blinds, and partition colors. Sex, shirt material, pants material, shirt type, and pants type did not significantly affect any of the errors in the combined datasets. Sex also did not have a significant effect in ODAH2 and ODAH3 combined. An overview of the mean errors and standard deviation of every variable category can be found in Appendix A.

All error metrics were significantly different on ODAH1 compared to ODAH2 and/or ODAH3. The t-tests show that ODAH2 gives a significantly lower scale error than ODAH3, but the MPJAE and PA-MPJPE do not show a significant difference between these two datasets (Table 3.2). The mean errors are lowest for ODAH1 followed by ODAH2 and then ODAH3 (Fig. 3.2). ODAH1 and ODAH3 have a difference in MPJAE, PA-MPJPE and scale error of 1.05°, 7.78 mm, and 1.01%, respectively.

| | MPJAE | | PA-MPJPE | | absolute scale error | |
|---|---|---|---|---|---|---|
| | F-stat | p-value | F-stat | p-value | F-stat | p-value |
| **ODAH1 + ODAH2 + ODAH3** | | | | | | |
| Dataset | **6.097** | **0.002** | **6.537** | **0.002** | **20.551** | **0.000** |
| Movement type | **17.987** | **0.000** | **30.796** | **0.000** | 1.446 | 0.097 |
| Subject | **7.087** | **0.000** | **4.623** | **0.000** | **30.656** | **0.000** |
| Sex | 1.931 | 0.165 | 0.109 | 0.742 | 2.201 | 0.139 |
| Skin tone | **2.796** | **0.002** | **2.716** | **0.002** | **4.493** | **0.000** |
| Shirt material | 0.980 | 0.451 | 1.754 | 0.084 | 1.387 | 0.200 |
| Pants material | 1.144 | 0.334 | 1.313 | 0.242 | 1.755 | 0.095 |
| Shirt type | 0.432 | 0.730 | 0.901 | 0.441 | **2.749** | **0.042** |
| Pants type | 0.102 | 0.959 | 0.653 | 0.581 | 0.613 | 0.607 |
| Floor color | 2.614 | 0.074 | 2.051 | 0.130 | **7.072** | **0.001** |
| Floormat color | 1.654 | 0.176 | 1.083 | 0.356 | **6.996** | **0.000** |
| Wall color | 1.933 | 0.123 | 1.418 | 0.237 | **5.139** | **0.002** |
| Blinds color | **7.719** | **0.001** | **6.913** | **0.001** | **11.575** | **0.000** |
| Partition color | 1.144 | 0.319 | 2.412 | 0.091 | **5.577** | **0.004** |
| **ODAH1 + ODAH2** | | | | | | |
| Skin tone | 1,564 | 0,109 | 1,210 | 0,280 | **3,429** | **0,000** |
| **ODAH2 + ODAH3** | | | | | | |
| Floor color | 0,70 | 0,50 | 0,36 | 0,70 | 1,43 | 0,24 |
| Floormat color | 0,52 | 0,67 | 0,19 | 0,90 | 1,81 | 0,14 |
| Wall color | 0,72 | 0,54 | 0,32 | 0,81 | 1,71 | 0,17 |
| Blinds color | **4,23** | **0,02** | **3,65** | **0,03** | **3,62** | **0,03** |
| Partition color | 0,06 | 0,94 | 0,51 | 0,60 | 0,75 | 0,48 |

**Table 3.1:** ANOVA results per variable and error metric. Significant results (p-value $\leq 0.05$) are highlighted in bold. The ANOVA is performed on all three datasets combined, on ODAH1 and ODAH2 combined, and on ODAH2 and ODAH3 combined.

| | | MPJAE | | PA-MPJPE | | scale error | |
|---|---|---|---|---|---|---|---|
| group A | group B | $t$ | p-value | $t$ | p-value | $t$ | p-value |
| **ODAH1 + ODAH2 + ODAH3** | | | | | | | |
| f | m | 1.393 | 0.164 | -0.331 | 0.741 | 1.487 | 0.138 |
| **ODAH1 + ODAH2** | | | | | | | |
| f | m | 1.164 | 0.245 | -0.749 | 0.454 | 0.915 | 0.361 |
| **Dataset comparison** | | | | | | | |
| ODAH1 | ODAH2 | **-2.229** | **0.027** | **-2.332** | **0.020** | -4.094 | **0.000** |
| ODAH1 | ODAH3 | **-3.318** | **0.001** | **-3.606** | **0.000** | -6.728 | **0.000** |
| ODAH2 | ODAH3 | -1.453 | 0.147 | -1.256 | 0.210 | **-2.305** | **0.022** |
| ODAH1 | ODAH2 + ODAH3 | **-3.141** | **0.002** | **-3.387** | **0.001** | **-5.866** | **0.000** |

**Table 3.2:** Welch's t-test results comparing two groups, group A and group B. A positive t-statistic indicates that group B has a higher error than group A. Significant results ($p < 0.05$) are highlighted in bold.

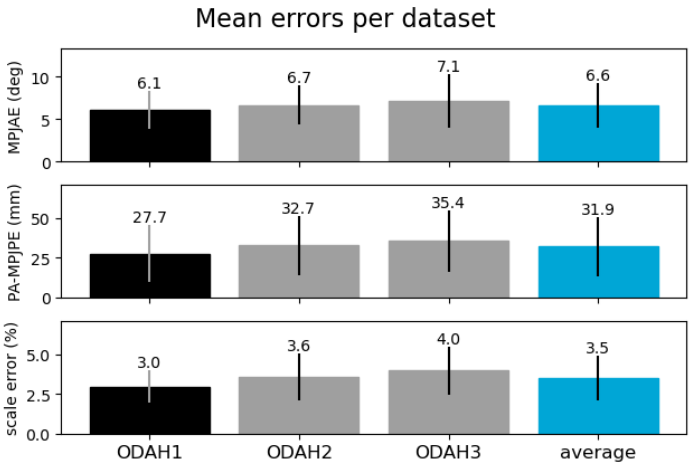


**Figure 3.2:** Mean errors and standard deviation for datasets. Bar height corresponds to the mean error, the vertical lines correspond to the standard deviation. The black bar denotes the test set corresponding to the train set used, gray bars denote datasets with increased diversity, the blue bar represents the average error and standard deviation of all datasets combined.

