TECHNISCHE UNIVERSITÄT
CHEMNITZ

Fakultät für Informatik

CSR-23-05

NaturalWalk: An Anatomy-based Synthesizer for Human Walking Motions

Samer Salamah · Guido Brunnett · Sabrina Bräuer ·
Tom Uhlmann · Oliver Rehren · Katharina Jahn ·
Thomas L. Milani · Günter Daniel Rey

März 2023

Chemnitzer Informatik-Berichte

NaturalWalk: An Anatomy-based Synthesizer for Human Walking Motions

Samer Salamah[1], Guido Brunnett[1], Sabrina Bräuer[2], Tom Uhlmann[1], Oliver Rehren[3], Katharina Jahn[4], Thomas L. Milani[2], Günter Daniel Rey[4]

[1] Faculty of Computer Science, Professorship of Computer Graphics and Visualization

[2] Faculty of Behavioural and Social Sciences, Professorship of Human Locomotion

[3] Faculty of Humanities, Institute for Media Research, Professorship of Media Psychology

[4] Faculty of Humanities, Institute for Media Research, Professorship of Psychology of Learning with Digital Media

Samer.Salamah@informatik.tu-chemnitz.de

Abstract: We present a novel data-driven approach for synthesizing human gait motions with individual style characteristics and natural appearance. Our approach is based on the concept of a *motion signature* that captures the essential characteristic of an individual walking motion. For each joint angle our motion model consists of a shape template and feature functions that describe the variation of that shape with the stride length. For the synthesis of a walking motion, the feature functions are evaluated for a desired stride length. Then the templates are adapted to match the computed features and used as progressions for the joint angles of the skeleton. We demonstrate our data driven approach using motion data captured from 12 individuals. We report on an experiment showing that the synthesized motions have a natural appearance and maintain the individual style.

1 Introduction

The realistic animation of virtual humans is still a challenging topic with many applications e.g. in computer games [1], simulations [2], ergonomic assessments [3] and robotics [4]. An application of particular current interest uses virtual reality (VR) technology for remote collaborations, meetings etc. [5]. The increasing use of such applications will certainly also raise the demand for high-quality self-representations in VR both with respect to the visual appearance and the motions of avatars. With regard to motions, an obvious desire is to display avatars with movements that are as similar as possible to one's own movements. Another requirement could be to use movements that differ from one's own movements only in very specific characteristics, for example to

disguise unwanted movement features. Having this application in mind, one can imagine that future VR users will not only use 3D scans of their body for the character creation but own motion capture data of key movements that are frequently performed in interactions. Walking motions play an essential role in this context because they occur frequently and convey a first impression about a person.

State-of-the-art methods of motion synthesis can be categorized as follows:

(1) stochastic or CNN-based methods for motion prediction allow the short-term continuation of a motion on a basis of a given sequence of frames but do not offer sufficient control for animation purposes.
(2) data-driven methods for neural style transfer allow to generate style adapted motions for character control but only refer to generic styles (e.g. “happy”) and not to individual motion characteristics.
(3) physics based methods allow the simulation of individual styles but require profound knowledge about the anatomical conditions of the individual. For a more detailed discussion see Section 2.

Our work is motivated by the apparent gap between the categories (2) and (3) characterized by the lack of a purely data-driven kinematic method that allows movements with individual style characteristics to be generated with high fidelity. In this paper, we develop such a method for the important case of gait motions based on a profound kinematic analysis. Our approach is based on the concept of an individual motion signature that for each joint angle consists of a shape template and feature functions that describe the variation of that shape. Templates can be defined to capture motion characteristics of individuals or groups of individuals, e.g. people of the same gender or age, and can be specified for different types of movements. However, in this paper only walking motions of individuals are considered.

We define template functions based on motion samples to describe whole body movements of individuals during walking. For our approach, we recorded motion data from 12 healthy subjects performing natural walking motions of different speeds. It is well known that for natural walking motions there is a linear relation between stride length and speed ([6] and [10]). Exploiting this relationship, we decided to study the motion variation in dependence of the stride length because this case appears more often in applications (e.g. motion planning).

As walking is a periodic movement, it suffices to define templates of strides. For each individual, we define templates for joint angle progressions by averaging the information contained in all recorded strides. This is done because averaging preserves shape features occurring in a significant large subset of instances while smoothing out minor variations. Now, the challenge is to transform the discrete representation of the template into a representation that has the flexibility to create progression functions for walking motions of different stride lengths and can be described with as few parameters as possible.

We investigated different possibilities to define C^1-continous shape-preserving approximations of templates using low degree polynomials. We obtained best results using a quintic spline T^c interpolating endpoints and extreme points of a given template T and determining the remaining degrees of freedom with a least squares approximation of the templates shape within each segment. Due to the monotonicity of its segments, T^c has exactly the same number of extreme points as the template T. Since endpoints and extreme points are the only (explicit) references to the shape of the template, we denote these as 'feature points'.

For each individual, we model 'feature functions' based on the recorded data which describe the variation of the feature points with the stride length. This involves the following steps. In the first step, we relate each recorded joint angle function to the corresponding motion template and in this way extract measurements of the template's feature points from the recordings. This results in measurement series both for the time values and angular values of the feature points in dependence of the stride length. In a second step, outliers are removed and best fitting quadratic polynomials to the measurements are computed which represent the feature functions.

To generate a progression function for a particular speed, we interpolate the features provided by the feature functions with an adaption of T^c that matches the feature points. This adaption is shape-preserving and can be easily obtained by exploiting the Bezier representation of the spline segments as follows. Let P_1, P_2 denote two consecutive extreme points of the template, Q_1, Q_2 the corresponding extreme points associated with a particular speed and A the 2d affine map that maps P_1 onto Q_1 and P_2 onto Q_2. If s denotes the segment of T^c joining P_1, P_2, then an adapted spline segment interpolating Q_1, Q_2 can be obtained by applying the affine map A to the Bezier points of s. Applying this transformation to each segment of T^c, one obtains a speed-adapted progression function with the same shape as the original template. This shape-preserving variation of the spline template is the core element of our motion generation on the level of joint angles.

In addition to the specification of joint angle progression, the motion generation requires to compute the global body position $\boldsymbol{p}_0^t$ of each pose that corresponds to the location of the pelvis node of the skeleton. To compute $\boldsymbol{p}_0^t$, we perform a forward kinematics of the skeleton, starting either from the position of the ankle joint or from the position of the toe joint of the supporting foot, i.e. for this computation we need to know the supporting foot and in which phase of movement it is at the moment. Our investigation of this problem shows that changes in the motion phase are indicated by certain minima of the joint angle functions. Since these minima are already contained in the feature vector, the times of the changes in the motion phases can easily be taken from this vector. This provides a particularly simple method for determining the current phase of motion.