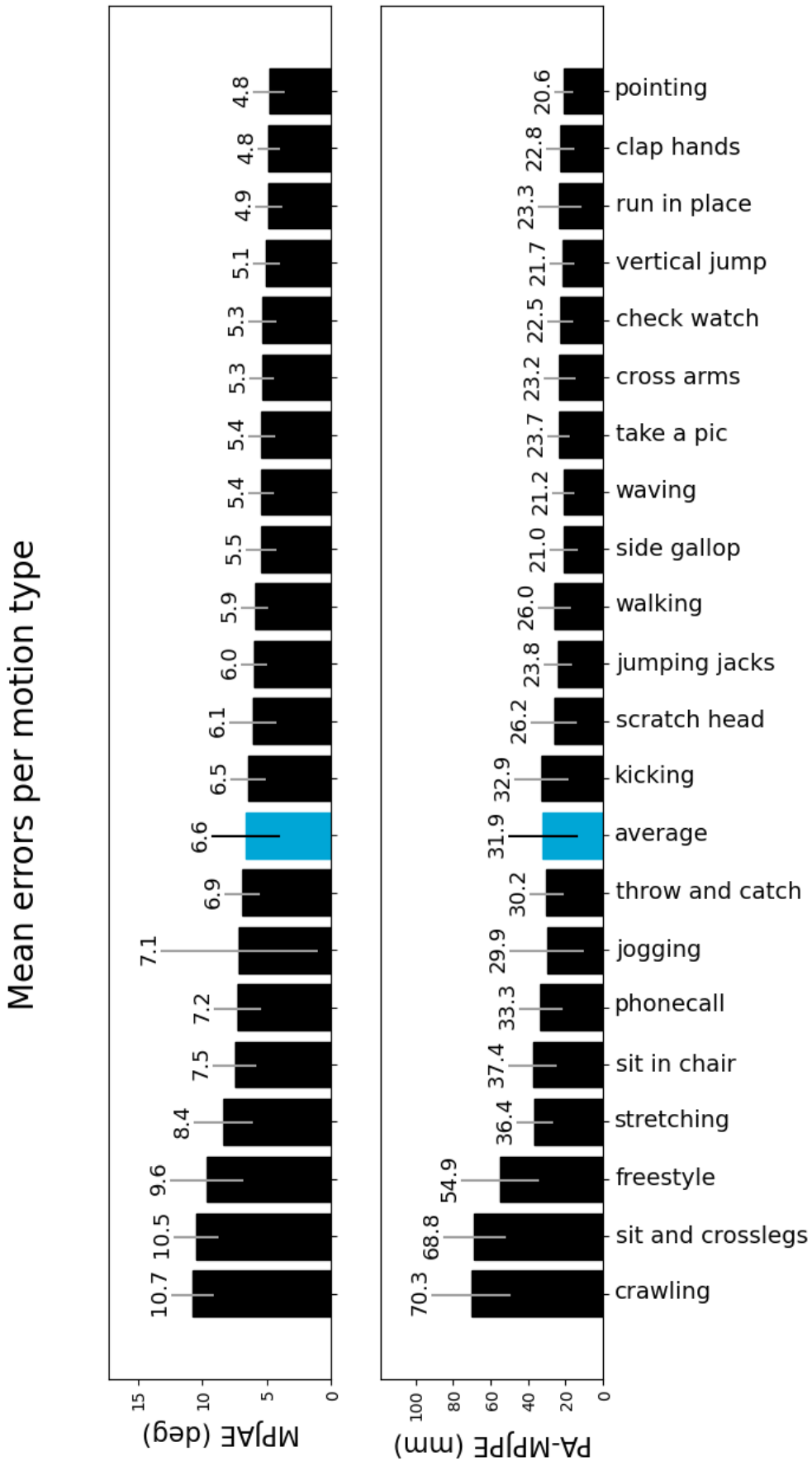


**Figure 3.3:** Mean errors and standard deviation for motion types. Bar height corresponds to the mean error, the vertical lines correspond to the standard deviation. Black bars denote variables included in the training data, gray bars denote unseen variables, the blue bar represents the total error and std of all datasets.

The movement type had the greatest differences in MPJAE and PA-MPJPE between variable categories. Figure 3.3 shows the mean and standard deviation of the MPJAE and PA-MPJPE for all motion types. The scale errors do not differ significantly. The MPJAE ranges from 4.82-10.74° and the PA-MPJAE ranges from 20.63-70.32 mm. For both metrics, the pointing motion resulted in the lowest error and crawling in the highest error.

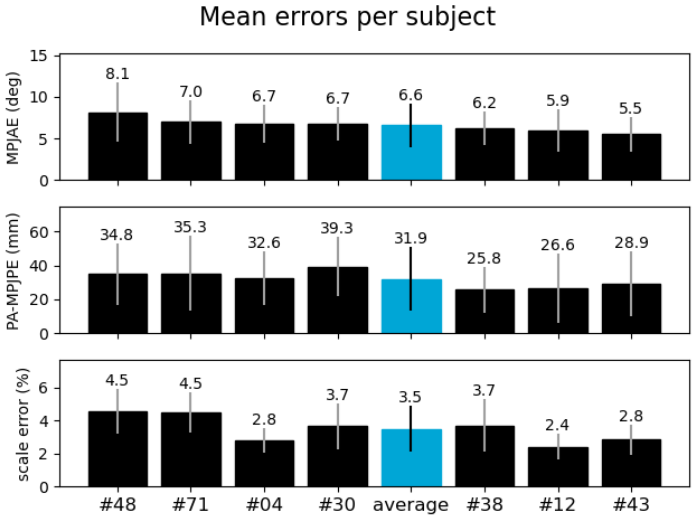


**Figure 3.4:** Mean errors + std for subjects. Bar height corresponds to the mean error, the lines correspond to the standard deviation. Black bars denote variables included in the trainin data, grey bars denote unseen variables, the blue bar represents the total error and std of all datasets.

The model performs significantly differently on the different subjects for all performance metrics (Fig. 3.4). The MPJAE, PA-MPJPE and scale error range from 5.52-8.14°, 25.76-35.33 mm, and 2.40-4.53%, respectively. The highest and lowest performing subjects differ per error metric.

When combining all three datasets, all performance metrics are significantly different for different skin tones (Fig. 3.5). The MPJAE, PA-MPJPE and scale error range from 5.90-8.57°, 26.77-43.93 mm, and 2.80-4.17%, respectively. The two original skin tones (og f and og m) have the lowest error for all metrics. When combining only ODAH1 and ODAH2, only the scale error shows a significant effect, with a difference of 1.6% between the best and worst performing skin tones (Table 3.1, Fig. 3.5).

Different shirt types had significantly different scale errors. The short sleeves resulted in the highest scale error (3.76%), and the half long sleeve had the lowest scale error (3.33%) (Fig. 3.6).

When all three datasets are combined, all environmental variables significantly affected the scale error (Fig. 3.7). The blinds, floor, floormat, wall, and partition colors had differences of 0.8%, 0.7%, 0.8%, 0.8%, and 0.6%, respectively, between the best and worst performing colors. Only the blinds color also showed significantly different MPJAE and PA-MPJPE per category, with differences of 1.4° and 9.6 mm, respectively (Fig. 3.8.a, Table 3.1). On ODAH2 and ODAH3 combined, the blinds color affected all three error metrics significantly (Table 3.1). The MPJAE, PA-MPJPE and scale error have maximum differences of 1.2°, 7.6 mm, and 0.6%, respectively (Fig. 3.8.b) No other environmental color variables had a significant effect.

| | **MPJAE** | | **PA-MPJPE** | | **scale error** | |
|---|---|---|---|---|---|---|
| | $r$ | p-value | $r$ | p-value | $r$ | p-value |
| weight | **0.105** | **0.028** | **0.140** | **0.003** | **0.123** | **0.010** |
| height | -0.051 | 0.287 | 0.006 | 0.897 | **0.142** | **0.003** |
| bmi | **0.151** | **0.002** | **0.113** | **0.018** | -0.030 | 0.526 |
| age | -0.043 | 0.365 | -0.023 | 0.625 | -0.032 | 0.506 |

**Table 3.3:** Pearson correlation test results for ODAH1, 2 and 3 combined. Significant results (p-value $\leq 0.05$) are highlighted in bold.

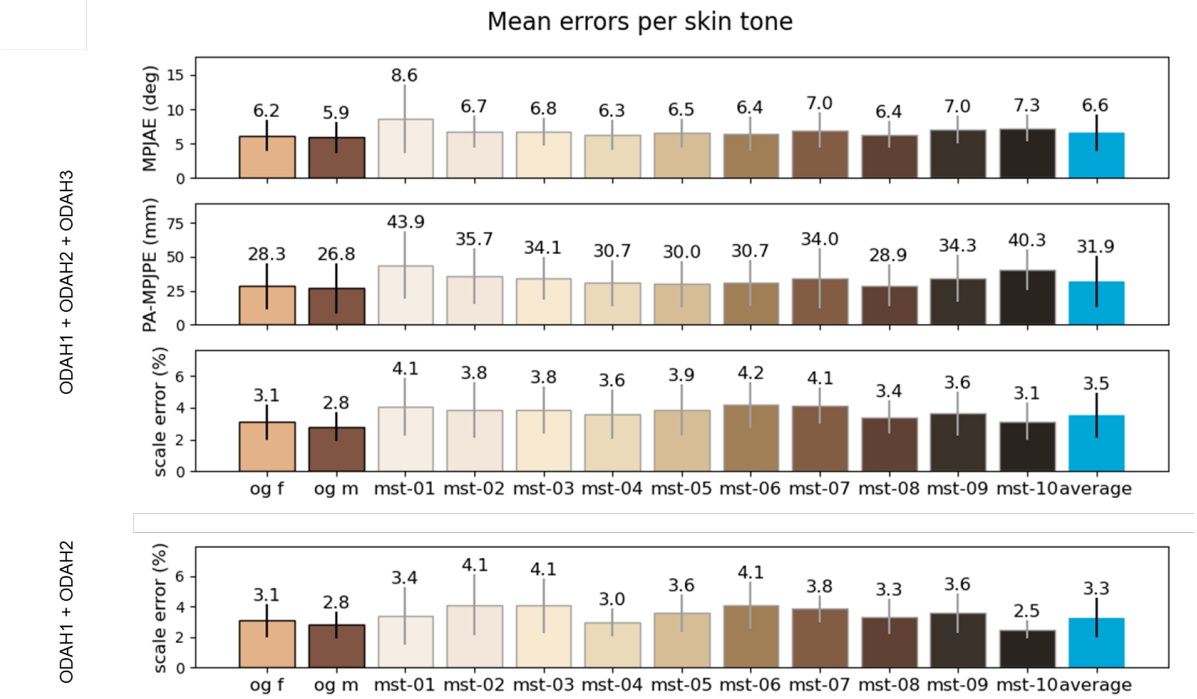


**Figure 3.5:** Mean errors and standard deviation for skin tones. Bar height corresponds to the mean error, the lines correspond to the standard deviation. Bar colors represent the skin colors. Black lines denote variables included in the training data, gray lines denote unseen variables.

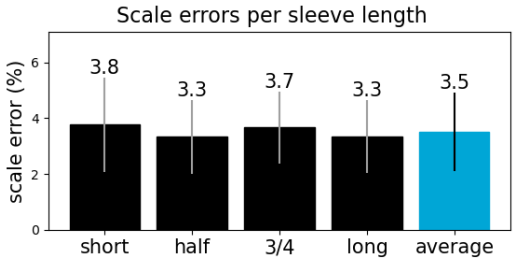


**Figure 3.6:** Mean scale errors standard deviation for shirt type. Bar height corresponds to the mean error, the lines correspond to the standard deviation. Black bars denote variables included in the training data, gray bars denote unseen variables, the blue bar represents the average error and standard deviation of all datasets.

The correlation coefficients $r$ of the t-test that describe the linear relationship between the quantitative variables and the different errors are shown in Table 3.3, along with the associated p-values. This shows that weight had a significant relationship with error types, height with the scale error, and BMI with the MPJAE and PA-MPJPE. Increased height, weight, and BMI led to a higher MPJAE (weight and BMI), a higher PA-MPJPE (weight, BMI), and scale error (weight, height), as can be seen from the positive r-values. The correlation coefficients range from 0.105 to 0.151 for the significant results.

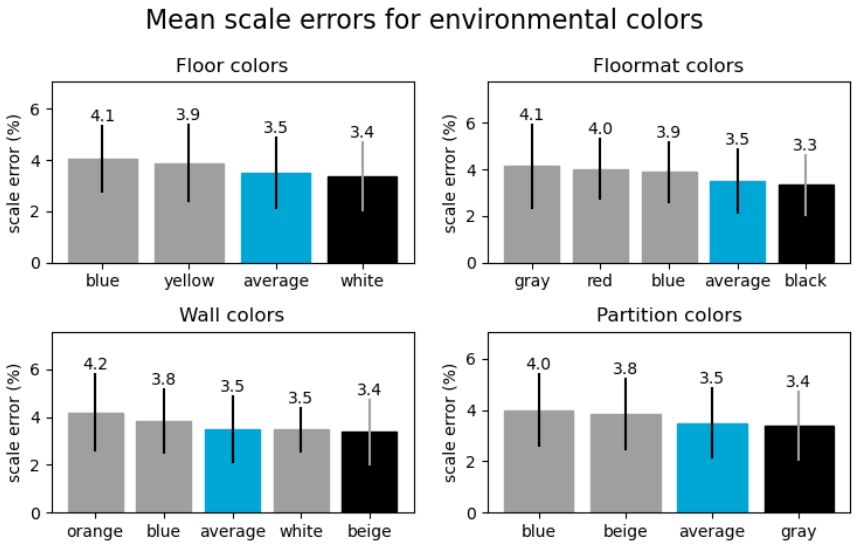


**Figure 3.7:** Mean scale errors and standard deviation for floor, floormat, wall, and partition colors. Bar height corresponds to the mean error, the lines correspond to the standard deviation. Black bars denote variables included in the training data, gray bars denote unseen variables, the blue bar represents the average error and standard deviation of all datasets combined.

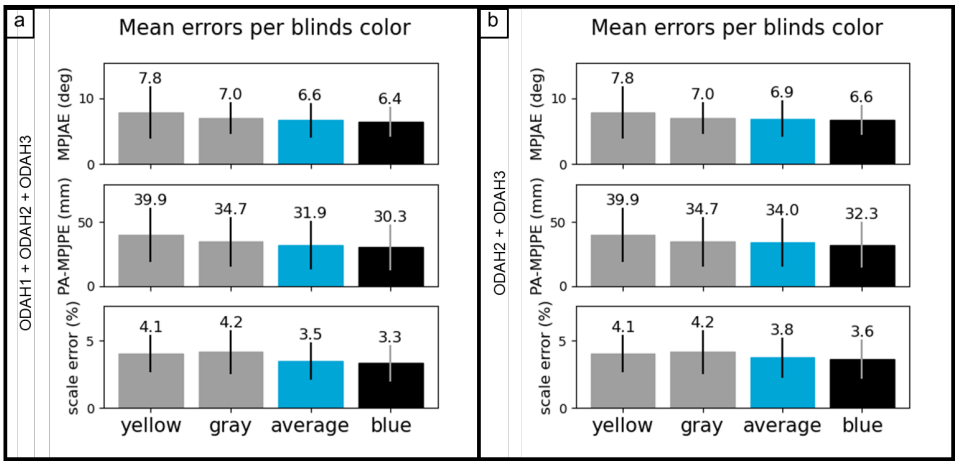


**Figure 3.8:** Mean errors standard deviation for blinds colors. Bar height corresponds to the mean error, the lines correspond to the standard deviation. Black bars denote variables included in the training data, gray bars denote unseen variables, the blue bar represents the average error and standard deviation of all datasets included. Figure a shows the results for all three datasets combined, and Figure b shows the results for ODAH2 and ODAH3 combined.

# 4

# Discussion

We show that the performance of a markerless motion capture model decreases when skin tone diversity is increased and when environmental colors are changed in the testing data with respect to the training data. Additionally, the model's accuracy is affected by the subject's motion, weight, height, BMI, and sleeve length.

In the results it can be seen that the dataset that matches the training data (ODAH1) performs the best on all performance metrics. ODAH2 only uses different variable categories for the skin tone variable, and this dataset shows a lower accuracy. ODAH3 has multiple new unseen variables for the model, and this dataset has lower performance than both ODAH1 and ODAH2. This shows that increasing the diversity of subjects' skin tones and adding variability to environmental colors will result in lower performance of a markerless motion capture model.

The differences in the MPJAE and PA-MPJPE for videos with different motion types were very large. The MPJAE doubles from the motion with the lowest MPJAE (pointing, MPJAE = $4.82°$) to the motion with the highest MPJAE (crawling, MPJAE = 10.74°). The lowest PA-MPJPE (pointing, PA-MPJPE = 20.63 mm) multiplies by almost 3.5 to the highest PA-MPJPE (crawling, PA-MPJPE = 70.32 mm). However, it is important to note that these errors are averaged over all joints, and these means should be interpreted with caution. Not all joints are equally relevant for every motion. For example, pointing, the motion with the lowest mean error, keeps most of the body static, while moving the shoulder, elbow, and wrist. Inspection of the joint errors of this motion has shown that the coordinate errors for the shoulder, elbow, and wrist are closer to the average for all motions compared to the errors for all coordinates combined. Additionally, the absolute sizes of the errors give more information when considering every coordinate separately. These are examples why it is important to study the coordinate specific errors more closely. An overview of all average coordinate angle errors and Procrustes aligned angle errors per motion type can be found in Appendix A (Table A.4, Table A.5).

Significant differences were found between subjects, and height, weight, and BMI also affected the model's performance. It is important to note that the test sets used only include seven different subjects, with fixed sex, height, weight, and body type per subject. Therefore, it is not possible to fully analyze the effects of these variables separately, as this would likely lead to a confounding bias. Furthermore, this study does not use a measurable variable for body type. Although BMI was included in the analysis, this is not a perfect metric to represent the different body types within the human population. BMI does not measure fat mass (Müller et al., 2010; Nuttall, 2015) or body mass distribution (Nuttall, 2015), which both affect different visual parameters of the body. Additionally, the synthetically created videos do not show the skin and fat movement that often increases with a person's weight. Assuming that these movements would increase the error, it is likely that this study underestimates the effect of weight and BMI on the model accuracy. Therefore, it is important to do more research on the effects of these variables. Ignoring these differences could lead to poorer model performance for different patients.

A possible additional limitation of the sex variable is that the OpenSim model, which is used both for rendering the different videos and as the ground truth for the accuracy calculation, is a male model

for both male and female subjects. Although this study did not find significant differences between the female and male subjects, more research is needed to analyze the effect of the model generalization.