To demonstrate the quality of the walking motions that are produced by our method, we conducted an extensive evaluation. A numerical comparison between 1680 recorded

strides and their syntheses showed that the mean joint displacement between two such motions varies between 0.5 and 3.6cm. Furthermore, the temporal coherence between similar poses is extremely high with an average deviation of 1.6 frames. While this evaluation shows a high mathematical similarity of the compared motions, it remains to evaluate how synthesized motions are perceived by observers. Therefore, we included a user study in our evaluation that consisted of two phases. In the first phase, users evaluated the perceived naturalness of the synthesized motions on an end-labeled bipolar 5-point Likert scale, ranging from artificial to natural. An equivalence test for the statistical analysis showed that participants do not perceive differences in naturalness between recorded and synthesized motions that are larger than $\Delta = .027$ with respect to a scale ranging from 0 to 1.

In the second phase, we tested the ability of the observer to recognize a reference motion among a set of three choices including the reference, its twin and a distractor motion. As a reference, either a recorded motion or a synthesized motion was shown. The twin of a recorded motion is defined as its synthesis and vice versa. For the evaluation, we only considered decisions for the reference R or the twin T. Under the assumption that the twin is complete identical to the reference, we would expect an equal number μ of decisions for the reference and the twin. Since NaturalWalk is a non-interpolating scheme that does not exactly reproduce the recordings, we cannot expect this result, but we did get a remarkably high 84% of the maximum possible. Taking into account, that humans have a high ability to recognize even slight differences between motions [39,40], we consider this result as a further confirmation of the high quality of the movements generated by our method.

The remainder of this paper is organized as follows. Following to the discussion of related work in section 2, preliminaries of the mathematics of motions are summarized in section 3. In section 4, we report on the acquisition and analysis of the data used as the basis of our approach. In section 5 shape templates and feature functions are introduced. The method of motion synthesis is described in section 6. Section 7 reports on the evaluation of the presented method.

2 Related Work

Due to their importance in applications, human gait has been widely studied in human movement sciences (e.g. [6], [7], [8], [9], [10]) or from the viewpoint of applications, e.g. [11]. Furthermore, numerous approaches for the synthesis of human walking motions have been proposed. Early kinematic approaches can be divided into two categories: procedural approaches and motion signal processing.

Based on experiment data, procedural approaches formulate equations on certain kinematic aspects of walking movement, e.g. the duration of the different phases within a stride. These equations are then used to compute key frames which are interpolated to

obtain the motion (e.g. [12, 13, 14]). These approaches laid the ground for more advanced methods, but do not allow modeling of individual walking styles due to the crude approximation of the motion kinematics.

Approaches of the second category are inspired by methods of signal processing and generate new motions from existing ones by applying global operations as motion blending, motion warping (e.g. [15, 16, 17]). Although suffering from the deficit to fulfil spatial constraints such as hand placements or avoidance of feet sliding, these proposals provided extremely useful tools for motion generation in general.

Applications of these techniques for walking motions include an online method that produces walking motions along a curved path [18]. However, in this work style aspects of the generated motion refer to the style of moving the mouse during the input of the path. In [19] motion warping is applied to generate motions specified by six style parameters: stride frequency, double support time, foot flexion, vertical hip excursion (apex and nadir) and pelvic swivel.

More recent work on motion synthesis fall into the categories of dynamics simulation and data-driven motion models. Approaches of the first category describe motion by a set of forces and torques acting upon a multi-body system. For an overview of such methods until 2010 see [20]. More recent methods are based on deep reinforcement learning which allows robust control policies to be learned by deep neural networks (e.g. [21]). Such methods even allow the simulation of pathological gait ([22]). However, for the reproduction of an individual walking style the anatomical conditions (e.g. mass distribution, body proportions, bone deformity, muscle conditions) must be known. This requirement excludes these methods for our vision of a simple and large-scale use of individual walking styles for avatar control.

Data-driven approaches build motion models from the analysis of large data sets of motion examples. Early approaches define motion spaces by interpolation of base motions [23] or by tessellation of the motion matrix containing the captured motions into its principle components [24]. These approaches have the inherent problems of lack of controllability and quality both resulting from the fact that the motion space is based on an abstract mathematical definition that has only an indirect relation to the kinematics of motion. Since then, more sophisticated motion models based on HMM or neural networks have been proposed, but they still struggle with the aforementioned problems.

In Computer Vision, neural networks have been used to learn motion models for the purposes of motion prediction and recognition. Prediction aims to compute the progression of a motion on the basis of a given sequence of poses. Both RNN- and CNN-based methods perform properly on short-term prediction (< 500ms) but tend to fail in the long-term prediction due to a collapse to a mean pose [25]. To reduce mode-collapsing, the use of stochastic models has been proposed [26]. Deep learning is also applied to improve the accuracy of gait recognition methods but relies on the availability of large data sets. To overcome the lack of training data, a framework to create synthetic

motion data has been presented [27]. It allows the variation of body shape, clothing, view and lighting conditions for the display of a virtual character but creates motion variations by sampling from a PCA motion space.

In Computer Graphics, neural network based motion models have been proposed to synthesize character movements based on high level parameters. In [28] the learned model is represented by the hidden units of a convolutional autoencoder that represents motion data in sparse components which can be combined to produce complex movements. However, the representation of motion has several disadvantages that make it necessary to apply a post process optimization to achieve a usable result. Even the rigidity of the skeleton has to be ensured by imposing bone length constraints in the optimization. More recent works focus on the problem of style transfer which is of interest in animation because it allows the reuse of recorded motion data in different contexts. However, the considered "styles" do not refer to specified kinematic properties of a movement but to a style represented by an example motion (frequently showing generic emotional states such as "happy" or "sad"; [29]). Typically, autoencoders are used that decompose an input motion into a content (i.e. the motion in a neutral style) and a style encoding during the learning phase. At runtime, the decoder combines the content of a presented motion with a previously learned style to produce the style-transferred output. In the contribution of [30] the style encoder also learns a common embedding from both 2D and 3D joint positions, which enables style extraction and transfer from videos. Output motions from the described methods have an appearance of a blend between the original motion and the example motion describing the style and currently there is no way to produce output with prescribed kinematic properties.

The quality evaluation of synthesized motions is a challenging issue that is left out in many works. Zhao et al. used [26] applied different metrics (measuring semantic consistency, diversity and distribution similarity) to show a statistical similarity between real data and data synthesized with their method. Such a similarity, however, can only be considered necessary for the quality of the synthesis. The statistical evaluation presented in this paper provides a stronger statement, because the deviations between real and synthesized movements are considered explicitly. However, any form of numerical evaluation can only give an indication of the quality of the data. The decisive proof of quality must relate to the perceived quality of the data and can therefore only be provided in a user study.

The only work we know of in which a user study was conducted to evaluate the quality of synthesized motions is by Tilmanne et al [31]. In their work the authors present an HMM-based motion model that allows to create style variations (in the sense of generic styles as "drunk") of a neutral walk encoded in the HMM. The study consisted of three phases two of which show some similarities to our work. In the first phase the naturalness of the motions was evaluated using a 3-point scale (real, undecided, synthetic). In this test 65% of the original walks and 50% of the synthetic walks were labeled as "real

walks". In the second test, the participants were asked to classify motions according to the displayed style. This task was successfully completed for 48% of the real motions and 36% of the synthesized motions. In the third test, participants were asked to rate the resemblance of simultaneously shown videos of real and synthesized motions in the same style using a 5-point scale. Here, participants achieved a global score of 3.15 which is close to maximum of 4. Unfortunately, for none of the tests the statistical significance of the results was investigated.

3 Preliminaries

3.1 Mathematics of motion

In computer animation, a kinematic human skeleton is an embedded graph (more precisely a tree) that describes the pose of a human body. The nodes $\boldsymbol{J}_i$ of this tree represent abstractions of joints in the human body and thus contain rotational information, while the edges of the tree correspond with bones. Since the graph is embedded in 3D space each joint $\boldsymbol{J}_i$ has a position $\boldsymbol{p}_i^t$ at the time *t* with respect to a global coordinate system. In this work, we use the skeleton of the Bio Vision Hierarchy (BVH) format [32] consisting of $nJ = 23$ joints. The root node $\boldsymbol{J}_0$ of this hierarchy is a virtual joint in the pelvis area. Fig. 1 shows this skeleton in a rest pose that is used as a reference for the pose description. Let $\boldsymbol{p}_i^{rest}$ $(i = 0, \ldots, nJ)$ denote the joint positions in the rest pose. Any joint $\boldsymbol{J}_i$ $(i = 0, \ldots, nJ)$ has an associated bone $\boldsymbol{T}_i$ connecting it to its unique father $\boldsymbol{J}_{f(i)}$, i.e. $\boldsymbol{T}_i = \boldsymbol{p}_i^{rest} - \boldsymbol{p}_{f(i)}^{rest}$. The lengths $|\boldsymbol{T}_i|$ of these bones specify the anthropometry of a particular body. Based on these definitions, an arbitrary pose $\boldsymbol{P}^t$ at the time *t* can be defined as

$$\boldsymbol{P}^t = (\boldsymbol{p}_0^t,\ \boldsymbol{q}_0^t, \boldsymbol{q}_1^t, \boldsymbol{q}_2^t, \ldots, \boldsymbol{q}_{nJ}^t),$$

where $\boldsymbol{p}_0^t$ specifies the global position of the skeleton, $\boldsymbol{q}_0^t$ its global orientation and $\boldsymbol{q}_i^t$ $(i = 0, \ldots, nJ)$ is the 3d rotation that brings the bone $\boldsymbol{T}_i$ into the desired orientation with respect to the local frame of $\boldsymbol{J}_{f(i)}$, i.e.

$$\boldsymbol{p}_i^t = \boldsymbol{q}_i^t \,.\, \boldsymbol{T}_i + \boldsymbol{p}_{f(i)}^t.$$

A motion *M* is a sequence of nP poses: $M = (\boldsymbol{P}^t),\ t = 0, \ldots, nP$.

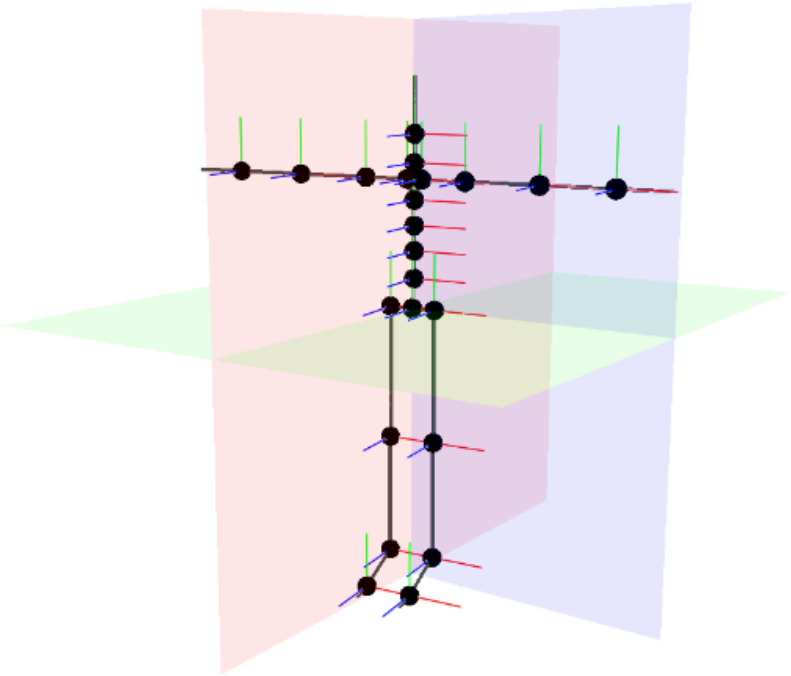

Figure 1: BVH skeleton with local coordinates systems at the joint positions: red for x-axis, blue for z-axis and green for y-axis. Anatomical body planes are shown in a similar shade as their normal vectors: red sagittal plane, blue frontal plane and green for transversal plane.

It is common, to express $\boldsymbol{q}_i^t$ in terms of elementary rotations about the axes of the local frame of $\boldsymbol{J}_{f(i)}$, i.e. one uses three angles $\alpha_i(t), \beta_i(t)$ and $\gamma_i(t)$ to specify the rotation at time value t about the x-, y- and z-axis of the local frame of $\boldsymbol{J}_{f(i)}$, respectively. The quantities α_i, β_i and γ_i are commonly referred to as joint angles (although not all of them correspond to a rotational degree of freedom of a human joint). The temporal variation of the joint angles in a motion constitutes discrete functions $\alpha_i(t), \beta_i(t)$ and $\gamma_i(t)$ $t = 0, \dots, nP$ which are called progression functions of the joint $\boldsymbol{J}_i$. We also say, that $\alpha_i(t)$ is the α-progression of joint $\boldsymbol{J}_i$ etc. We use the symbol θ for an arbitrary joint angle (i.e. without specifying the associated rotational axis). Note, that a motion is completely specified when $\boldsymbol{p}_0^t$, $\boldsymbol{q}_0^t$ and the progression functions of all $3nJ$ joint angles θ are given.