Noticeable in the accuracy results per skin color is that both the lightest and darkest skin tones have the lowest accuracy, with the accuracy improving toward the middle skin tones. This is likely caused by the middle skin tones' similarity to the original skin tones that the model was trained on. The mean MPJAE of a skin tone ranges from $5.9° \pm 2.3°$ to $8.6° \pm 5.0°$, and the PA-MPJPE ranges from $27 \pm 18$ to $44 \pm 25$ mm. These are large differences that could affect the usability of a markerless motion capture model. This shows the importance of using a training dataset with a wide variety of skin tones and of testing models more extensively.

Most clothing variables did not significantly affect the model's performance. Only the shirt type had a significant effect, and only on the scale error. However, it is important to note that the garments only varied in sleeve/pants length. All clothing was skin tight, so the possible effect of different clothes fittings has not been studied.

All edited environmental variables significantly affected the scale error, and the color of the blinds affected all performance metrics. The blinds affected the scale error more strongly than the other environmental variables. All seen variable categories had lower errors than unseen variables.

As is clearly shown in the reduced accuracy for ODAH2 and ODAH3, the markerless motion capture model performs worse on unseen skin tones and environmental colors. This shows that these variables do not generalize well, and it is necessary to mitigate this issue by increasing the diversity in the training data. The model performed better on seen variable categories than unseen variable categories, implying that training with more diverse data could improve the model's final performance. One way to do this is by creating diverse synthetic data with methods described in this study, as this improves the ease with which you can change and keep track of variables. When using real video data, it is important to keep track of the variables used. The Monk Skin Tone scale can be used for categorizing skin tones from real people (Monk, 2019). Additionally, future research should focus on retraining the neural network with more diverse training data to definitively show the effects of training data diversity on model performance.

The results have also shown significant effects of variables that were not changes with respect to the original test data, like motion type and subjects and their characteristics. These differences are not caused by a lack of training data diversity, but are likely the result of some bias within the neural network. Therefore, reducing this effect requires a different approach. Knowledge of these biases in the model is the first step towards reducing them. However, most of these variables cannot be changed, and while motion types could be changed to an extent, in a medical setting, one motion type will not always be interchangeable with a different motion type because they might give different insights into a patient's movement. Additionally, height, weight, and BMI affected the performance of the model, while these variables of the test subjects were within the range of those of the training subjects. Analyses of the effects of these body variables on model performance are not commonly included in the testing of markerless motion capture models, despite the medical, biomechanical, and societal relevance. This study shows the importance of including this analysis in future studies. When more insight is gained into this bias, additional research should be done on reducing it.

It is noticeable that the scale error was significantly different for more variables than the MPJAE and PA-MPJPE. This can be explained by the fact that MPJAE and PA-MPJAE are not directly affected by scale errors. The MPJAE only tests angles, and the PA-MPJPE aligns with the ground truth for global translation, rotation, and scale. The Procrustes alignment of the MPJPE is necessary because no starting point within the global coordinate system is specified within the used model, so the non-Procrustes aligned MPJPE has average values over 200 mm. This means that it is important to evaluate the scale error separately, as this would affect the joint positions when using the model, when ground-truth data is not available to perform Procrustes alignment. The model used in this study does not fix the scale between video frames, which could affect the size of the errors.

# 5

# Conclusion

This study has found that multiple subject and environmental variables significantly affect the accuracy of a markerless motion capture model. We show the importance of using a diverse dataset for training and testing motion capture models. This study also shows the importance of analyzing motion capture accuracy more closely, because the accuracy for people of different skin tones, sexes, and body types varies. It is important for future motion capture research to take these findings into account when creating datasets and when creating, training, and testing neural networks. Bigger datasets should be created using the methods from this study, so networks can be (re-)trained and bigger tests can be performed. Additional variables, including body type and clothing type, should be further studied. Diversity matters, and this knowledge should be applied for better, more equal motion capture.

# References


Black, M. J., Patel, P., Tesch, J., & Yang, J. (2023). Bedlam: A synthetic dataset of bodies exhibiting detailed lifelike animated motion. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8726–8737.

Cappozzo, A., Catani, F., Leardini, A., Benedetti, M. G., & Della Croce, U. (1996). Position and orientation in space of bones during movement: Experimental artefacts. *Clinical Biomechanics*, *11*(2), 90–100. https://doi.org/https://doi.org/10.1016/0268-0033(95)00046-1

Colyer, S. L., Evans, M., Cosker, D. P., & Salo, A. I. T. (2018). A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system. *Sports Med Open*, *4*(1), 24. https://doi.org/10.1186/s40798-018-0139-y

Delp, S. L., Anderson, F. C., Arnold, A. S., Loan, P., Habib, A., John, C. T., Guendelman, E., & Thelen, D. G. (2007). Opensim: Open-source software to create and analyze dynamic simulations of movement. *IEEE Trans Biomed Eng*, *54*(11), 1940–50. https://doi.org/10.1109/tbme.2007.901024

FitzGerald, C., & Hurst, S. (2017). Implicit bias in healthcare professionals: A systematic review. *BMC Med Ethics*, *18*(1), 19. https://doi.org/10.1186/s12910-017-0179-8

Géron, A. (2023). Hands-on machine learning with scikit-learn, keras and tensorflow : Concepts, tools, and techniques to build intelligent systems. https://www.oreilly.com/library/view/-/9781098125967/

Ghorbani, S., Mahdaviani, K., Thaler, A., Kording, K., Cook, D. J., Blohm, G., & Troje, N. F. (2021). Movi: A large multi-purpose human motion and video dataset. *PLOS ONE*, *16*(6), e0253157. https://doi.org/10.1371/journal.pone.0253157

Gorton, G. E., Hebert, D. A., & Gannotti, M. E. (2009). Assessment of the kinematic variability among 12 motion analysis laboratories. *Gait & Posture*, *29*(3), 398–402. https://doi.org/https://doi.org/10.1016/j.gaitpost.2008.10.060

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, *40*, 33–51. https://doi.org/https://doi.org/10.1007/BF02291478

Habermann, M., Xu, W., Zollhofer, M., Pons-Moll, G., & Theobalt, C. (2020). Deepcap: Monocular human performance capture using weak supervision. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5052–5063.

Ionescu, C., Papava, D., Olaru, V., & Sminchisescu, C. (2014). Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*(7), 1325–1339. https://doi.org/10.1109/TPAMI.2013.248

Kanko, R., Laende, E. K., Davis, E. M., Selbie, W. S., & Deluzio, K. J. (2021). Concurrent assessment of gait kinematics using marker-based and markerless motion capture. *bioRxiv*, 2020.12.10.420075. https://doi.org/10.1101/2020.12.10.420075

Kanko, R. M., Laende, E., Selbie, W. S., & Deluzio, K. J. (2021). Inter-session repeatability of markerless motion capture gait kinematics. *Journal of Biomechanics*, *121*, 110422. https://doi.org/https://doi.org/10.1016/j.jbiomech.2021.110422

Lawrence, J., Bernal, J., & Witzgall, C. (2019). A purely algebraic justification of the kabsch-umeyama algorithm. *Journal of research of the National Institute of Standards and Technology*, *124*, 1. https://doi.org/10.6028/jres.124.028

Lin, Z.-Y., Lyu, B., Fernandez, J. C., Van Der Kruk, E., Seth, A., & Zhang, X. (2024). 3d kinematics estimation from video with a biomechanical model and synthetic training data. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1441–1450.

Lyu, B. (2023). *Synthetic human motion video generation based on biomechanical model* [Thesis]. Delft University of Technology.

Mahmood, N., Ghorbani, N., Troje, N. F., Pons-Moll, G., & Black, M. J. (2019). Amass: Archive of motion capture as surface shapes. *Proceedings of the IEEE/CVF international conference on computer vision*, 5442–5451.

Monk, E. (2019). Monk skin tone scale. https://skintone.google

Müller, M. J., Bosy-Westphal, A., & Krawczak, M. (2010). Genetic studies of common types of obesity: A critique of the current use of phenotypes. *Obesity Reviews*, *11*(8), 612–618. https://doi.org/https://doi.org/10.1111/j.1467-789X.2010.00734.x

Nuttall, F. Q. (2015). Body mass index: Obesity, bmi, and health: A critical review. *Nutrition Today*, *50*(3), 117–128. https://doi.org/10.1097/nt.0000000000000092

Pavlakos, G., Choutas, V., Ghorbani, N., Bolkart, T., Osman, A. A., Tzionas, D., & Black, M. J. (2019). Expressive body capture: 3d hands, face, and body from a single image. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10975–10985.