3.2 Walking motions

Human walking is a periodic form of locomotion where at any time there is at least one foot in contact with the ground [6]. In this work, we consider only straight walking on a flat horizontal ground. Considering one foot, a period consists of two phases: (1) stance phase, where the considered foot is in contact with the ground and (2) swing phase, where

the considered foot has no contact with the ground [10]. The stance phase begins with a heel-strike pose, which is the first pose in which a swinging foot touches the ground, and ends with a toe-off pose, which is the last pose before the foot loses contact to the ground (see Fig. 2). Commonly, one defines a step as a portion of a walking motion that starts with a heel strike pose of one foot and ends with the next heel strike pose of the other foot. The stance phase of one foot lasts about 60% of time for one step, while the swing phase lasts about 40% of this. However, the faster the motion becomes, the shorter the contact phase and the longer the swing phase last (see [6], [10]). A stride is generally defined as a portion of a walking motion starting by a heel strike pose of one foot and ending by the next heel strike pose of the same foot. Within each stride, both feet touch the ground in a short interval. This double support phase occurs in between the swinging phases of the different legs. From a kinematic point of view, a walking motion is produced by flexion and extension of the lower extremities. Thus, it can be described as angular movements of hips, knees and feet in the sagittal plane [33] which correspond to the α-progressions of these joints.

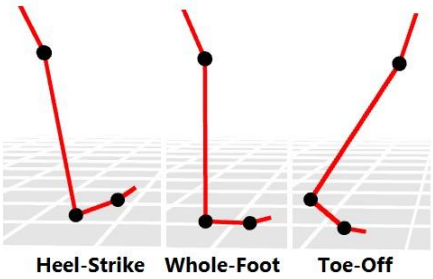

Figure 2: Subphases of the stance of the right foot.

4 Data acquisition and analysis

In this section, we report on the acquisition and analysis of the data used as the basis of our approach.

Assessing joint angle progressions of the lower extremities for different walking humans, it is apparent that these functions show many similarities and even the variations seem to follow distinguishable patterns. For illustration, Fig. 3 shows the α-progressions of the left knee for 3 different subjects walking with the same speed.

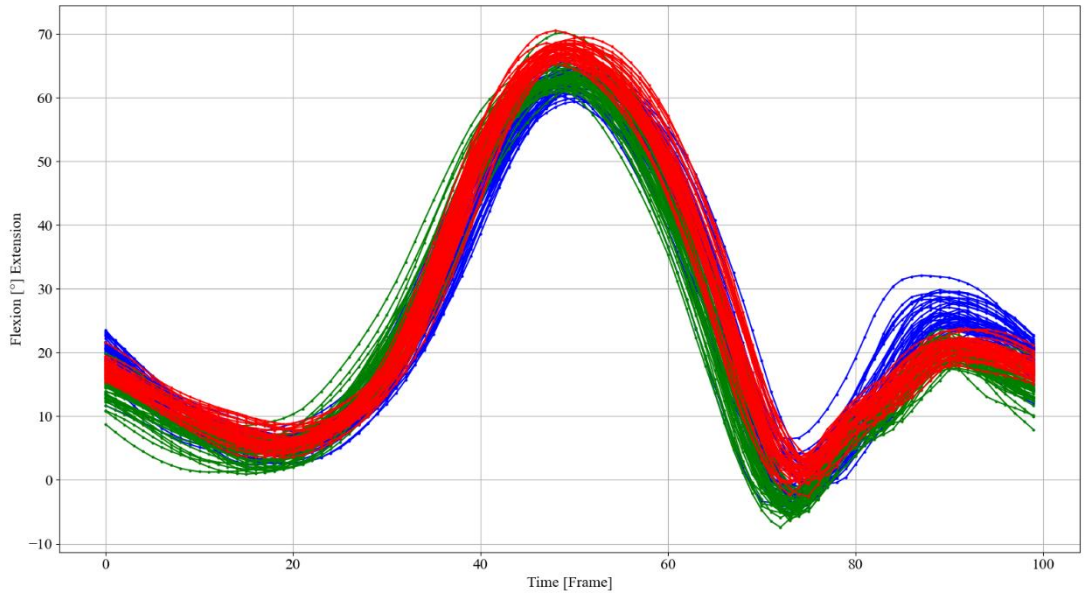

Figure 3: α-progressions of the left knee for 3 different subjects walking with the same speed. Progressions of the same color belong to the same subject.

To investigate the occurring shapes of such progressions, we recorded motion data for 12 healthy subjects without gait disorders (6 male and 6 female, aged between 26 and 54 years). A 17-wireless IMU-based sensor system (Xsens [34]) was used to collect 3-dimensional (3D) kinematic data at 60 Hz on a 15 m walkway. IMUs were attached to velcro bands respectively taps on 17 anatomical landmarks: head, sternum, shoulders, upper and lower arms, pelvis, hands, upper and lower legs and feet [35]. Overall, these markers represented 23 rigid segments and 13 joints in all three planes. The sensor-to-segment calibration was conducted in the recommended manner by Xsens. This procedure was done in two calibration steps Npose (10 s) + Walk (20 s) and was sufficient, when the quality indication was on a “good” respectively “acceptable” level. These static and dynamic calibration trials were followed by the kinematic data capturing while subjects walked at three different walking speeds (1.0 m/s, 1.5 m/s and 2.0 m/s) on a 15m walkway wearing their own shoes. For each speed, 20 sequences were recorded to create sufficient data to study intrapersonal variations, i.e. in total, we recorded about 720 motion clips consists of approximately 4 - 8 consecutive strides. Subjects were instructed to start walking from a standing still position (3 s), walk 15 m at the predetermined speed and stand still for 3 s at the end. As the recording was stopped, subjects went back to the starting position. The mean speed was controlled by two photoelectric barriers (Alge-Timing, Lustenau, Austria) – a deviation of ten percent of the predetermined speed was allowed, otherwise, the trial was invalid.