Pearson, K., & Galton, F. (1895). Vii. note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, *58*(347-352), 240–242. https://doi.org/10.1098/rspl.1895.0041

Rajagopal, A., Dembia, C. L., DeMers, M. S., Delp, D. D., Hicks, J. L., & Delp, S. L. (2016). Full-body musculoskeletal model for muscle-driven simulation of human gait. *IEEE Trans Biomed Eng*, *63*(10), 2068–79. https://doi.org/10.1109/tbme.2016.2586891

Ross, A., & Willson, V. L. (2017). One-way anova. In A. Ross & V. L. Willson (Eds.), *Basic and advanced statistical tests: Writing results sections and creating tables and figures* (pp. 21–24). SensePublishers. https://doi.org/10.1007/978-94-6351-086-8_5

Salisu, S., Ruhaiyem, N. I. R., Eisa, T. A. E., Nasser, M., Saeed, F., & Younis, H. A. (2023). Motion capture technologies for ergonomics: A systematic literature review. *Diagnostics*, *13*(15), 2593. https://doi.org/https://doi.org/10.3390/diagnostics13152593

Scott, B., Seyres, M., Philp, F., Chadwick, E. K., & Blana, D. (2022). Healthcare applications of single camera markerless motion capture: A scoping review. *PeerJ*, *10*, e13517. https://doi.org/10.7717/peerj.13517

Seth, A., Hicks, J. L., Uchida, T. K., Habib, A., Dembia, C. L., Dunne, J. J., Ong, C. F., DeMers, M. S., Rajagopal, A., Millard, M., Hamner, S. R., Arnold, E. M., Yong, J. R., Lakshmikanth, S. K., Sherman, M. A., Ku, J. P., & Delp, S. L. (2018). Opensim: Simulating musculoskeletal dynamics and neuromuscular control to study human and animal movement. *PLoS Comput Biol*, *14*(7), e1006223. https://doi.org/10.1371/journal.pcbi.1006223

Sigal, L., Balan, A. O., & Black, M. J. (2010). Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. *International Journal of Computer Vision*, *87*(1), 4–27. https://doi.org/10.1007/s11263-009-0273-6

Sinclair, J., Hebron, J., & Taylor, P. J. (2014). The influence of tester experience on the reliability of 3d kinematic information during running. *Gait & Posture*, *40*(4), 707–711. https://doi.org/https://doi.org/10.1016/j.gaitpost.2014.06.004

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Welch, B. L. (1947). The generalisation of student's problems when several different population variances are involved. *Biometrika*, *34*(1-2), 28–35. https://doi.org/10.1093/biomet/34.1-2.28

Yucer, S., Tektas, F., Moubayed, N. A., & Breckon, T. (2024). Racial bias within face recognition: A survey. *ACM Comput. Surv.*, *57*(4), Article 105. https://doi.org/10.1145/3705295

# A

# Additional tables

## A.1. Mean errors and standard deviation for every variable category

### A.1.1. ODAH1 + ODAH2 + ODAH3

[!]

| | | MPJAE (deg) | | PA-MPJPE (mm) | | Scale error (%) | |
|---|---|---|---|---|---|---|---|
| Category | Count | mean | std | mean | sd | mean | std |
| **Dataset** | | | | | | | |
| ODAH1 | 146 | 6.07 | 2.26 | 27.66 | 17.73 | 2.96 | 1.02 |
| ODAH2 | 146 | 6.66 | 2.28 | 32.65 | 18.83 | 3.57 | 1.48 |
| ODAH3 | 146 | 7.12 | 3.12 | 35.44 | 19.11 | 3.97 | 1.50 |
| **Motion type** | | | | | | | |
| crawling | 21 | 10.74 | 1.65 | 70.32 | 20.93 | 4.32 | 1.78 |
| sit and cross legs | 21 | 10.50 | 1.75 | 68.82 | 16.71 | 4.28 | 1.35 |
| freestyle | 21 | 9.64 | 2.84 | 54.89 | 20.72 | 3.38 | 1.51 |
| stretching | 21 | 8.39 | 2.29 | 36.42 | 9.90 | 3.81 | 1.61 |
| sit chair | 21 | 7.47 | 1.71 | 37.44 | 13.08 | 3.78 | 1.50 |
| phonecall | 21 | 7.23 | 1.84 | 33.27 | 11.58 | 3.43 | 1.30 |
| jogging | 18 | 7.13 | 6.12 | 29.87 | 20.01 | 3.72 | 1.44 |
| throw and catch | 21 | 6.92 | 1.38 | 30.19 | 9.19 | 3.12 | 0.88 |
| kicking | 21 | 6.45 | 1.37 | 32.88 | 14.79 | 3.13 | 1.33 |
| scratch head | 21 | 6.07 | 1.88 | 26.18 | 12.13 | 3.73 | 1.39 |
| jumping jacks | 21 | 5.99 | 0.98 | 23.82 | 7.58 | 3.71 | 1.27 |
| walking | 21 | 5.93 | 1.05 | 26.03 | 8.92 | 3.30 | 0.88 |
| side gallop | 21 | 5.46 | 1.19 | 20.97 | 7.61 | 3.55 | 1.43 |
| waving | 21 | 5.44 | 0.98 | 21.18 | 6.09 | 3.44 | 1.34 |
| take a picture | 21 | 5.40 | 1.04 | 23.68 | 5.94 | 3.53 | 1.52 |
| cross arms | 21 | 5.35 | 0.96 | 23.15 | 8.29 | 3.39 | 1.69 |
| check watch | 21 | 5.33 | 1.07 | 22.51 | 6.91 | 3.33 | 1.18 |
| jump vertically | 21 | 5.05 | 1.06 | 21.66 | 6.68 | 2.81 | 0.76 |
| run in place | 21 | 4.87 | 1.05 | 23.29 | 11.61 | 3.34 | 1.53 |
| clap hands | 21 | 4.84 | 0.87 | 22.77 | 7.61 | 3.23 | 1.49 |
| pointing | 21 | 4.82 | 1.25 | 20.63 | 5.10 | 3.15 | 0.91 |
| **Subject** | | | | | | | |
| 48 | 63 | 8.14 | 3.58 | 34.77 | 17.94 | 4.53 | 1.37 |
| 71 | 63 | 6.99 | 2.61 | 35.33 | 22.03 | 4.49 | 1.21 |
| 4 | 63 | 6.74 | 2.28 | 32.55 | 15.98 | 2.82 | 0.75 |
| 30 | 63 | 6.74 | 2.04 | 39.31 | 17.46 | 3.65 | 1.38 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 38 | 63 | 6.22 | 1.98 | 25.76 | 13.58 | 3.70 | 1.58 |
| 12 | 60 | 5.94 | 2.58 | 26.55 | 20.57 | 2.40 | 0.77 |
| 43 | 63 | 5.52 | 2.05 | 28.90 | 19.10 | 2.85 | 0.91 |
| **Sex** | | | | | | | |
| f | 249 | 6.77 | 2.76 | 31.66 | 18.43 | 3.59 | 1.52 |
| m | 189 | 6.42 | 2.41 | 32.26 | 19.38 | 3.38 | 1.25 |
| **Skin tone** | | | | | | | |
| og f | 83 | 6.20 | 2.22 | 28.33 | 17.12 | 3.08 | 1.09 |
| og m | 63 | 5.90 | 2.31 | 26.77 | 18.47 | 2.80 | 0.89 |
| mst-01 | 34 | 8.57 | 4.99 | 43.93 | 24.72 | 4.07 | 1.80 |
| mst-02 | 27 | 6.75 | 2.39 | 35.66 | 20.52 | 3.84 | 1.74 |
| mst-03 | 31 | 6.78 | 2.03 | 34.10 | 15.74 | 3.84 | 1.50 |
| mst-04 | 39 | 6.27 | 2.18 | 30.67 | 16.97 | 3.58 | 1.51 |
| mst-05 | 29 | 6.54 | 2.11 | 30.03 | 16.86 | 3.88 | 1.62 |
| mst-06 | 25 | 6.45 | 2.51 | 30.71 | 16.91 | 4.17 | 1.44 |
| mst-07 | 23 | 6.97 | 2.51 | 33.95 | 21.93 | 4.11 | 1.17 |
| mst-08 | 32 | 6.36 | 1.91 | 28.87 | 15.57 | 3.40 | 1.02 |
| mst-09 | 34 | 7.03 | 1.98 | 34.25 | 17.39 | 3.62 | 1.37 |
| mst-10 | 18 | 7.27 | 1.90 | 40.25 | 14.83 | 3.13 | 1.20 |
| **shirt material** | | | | | | | |
| 7: blue | 16 | 6.63 | 1.59 | 31.93 | 8.89 | 4.03 | 1.42 |
| 8: green-plaid | 17 | 6.49 | 1.47 | 27.07 | 7.65 | 3.91 | 1.08 |
| 2: black-white | 59 | 7.18 | 2.69 | 37.50 | 22.42 | 3.76 | 1.70 |
| 4: gray | 61 | 6.81 | 2.79 | 34.51 | 22.86 | 3.68 | 1.53 |
| 9: white-beige | 13 | 6.90 | 1.95 | 34.13 | 16.68 | 3.46 | 1.18 |
| 3: red-black | 73 | 6.11 | 1.78 | 27.71 | 13.02 | 3.45 | 1.36 |
| 5: blue-striped | 62 | 6.25 | 1.91 | 28.47 | 14.69 | 3.40 | 1.37 |
| 6: black | 61 | 6.88 | 2.70 | 33.65 | 21.04 | 3.32 | 1.24 |
| 1: gray-red | 76 | 6.58 | 3.67 | 31.68 | 20.01 | 3.22 | 1.26 |
| **pants material** | | | | | | | |
| 6: floral | 13 | 7.56 | 2.65 | 40.42 | 23.01 | 4.03 | 1.49 |
| 8: brown | 21 | 7.49 | 2.34 | 39.09 | 20.25 | 4.18 | 1.51 |
| 5: blue-white | 82 | 6.84 | 2.53 | 32.86 | 19.58 | 3.34 | 1.29 |
| 4: black | 95 | 6.78 | 3.50 | 32.80 | 19.97 | 3.54 | 1.43 |
| 3: yellow | 74 | 6.42 | 2.41 | 31.32 | 19.35 | 3.49 | 1.38 |
| 1: black-white | 70 | 6.36 | 2.07 | 30.16 | 17.06 | 3.22 | 1.21 |
| 7: green | 16 | 6.36 | 1.20 | 30.63 | 11.06 | 3.91 | 1.64 |
| 2: gray | 67 | 6.19 | 2.14 | 28.43 | 16.13 | 3.53 | 1.52 |
| **shirt type** | | | | | | | |
| short | 95 | 6.43 | 2.16 | 29.71 | 16.43 | 3.76 | 1.69 |
| half | 153 | 6.72 | 3.29 | 33.18 | 21.66 | 3.33 | 1.32 |
| 3/4 | 97 | 6.48 | 1.89 | 30.87 | 15.37 | 3.66 | 1.29 |
| long | 93 | 6.77 | 2.44 | 33.19 | 19.20 | 3.34 | 1.30 |
| **pants type** | | | | | | | |
| short | 106 | 6.52 | 2.18 | 30.09 | 16.14 | 3.35 | 1.43 |
| half | 112 | 6.65 | 3.30 | 31.39 | 20.30 | 3.48 | 1.46 |
| 3/4 | 99 | 6.57 | 2.35 | 32.71 | 19.42 | 3.55 | 1.34 |
| long | 121 | 6.70 | 2.47 | 33.36 | 19.01 | 3.59 | 1.40 |
| **Floor color** | | | | | | | |
| floor blue | 45 | 7.05 | 2.29 | 35.76 | 16.62 | 4.06 | 1.33 |
| floor yellow | 49 | 7.25 | 4.20 | 35.05 | 21.01 | 3.89 | 1.54 |
| floor white | 344 | 6.47 | 2.33 | 30.97 | 18.68 | 3.37 | 1.37 |
| **Floormat color** | | | | | | | |
| floormat gray | 38 | 6.81 | 2.19 | 33.73 | 15.77 | 4.14 | 1.83 |
| floormat red | 37 | 7.41 | 4.46 | 36.28 | 18.15 | 4.03 | 1.35 |