Since we were interested to specify basic shapes of the joint angle progressions for a complete gait cycle, we segmented the recorded data into strides. However, we found the

common stride definition (that is based on the notion of a heel-strike pose) not to be useful for the automatic processing of our data. Due to inaccuracies in the motion data, it is difficult to precisely recognize the first contact of a swinging foot with the ground. Consequently, a segmentation of the motion data based on this criterion leads to strides that start (or end) with clearly differing poses, which means that the segments represent different sections of the walking motion. To overcome this problem, we based our segmentation on the requirement that start and end poses (regarding the lower extremities) of all segments are as similar as possible. More precisely, we use the following stride definition for our segmentation:

A stride is a portion of a walking motion that begins with a pose where the left foot touches the ground and the flexions of the left and right hip in the sagittal plane are approximately equal and ends with the next pose that fulfils these conditions.

To segment a given motion clip $M = (\boldsymbol{P}^t),\ t = 0, \dots, nP$ into strides according to this definition, we first compute the local minima of the discrete function $f(t) = \left|\alpha_{lhip}(t) - \alpha_{rhip}(t)\right|,\ t = 0, \dots nP$. The minimum points $\Omega = (t_l), l = 0, \dots, nM$ of f are used to segment the motion into nM strides S_l defined as a sub-sequence of poses of M, i.e., $S_l = (\boldsymbol{P}^t),\quad t \in [t_l, t_{l+1}]$ for $l = 0, \dots, nM - 1$. Segments containing less than 20 poses are discarded as they do not correspond to strides but variations of a standing pose.

For each subject and joint angle, the segmentation creates at least 240 progressions (20 sequences per speed, at least 4 strides per sequence). For the shape analysis of these functions, we compared the number of their extremal points (max or min). Table 1 shows the results of the joint angles α_{lhip}, α_{lknee} and α_{lfoot}. It is apparent that for each of these joint angles a large majority of progressions contains the same number of extremal points (the majority vote). Furthermore, for most subjects and joints a second cluster exists that contains progressions with the same number of extremal points different from the majority vote. For example, there are 253 instances of the α-progression of the left hip of subject 2. 192 instances have 4 extrema while 57 have only 2 ones. Only 4 instances have a different count.

Joint Angle	$\boldsymbol{\alpha_{lhip}(t)}$			$\boldsymbol{\alpha_{lknee}(t)}$			$\boldsymbol{\alpha_{lfoot}(t)}$		
Features Count	4	2	other	4	2	other	4	6	other
Subject 1	235	6	0	241	0	0	180	61	0
Subject 2	192	57	4	214	37	2	238	12	3
Subject 3	188	57	4	248	1	0	204	42	3
Subject 4	197	35	8	240	0	0	231	7	2
Subject 5	253	1	8	262	0	0	227	32	3

Subject 6	289	45	19	319	34	0	346	5	2
Subject 7	238	68	0	304	1	1	304	1	1
Subject 8	257	16	5	244	33	1	199	76	3
Subject 9	283	72	0	349	6	0	352	3	0
Subject 10	228	65	4	293	4	0	289	8	0
Subject 11	423	14	2	341	96	2	435	2	2
Subject 12	353	88	0	322	117	2	438	1	2

Table 1: Numbers of extremal points in joint angle progressions for all subjects considered in this work.

Fig. 4 shows the shapes of two α_{lhip} progressions chosen from the two major clusters of subject 2. Obviously, the shape of the progression with 2 extrema can be considered a variation of the more general shape with 4 extrema.

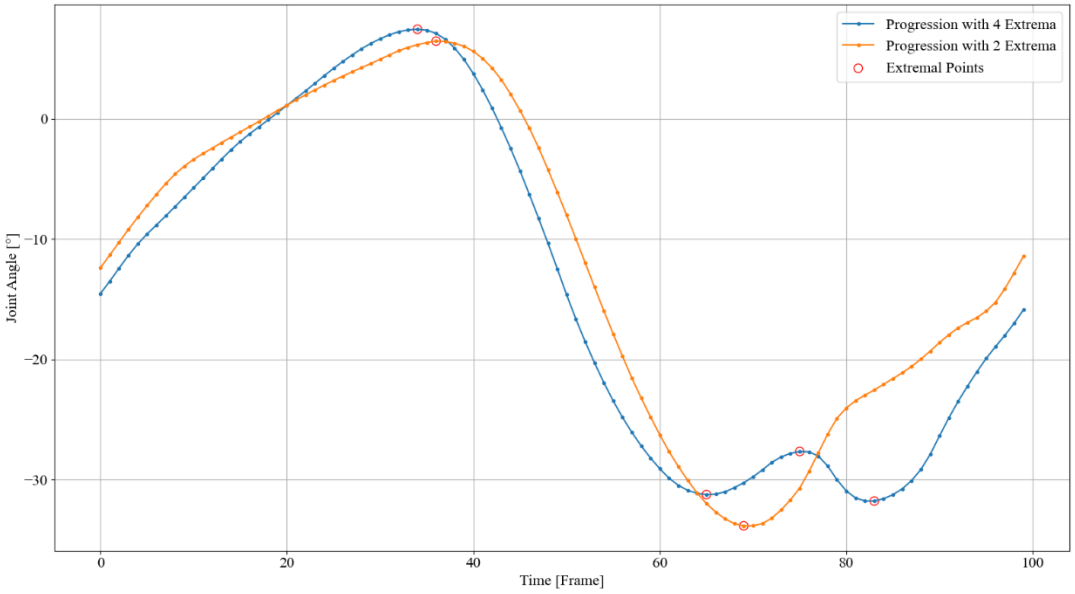

Figure 4: Two different shapes of the α-progression of the left hip of subject 2. The orange shape can be considered a special case of the blue shape in which two extrema are missing.

We analyzed whether the variations in the number of extrema depends on the stride length, but at least an obvious relation could not be established. Fig. 5 shows that variations in the number of extrema occur over the whole range of stride lengths.

Based on these findings, we decided to define one template shape for each joint angle which is used to create walking motions of different speeds in the range of the recorded speeds.

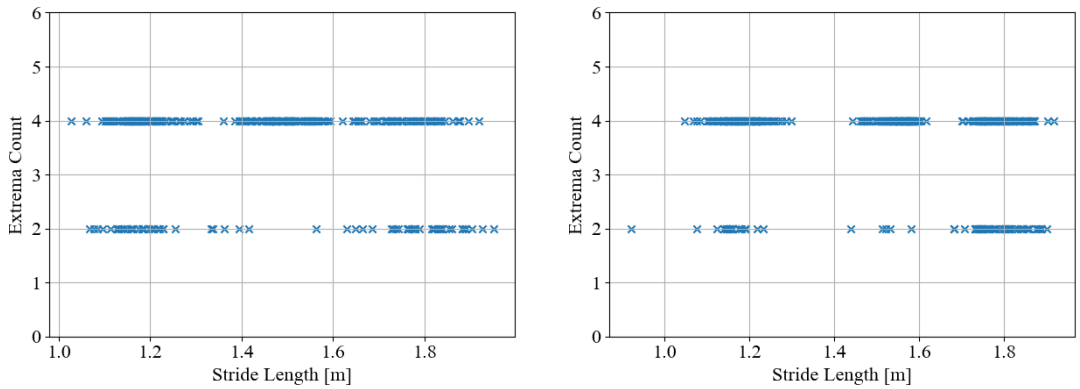

Figure 5: Number of extrema per stride length for α-progressions of left hip in two different subjects: subject 7 on the left and subject 9 on the right.

5. Shape templates and feature functions

In this section, we define shape templates for the progression functions of joint angles of a walking human. Depending on the application, shape templates can be used to represent motion characteristics of individuals or groups of individuals. These cases differ in the choice of segmented strides used in the template definition while the process to create the templates remains the same. The template definition is based on shape describing feature points extracted from the considered data base. A model is presented that describes the variation of these feature points as a function of the stride length. This allows to create progression functions realizing walking motions of a particular stride length by adapting the template functions.

5.1 Definition of template functions

We will define features of walking motions based on a set of segmented strides. For a correct assignment of different measurements to a particular feature, it is necessary to normalize the duration of the motion segments. This is done by computing a continuous representation of each motion segment and resampling these representations with a constant number of 101 frames. For the continuous representation we employ linear interpolation.

The template T_θ of a joint angle θ is a (discrete) function that serves as a shape model for the progression of θ, i.e. it shows the main shape characteristics of the progressions that appear for varying stride lengths but abstracts from minor shape variations. Since an

averaging procedure has the desired effect of preserving shape features occurring in a relevant subset of instances and smoothing out minor variations, we define the template T_θ as follows:

$$T_\theta(t) = \frac{1}{nS}\sum_{k=1}^{nS} \theta_k(t) \quad \text{for } t = 0, \dots, 100$$

where nS is the number of strides considered and $\theta_k(t)$ is the progression of joint angle θ in the stride k. Ends points and extremal points constitute the feature vector F of the template T_θ, i.e.

$$F(T_\theta) = \left((t_1, T_\theta(t_1)), \dots, (t_m, T_\theta(t_m)) \right), \tag{1}$$

where $m-2$ is the number of extremal points of T_θ, i.e., $t_1 = 0$ and $t_m = 100$ and $\left(t_j, T_\theta(t_j)\right)$ with $t_j \in \{2, \dots, 99\}$ is an extremal point for $j = 1, \dots, m-1$. The number $m(\theta)$ is called the feature count of T_θ. The number $r(\theta) \in \{1, 2\}$ specifies whether the sequence of extremal points of T_θ begins with maximum (r=1) or minimum (r=2). The pair $(m(\theta), r(\theta))$ is called feature signature of T_θ. Fig. 6 shows examples of shape templates with highlighted features.

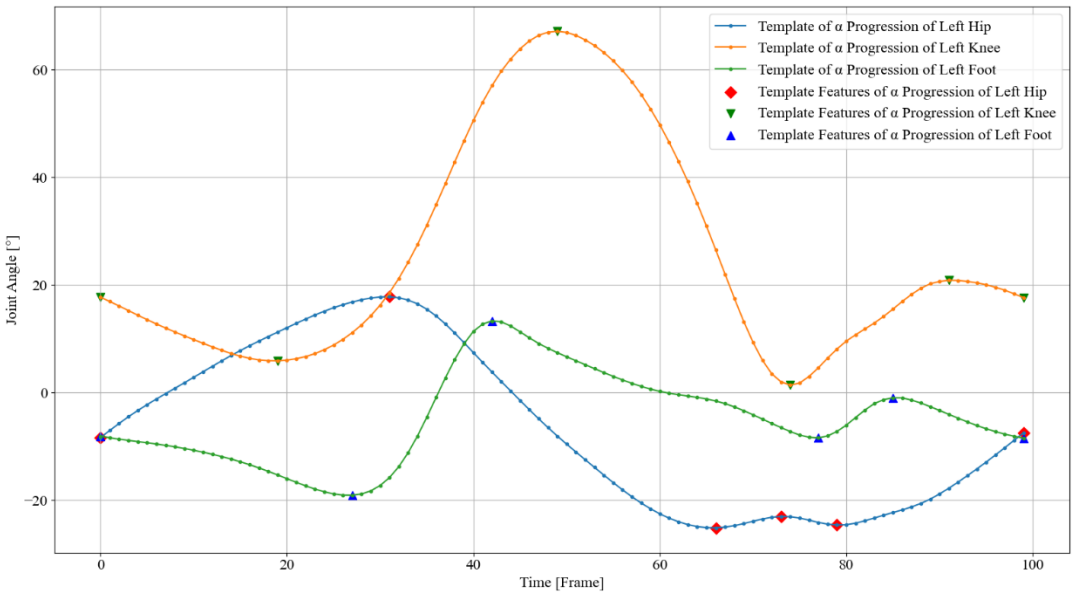

Figure 6: Shape templates of α_{lhip} (blue), α_{lknee} (orange) and α_{lfoot} (green) of subject 9 with feature points.

In the motion generation, we will compute progression functions for given instances of feature vectors. For this, we employ a continuous representation of the template function that is adapted to the particular values of the feature vector. In the next section, we begin by computing continuous representations of the templates.

5.2 Continuous representation of template functions

Let T_θ be the template of a joint angle θ and $F(\theta)$ the associated feature vector according to (1). We compute a C^1-continuous function T_θ^c: $[0,100] \rightarrow \boldsymbol{R}$ such that

1. $T_\theta^c(t_j) = T_\theta(t_j)$ for $j = 1, \dots, m$.
2. $t_2, \dots, t_{m-1}$ are the only extremal points of T_θ^c in $]0,100[$. (2)

Condition 1 of the above problem can be easily fulfilled by constructing a low degree polynomial Hermite spline, i.e. by fitting polynomials into each segment $[t_j, t_{j+1}], j = 1, \dots m-1$. We used different polynomial degrees to inspect how well these functions approximate the shapes of the template segments. For simplicity, we will restrict the following discussion to an interior segment $[t_j, t_{j+1}], \mathrm{j} = 2, \dots m-2$.