| | | MPJAE | | PA-MPJPE | | Absolute scale error | |
|---|---|---|---|---|---|---|---|
| Category | Count | mean | std | mean | sd | mean | std |
| **Skin tone** | | | | | | | |
| og f | 83 | 6,20 | 2,22 | 28,33 | 17,12 | 3,08 | 1,09 |
| og m | 63 | 5,90 | 2,31 | 26,77 | 18,47 | 2,80 | 0,89 |
| mst-01 | 11 | 7,49 | 2,43 | 35,54 | 15,81 | 3,38 | 1,89 |
| mst-02 | 14 | 6,85 | 2,54 | 36,55 | 22,67 | 4,10 | 2,00 |
| mst-03 | 19 | 7,29 | 2,04 | 37,37 | 17,78 | 4,06 | 1,76 |
| mst-04 | 17 | 5,51 | 1,83 | 27,44 | 14,01 | 2,96 | 0,89 |
| mst-05 | 15 | 6,59 | 2,25 | 29,49 | 18,75 | 3,58 | 1,23 |
| mst-06 | 11 | 5,71 | 2,08 | 26,38 | 16,39 | 4,05 | 1,54 |
| mst-07 | 11 | 7,12 | 2,60 | 34,93 | 22,54 | 3,83 | 0,86 |
| mst-08 | 19 | 6,44 | 2,26 | 29,51 | 18,79 | 3,33 | 1,17 |
| mst-09 | 21 | 7,18 | 2,28 | 37,28 | 20,96 | 3,58 | 1,30 |
| mst-10 | 8 | 6,08 | 0,78 | 28,47 | 3,52 | 2,48 | 0,59 |

**Table A.2:** Mean errors and standard deviation for all videos with the shown variable categories. Videos from ODAH1 and ODAH2 are included in this calculation.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| floormat blue | 29 | 6.92 | 2.18 | 33.68 | 16.64 | 3.89 | 1.35 |
| floormat black | 334 | 6.48 | 2.40 | 31.08 | 19.32 | 3.33 | 1.32 |
| **Wall color** | | | | | | | |
| wall orange | 43 | 7.45 | 4.28 | 36.46 | 19.34 | 4.20 | 1.62 |
| wall blue | 36 | 6.89 | 2.56 | 34.67 | 19.49 | 3.82 | 1.38 |
| wall white | 32 | 6.70 | 2.14 | 32.86 | 17.47 | 3.47 | 0.97 |
| wall beige | 327 | 6.47 | 2.34 | 30.93 | 18.71 | 3.37 | 1.39 |
| **Blinds color** | | | | | | | |
| blinds yellow | 56 | 7.81 | 4.00 | 39.95 | 21.19 | 4.05 | 1.41 |
| blinds gray | 37 | 6.97 | 2.40 | 34.72 | 19.23 | 4.19 | 1.64 |
| blinds blue | 345 | 6.39 | 2.28 | 30.31 | 18.01 | 3.33 | 1.34 |
| **Partition color** | | | | | | | |
| partition blue | 47 | 7.00 | 2.53 | 34.82 | 18.67 | 3.98 | 1.43 |
| partition beige | 52 | 6.94 | 2.28 | 36.16 | 18.45 | 3.84 | 1.42 |
| partition gray | 339 | 6.51 | 2.67 | 30.87 | 18.79 | 3.38 | 1.38 |

**Table A.1:** Mean errors and standard deviation for all videos with the shown variable categories. Videos from all three datasets are included in this calculation.