Interpolating the conditions $T_\theta^c(t_j) = T_\theta(t_j), T_\theta^{c'}(t_j) = 0$, $T_\theta^c(t_{j+1}) = T_\theta(t_{j+1})$ and $T_\theta^{c'}(t_{j+1}) = 0$ with cubic polynomials yields a C^1 cubic spline. However, some segments of this spline tend to be far off the corresponding segment of the template (see Fig. 7). Using quartic polynomials allows to interpolate an inflection point within the segment. However, as shown in Fig. 7 complicated segments tend to have several inflections points and the interpolation of one of them improves the interpolation only slightly. Interpolating all inflection points would greatly improve the approximation but such an approach comes with the following disadvantage. Using inflection points as features would require to develop a model that describes how their numbers and locations change with the stride length (see section 5.4 for such a model of extremal points). However, spurious inflection points can easily occur in noisy curve data and the inclusion of inflections as features would thus introduce inaccuracies into our method. We also used a quintic Hermite approach where in addition to the cubic case also second derivatives at the knots (which are estimated from the template functions) are interpolated. Fig. 7 shows that this another approach that improves the approximation only slightly.

Based on these experiments, we decided to follow an approach that combines interpolation and approximation. More precisely, we approximate each segment of T_θ with a quintic polynomial p_j that interpolates positions and derivatives at the end points of the interval $[t_j, t_{j+1}]$ and determine the remaining two degrees of freedom with a least square approximation of the points of T_θ within the interval.

For this, we use a Bezier representation of p_j on the interval $[t_j, t_{j+1}]$, i.e.

$$p_j(t) = b_{0,j}(t)B_{i,j}(t) + \cdots + b_{5,j}(t)B_{5,j}(t),\ t \in [t_j, t_{j+1}],\ j = 1, \dots m-1$$

where $B_{k,j}(t), k = 0, \dots 5$ denote the Bernstein polynomials of degree 5 on $[t_j, t_{j+1}]$. Here, we set

$$b_{0,\mathrm{j}} = b_{1,\mathrm{j}} = T_\theta(t_j) \text{ and } b_{4,\mathrm{j}} = b_{5,\mathrm{j}} = T_\theta(t_{j+1}),$$

which guarantees that p_j interpolates $T_\theta(t_j)$ and $T_\theta(t_{j+1})$ and has vanishing derivatives at both end points. The remaining Bezier coefficients $b_{2,j}, b_{3,j}$ are determined by solving the least squares problem

$$h(b_{2,j}, b_{3,j}) = \sum_{i=0}^{n} \left| T_\theta(t_j + u_i) - p_j(t_j + u_i) \right|^2 \to \min$$

with $u_i = \frac{t_{j+1} - t_j}{n} * i$. The solution of this minimization task is obtained by solving the following system of equations using the Moore-Penrose inverse:

$$\begin{bmatrix} B_{2,j}(u_0) & B_{3,j}(u_0) \\ B_{2,j}(u_1) & B_{3,j}(u_1) \\ \dots & \dots \\ B_{2,j}(u_n) & B_{3,j}(u_n) \end{bmatrix} \begin{bmatrix} b_{2,j} \\ b_{3,j} \end{bmatrix} = \begin{bmatrix} T_\theta(t_j + u_0) - p_j(t_j + u_0) \\ T_\theta(t_j + u_1) - p_j(t_j + u_1) \\ \dots \\ T_\theta(t_j + u_n) - p_j(t_j + u_n) \end{bmatrix}$$

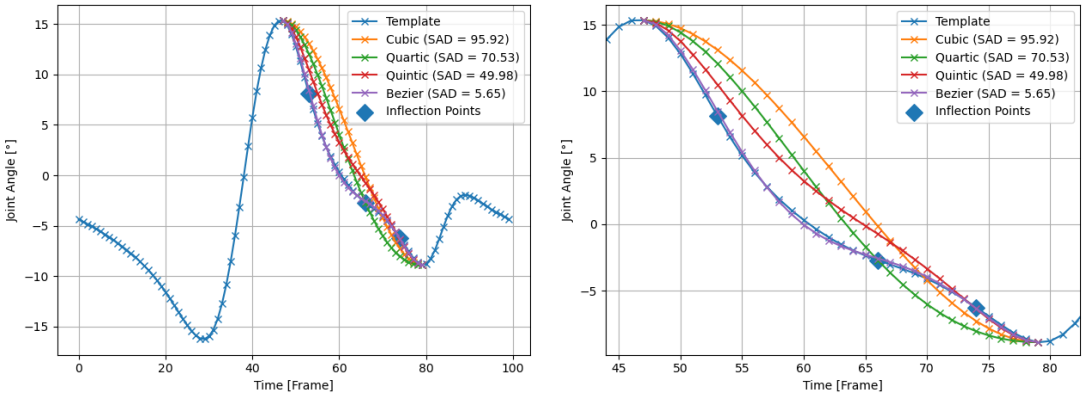

Figure 7: Left: different approximation methods applied to a segment of the template progression of $\boldsymbol{\alpha_{lfoot}}$. Right: zoomed view.

Since the approximation is computed with respect to a monotone subsequence of T_θ, each polynomial p_j is itself monotone which means that T_θ^c has no further extremal points in $]0,100[$ besides $t_2, \dots, t_{m-1}$. Thus, the function $T_\theta^c: [0,100] \to \boldsymbol{R}$ obtained in this way is a quintic C^1spline satisfying (2).

5.3 Building the feature functions

In the motion generation, we will synthesize joint angle progressions for any given stride length l. Thus, for any joint angle θ we need a function that describes the variation of the feature points of T_θ with the stride length. These feature functions are obtained by interpolation of feature data extracted from the considered set of segmented strides.