### A.1.2. ODAH1+ODAH2

### A.1.3. ODAH2+ODAH3

## A.2. Mean coordinate errors per motion type

| | | MPJAE | | PA-MPJPE | | Absolute scale error | |
|---|---|---|---|---|---|---|---|
| Category | Count | mean | std | mean | sd | mean | std |
| **Sex** | | | | | | | |
| f | 166 | 7,05 | 2,95 | 33,32 | 18,82 | 3,84 | 1,63 |
| m | 126 | 6,68 | 2,42 | 35,00 | 19,24 | 3,68 | 1,30 |
| **Floor color** | | | | | | | |
| floor blue | 45 | 7,05 | 2,29 | 35,76 | 16,62 | 4,06 | 1,33 |
| floor yellow | 49 | 7,25 | 4,20 | 35,05 | 21,01 | 3,89 | 1,54 |
| floor white | 198 | 6,77 | 2,34 | 33,41 | 18,98 | 3,67 | 1,52 |
| **Floormat color** | | | | | | | |
| floormat gray | 38 | 6,81 | 2,19 | 33,73 | 15,77 | 4,14 | 1,83 |
| floormat red | 37 | 7,41 | 4,46 | 36,28 | 18,15 | 4,03 | 1,35 |
| floormat blue | 29 | 6,92 | 2,18 | 33,68 | 16,64 | 3,89 | 1,35 |
| floormat black | 188 | 6,80 | 2,45 | 33,73 | 20,08 | 3,62 | 1,45 |
| **Wall color** | | | | | | | |
| wall orange | 43 | 7,45 | 4,28 | 36,46 | 19,34 | 4,20 | 1,62 |
| wall blue | 36 | 6,89 | 2,56 | 34,67 | 19,49 | 3,82 | 1,38 |
| wall white | 32 | 6,70 | 2,14 | 32,86 | 17,47 | 3,47 | 0,97 |
| wall beige | 181 | 6,79 | 2,36 | 33,56 | 19,06 | 3,71 | 1,55 |
| **Blinds color** | | | | | | | |
| blinds yellow | 56 | 7,81 | 4,00 | 39,95 | 21,19 | 4,05 | 1,41 |
| blinds gray | 37 | 6,97 | 2,40 | 34,72 | 19,23 | 4,19 | 1,64 |
| blinds blue | 199 | 6,62 | 2,27 | 32,26 | 17,97 | 3,61 | 1,47 |
| **Partition color** | | | | | | | |
| partition blue | 47 | 7,00 | 2,53 | 34,82 | 18,67 | 3,98 | 1,43 |
| partition beige | 52 | 6,94 | 2,28 | 36,16 | 18,45 | 3,84 | 1,42 |
| partition gray | 193 | 6,85 | 2,90 | 33,29 | 19,21 | 3,70 | 1,53 |

**Table A.3:** Mean errors and standard deviation for all videos with the shown variable categories. Videos from ODAH2 and ODAH3 are included in this calculation.

| | pelvis tilt | pelvis list | pelvis rotation | hip flexion | hip adduction | hip rotation | knee angle | ankle angle | subtalar angle | mtp angle |
|---|---|---|---|---|---|---|---|---|---|---|
| check watch | 1,03 | 3,80 | 2,92 | 4,50 | 1,08 | 3,71 | 1,77 | 3,07 | 2,62 | 0,10 |
| clap hands | 0,94 | 3,68 | 4,92 | 4,32 | 1,06 | 3,69 | 1,66 | 4,29 | 2,95 | 0,08 |
| crawling | 10,42 | 8,07 | 43,59 | 13,55 | 6,37 | 9,87 | 8,93 | 11,69 | 7,05 | 0,24 |
| cross arms | 1,05 | 3,56 | 2,93 | 3,89 | 1,13 | 3,72 | 1,33 | 4,22 | 3,24 | 0,09 |
| sit cross legs | 9,91 | 9,00 | 12,98 | 12,44 | 10,70 | 10,62 | 6,25 | 9,60 | 9,97 | 0,22 |
| freestyle | 4,32 | 5,42 | 9,41 | 9,48 | 5,02 | 8,62 | 9,29 | 9,16 | 6,63 | 0,15 |
| jogging | 2,95 | 3,57 | 25,51 | 5,32 | 3,31 | 5,67 | 5,01 | 6,98 | 4,58 | 0,14 |
| jump vert | 1,82 | 3,94 | 3,80 | 4,60 | 1,95 | 3,58 | 3,13 | 4,14 | 3,63 | 0,16 |
| jumping jacks | 1,34 | 4,55 | 2,75 | 5,35 | 1,73 | 5,11 | 3,18 | 5,26 | 4,12 | 0,16 |
| kicking | 2,89 | 4,24 | 4,97 | 6,59 | 2,97 | 6,18 | 4,65 | 5,68 | 4,54 | 0,11 |
| phonecall | 2,43 | 2,66 | 6,60 | 3,97 | 2,03 | 6,07 | 3,56 | 5,33 | 3,37 | 0,13 |
| pointing | 1,62 | 3,77 | 3,47 | 4,24 | 1,25 | 3,08 | 1,47 | 3,64 | 2,67 | 0,08 |
| run in place | 1,85 | 3,42 | 4,23 | 4,31 | 2,04 | 4,08 | 3,20 | 4,41 | 3,21 | 0,13 |
| scratch head | 1,67 | 3,50 | 3,59 | 4,05 | 1,43 | 3,43 | 1,58 | 4,12 | 2,58 | 0,10 |
| side gallop | 2,14 | 4,02 | 4,14 | 4,99 | 2,57 | 4,59 | 2,92 | 4,36 | 3,63 | 0,10 |
| sit chair | 3,05 | 4,77 | 5,50 | 7,00 | 2,74 | 5,38 | 6,68 | 6,04 | 4,95 | 0,23 |
| stretching | 2,39 | 3,88 | 4,85 | 4,73 | 2,43 | 4,90 | 3,22 | 5,01 | 3,53 | 0,10 |
| take pic | 1,75 | 3,30 | 4,58 | 3,65 | 1,41 | 4,64 | 1,81 | 4,50 | 3,06 | 0,09 |
| throw catch | 2,11 | 3,91 | 5,16 | 4,77 | 2,34 | 5,64 | 3,38 | 5,10 | 3,36 | 0,10 |
| walking | 2,99 | 3,35 | 26,38 | 5,26 | 3,14 | 6,16 | 3,73 | 4,74 | 4,21 | 0,12 |
| waving | 1,09 | 3,82 | 4,03 | 4,32 | 1,07 | 3,34 | 1,73 | 4,03 | 2,48 | 0,08 |

| | lumbar extension | lumbar bending | lumbar rotation | arm flex | arm add | arm rot | elbow flex | pro sup | wrist flex | wrist dev |
|---|---|---|---|---|---|---|---|---|---|---|
| check watch | 5,79 | 1,23 | 2,42 | 5,88 | 3,14 | 10,62 | 6,84 | 16,33 | 0,20 | 0,51 |
| clap hands | 5,75 | 1,11 | 2,61 | 5,53 | 2,68 | 6,29 | 5,50 | 13,26 | 0,19 | 0,58 |
| crawling | 9,02 | 6,47 | 6,73 | 10,72 | 5,67 | 16,86 | 11,59 | 15,32 | 0,31 | 0,82 |
| cross arms | 5,31 | 1,28 | 2,49 | 5,92 | 3,68 | 8,18 | 6,78 | 16,90 | 0,23 | 0,62 |
| sit cross legs | 9,28 | 6,49 | 10,22 | 12,27 | 6,06 | 16,23 | 10,98 | 15,06 | 0,25 | 0,73 |
| freestyle | 6,91 | 4,20 | 4,74 | 10,57 | 5,42 | 15,91 | 12,53 | 18,94 | 0,27 | 0,67 |
| jogging | 6,46 | 2,79 | 3,14 | 4,96 | 3,61 | 6,77 | 10,92 | 17,87 | 0,23 | 0,83 |
| jump vert | 6,20 | 2,16 | 1,89 | 4,56 | 1,97 | 7,55 | 5,70 | 13,23 | 0,21 | 0,54 |
| jumping jacks | 6,31 | 1,46 | 2,44 | 47,43 | 3,71 | 20,87 | 7,51 | 12,99 | 0,31 | 0,69 |
| kicking | 5,79 | 2,85 | 3,78 | 4,68 | 2,62 | 8,30 | 6,86 | 16,74 | 0,20 | 0,53 |
| phonecall | 5,29 | 2,09 | 4,00 | 9,90 | 4,44 | 11,22 | 10,61 | 21,03 | 0,25 | 0,63 |
| pointing | 6,25 | 1,51 | 3,00 | 3,87 | 3,11 | 10,63 | 6,25 | 11,16 | 0,23 | 0,49 |
| run in place | 5,05 | 2,11 | 2,34 | 5,00 | 2,72 | 5,92 | 5,43 | 12,42 | 0,22 | 0,56 |
| scratch head | 5,41 | 1,86 | 2,85 | 8,00 | 3,08 | 13,62 | 7,61 | 19,67 | 0,26 | 0,65 |
| side gallop | 5,38 | 2,31 | 2,45 | 3,77 | 2,01 | 8,96 | 6,23 | 14,96 | 0,19 | 0,49 |
| sit chair | 6,59 | 3,08 | 2,89 | 7,83 | 4,13 | 12,15 | 9,62 | 17,62 | 0,20 | 0,61 |
| stretching | 5,11 | 2,60 | 3,97 | 44,66 | 6,00 | 26,86 | 13,31 | 21,24 | 0,29 | 0,67 |
| take pic | 5,26 | 1,73 | 3,46 | 5,39 | 3,25 | 8,71 | 6,14 | 16,46 | 0,21 | 0,64 |
| throw catch | 5,65 | 2,60 | 4,04 | 7,88 | 4,23 | 13,00 | 9,18 | 18,76 | 0,24 | 0,64 |
| walking | 5,26 | 2,70 | 3,04 | 4,56 | 2,32 | 10,28 | 6,84 | 12,84 | 0,19 | 0,56 |
| waving | 5,70 | 1,48 | 3,24 | 5,17 | 2,94 | 9,79 | 5,20 | 20,05 | 0,23 | 0,58 |

**Table A.4:** The mean angle error per coordinate for every motion type

| | pelvis tilt | pelvis list | pelvis rotation | hip flexion | hip adduction | hip rotation | knee angle | ankle angle | subtalar angle | mtp angle |
|---|---|---|---|---|---|---|---|---|---|---|
| check watch | 0,90 | 3,72 | 4,15 | 4,50 | 1,08 | 3,71 | 1,77 | 3,07 | 2,62 | 0,10 |
| clap hands | 0,91 | 3,71 | 2,57 | 4,32 | 1,06 | 3,69 | 1,66 | 4,29 | 2,95 | 0,08 |
| crawling | 9,27 | 7,51 | 36,80 | 13,55 | 6,37 | 9,87 | 8,93 | 11,69 | 7,05 | 0,24 |
| cross arms | 1,17 | 3,46 | 2,74 | 3,89 | 1,13 | 3,72 | 1,33 | 4,22 | 3,24 | 0,09 |
| sit cross legs | 9,35 | 8,93 | 12,41 | 12,44 | 10,70 | 10,62 | 6,25 | 9,60 | 9,97 | 0,22 |
| freestyle | 4,96 | 6,06 | 8,51 | 9,48 | 5,02 | 8,62 | 9,29 | 9,16 | 6,63 | 0,15 |
| jogging | 3,33 | 3,49 | 12,64 | 5,32 | 3,31 | 5,67 | 5,01 | 6,98 | 4,58 | 0,14 |
| jump vert | 1,64 | 4,06 | 2,57 | 4,60 | 1,95 | 3,58 | 3,13 | 4,14 | 3,63 | 0,16 |
| jumping jacks | 1,27 | 4,35 | 3,08 | 5,35 | 1,73 | 5,11 | 3,18 | 5,26 | 4,12 | 0,16 |
| kicking | 2,90 | 4,11 | 4,77 | 6,59 | 2,97 | 6,18 | 4,65 | 5,68 | 4,54 | 0,11 |
| phonecall | 2,11 | 2,59 | 5,56 | 3,97 | 2,03 | 6,07 | 3,56 | 5,33 | 3,37 | 0,13 |
| pointing | 1,37 | 3,82 | 3,10 | 4,24 | 1,25 | 3,08 | 1,47 | 3,64 | 2,67 | 0,08 |
| run in place | 1,70 | 3,37 | 3,14 | 4,31 | 2,04 | 4,08 | 3,20 | 4,41 | 3,21 | 0,13 |
| scratch head | 1,49 | 3,65 | 3,76 | 4,05 | 1,43 | 3,43 | 1,58 | 4,12 | 2,58 | 0,10 |
| side gallop | 2,07 | 3,77 | 2,92 | 4,99 | 2,57 | 4,59 | 2,92 | 4,36 | 3,63 | 0,10 |
| sit chair | 3,04 | 4,38 | 3,52 | 7,00 | 2,74 | 5,38 | 6,68 | 6,04 | 4,95 | 0,23 |
| stretching | 2,04 | 3,96 | 5,60 | 4,73 | 2,43 | 4,90 | 3,22 | 5,01 | 3,53 | 0,10 |
| take pic | 1,33 | 3,21 | 4,10 | 3,65 | 1,41 | 4,64 | 1,81 | 4,50 | 3,06 | 0,09 |
| throw catch | 1,98 | 3,84 | 4,38 | 4,77 | 2,34 | 5,64 | 3,38 | 5,10 | 3,36 | 0,10 |
| walking | 3,07 | 3,07 | 14,74 | 5,26 | 3,14 | 6,16 | 3,73 | 4,74 | 4,21 | 0,12 |
| waving | 0,98 | 3,75 | 2,42 | 4,32 | 1,07 | 3,34 | 1,73 | 4,03 | 2,48 | 0,08 |

| | lumbar extension | lumbar bending | lumbar rotation | arm flex | arm add | arm rot | elbow flex | pro sup | wrist flex | wrist dev |
|---|---|---|---|---|---|---|---|---|---|---|
| check watch | 5,79 | 1,23 | 2,42 | 5,88 | 3,14 | 10,62 | 6,84 | 16,33 | 0,20 | 0,51 |
| clap hands | 5,75 | 1,11 | 2,61 | 5,53 | 2,68 | 6,29 | 5,50 | 13,26 | 0,19 | 0,58 |
| crawling | 9,02 | 6,47 | 6,73 | 10,72 | 5,67 | 16,86 | 11,59 | 15,32 | 0,31 | 0,82 |
| cross arms | 5,31 | 1,28 | 2,49 | 5,92 | 3,68 | 8,18 | 6,78 | 16,90 | 0,23 | 0,62 |
| sit cross legs | 9,28 | 6,49 | 10,22 | 12,27 | 6,06 | 16,23 | 10,98 | 15,06 | 0,25 | 0,73 |
| freestyle | 6,91 | 4,20 | 4,74 | 10,57 | 5,42 | 15,91 | 12,53 | 18,94 | 0,27 | 0,67 |
| jogging | 6,46 | 2,79 | 3,14 | 4,96 | 3,61 | 6,77 | 10,92 | 17,87 | 0,23 | 0,83 |
| jump vert | 6,20 | 2,16 | 1,89 | 4,56 | 1,97 | 7,55 | 5,70 | 13,23 | 0,21 | 0,54 |
| jumping jacks | 6,31 | 1,46 | 2,44 | 47,43 | 3,71 | 20,87 | 7,51 | 12,99 | 0,31 | 0,69 |
| kicking | 5,79 | 2,85 | 3,78 | 4,68 | 2,62 | 8,30 | 6,86 | 16,74 | 0,20 | 0,53 |
| phonecall | 5,29 | 2,09 | 4,00 | 9,90 | 4,44 | 11,22 | 10,61 | 21,03 | 0,25 | 0,63 |
| pointing | 6,25 | 1,51 | 3,00 | 3,87 | 3,11 | 10,63 | 6,25 | 11,16 | 0,23 | 0,49 |
| run in place | 5,05 | 2,11 | 2,34 | 5,00 | 2,72 | 5,92 | 5,43 | 12,42 | 0,22 | 0,56 |
| scratch head | 5,41 | 1,86 | 2,85 | 8,00 | 3,08 | 13,62 | 7,61 | 19,67 | 0,26 | 0,65 |
| side gallop | 5,38 | 2,31 | 2,45 | 3,77 | 2,01 | 8,96 | 6,23 | 14,96 | 0,19 | 0,49 |
| sit chair | 6,59 | 3,08 | 2,89 | 7,83 | 4,13 | 12,15 | 9,62 | 17,62 | 0,20 | 0,61 |
| stretching | 5,11 | 2,60 | 3,97 | 44,66 | 6,00 | 26,86 | 13,31 | 21,24 | 0,29 | 0,67 |
| take pic | 5,26 | 1,73 | 3,46 | 5,39 | 3,25 | 8,71 | 6,14 | 16,46 | 0,21 | 0,64 |
| throw catch | 5,65 | 2,60 | 4,04 | 7,88 | 4,23 | 13,00 | 9,18 | 18,76 | 0,24 | 0,64 |
| walking | 5,26 | 2,70 | 3,04 | 4,56 | 2,32 | 10,28 | 6,84 | 12,84 | 0,19 | 0,56 |
| waving | 5,70 | 1,48 | 3,24 | 5,17 | 2,94 | 9,79 | 5,20 | 20,05 | 0,23 | 0,58 |

**Table A.5:** The mean Procrustes aligned angle error per coordinate for every motion type