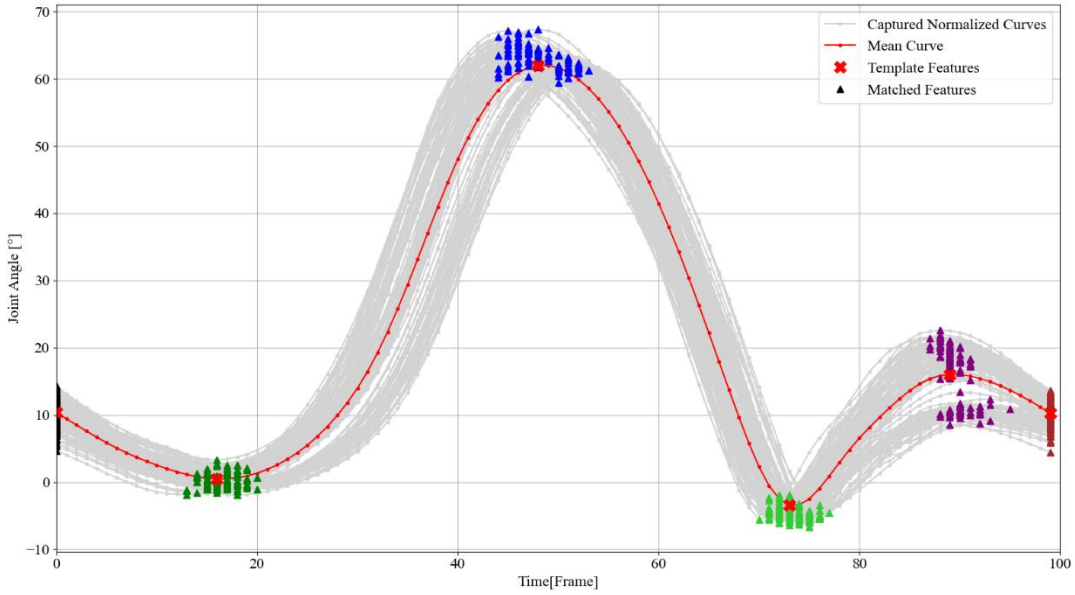

Figure 8: Template (in red) computed from a set of $\boldsymbol{\alpha_{lknee}}$ progressions (in grey). Feature points of progressions marked with triangles.

First, we will consider the extraction of the feature data. Fig 8 illustrates the problem for a set of α_{lknee} progressions of subject 9 (shown in grey). The template of α_{lknee} is shown in red with crosses indicating the feature points. The locations of the feature points in the individual progressions are marked with colored triangles. Each of these triangles corresponds to a measurement of a feature point and we want to extract these measurements.

Let Σ denote a set of nS segmented and normalized strides and Σ_θ be the set of progressions of a joint angle θ contained in Σ. For each progression $s \in \Sigma_\theta$ we compute its feature signature and recorded length of that stride from which it has been extracted. This is done because we want to create a data base of feature points in dependence of the recorded stride length. For this, we need to find a correspondence between the extremal points of each $s \in \Sigma_\theta$ and those of the template T_θ because only the latter are considered in our model. Since we have sufficient data, we simplify this fitting process by only considering those progressions that have the same feature signature as T_θ. Then, we simply associate the extremal points of the two functions according to the order of appearance.

For each joint angle θ, the described procedure extracts recorded feature data from the set of strides in the format

$$Z_j^k(\theta) = \{(l^k, t_j^k, y_j^k)\} \;\; j = 1,..,m \;\text{ and }\; k = 1,\dots,nR(\theta)$$

where (t_j^k, y_j^k) is the location of the j-th feature point of T_θ in the k-th stride and l^k is the length of that stride. Note, that the number of strides $nR(\theta)$ from which data can be extracted varies with θ.

We approximate the extracted feature data separately in the time and angle domain, i.e., we consider the sets

$$X_j^k(\theta) = \{(l^k, t_j^k)\} \text{ and } Y_j^k(\theta) = \{(l^k, y_j^k)\} \; j = 1,..,m \text{ and } k = 1,\dots,nR(\theta).$$

We use the DBSCAN algorithm [36] to remove outliers in these data sets and then compute best fitting quadratic polynomials. In this way, we obtain a function $\chi_{\theta,j}(\ell)$ approximately describing the variation of the j^{th} feature of T_θ in the time domain and a function $\psi_{\theta,j}(\ell)$ to denote its variation in the angular domain. Fig. 9 and Fig. 10 show the data sets $X_3^k(\alpha_{lknee})$ and $Y_3^k(\alpha_{lknee})$ for the third feature point (corresponding to the absolute maximum) of α_{lknee} of subject 9 and the quadratic fits to these sets.

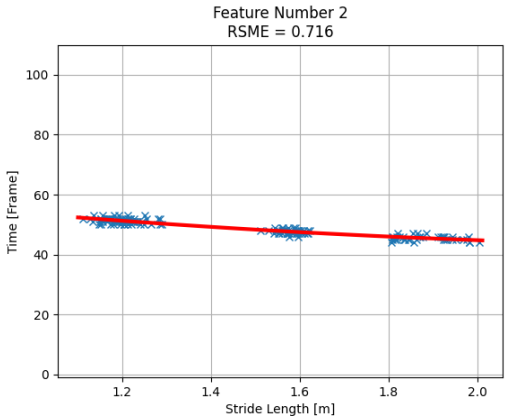

Figure 9: Variation of time values of the third feature of subject 9.

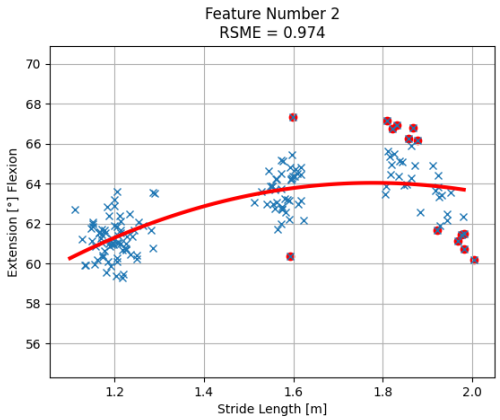

Figure 10: Variation of angular values of the third feature of subject 9 (outlier marked in red).

6 Motion Generation

In this section, we describe how a walking motion of a particular stride length l and stride duration d is computed for a skeleton of specified anthropometry and an associated set of feature functions $\chi_{\theta,j}$ and $\psi_{\theta,j}, j = 1,..,m$, given for each joint angle θ. The motion generation proceeds in the following steps.

1. For each joint angle θ, the evaluation of the feature functions at stride length l provides a feature vector

$F(\theta) = ((s_1, e_1), \dots, (s_m, e_m))$ with $s_j := \chi_{\theta,j}(l)$ and $e_j := \psi_{\theta,j}(l), j = 1,..,m.$

2. For each joint angle θ, a progression function T_θ^F is generated by adapting the template function T_θ^c to match the feature vector $F(\theta)$. This step is described in section 6.1.
3. Each pose $\boldsymbol{P}^t$ of the motion is computed by evaluating the progression functions T_θ^F for all joint angle θ at t and using the obtained values to specify the orientations of the joints of the skeleton. This step is described in section 6.2.

6.1 Adaption of template functions

Let θ be any joint angle and (m, r) the feature signature of its template T_θ^c. A vector $F = ((s_1, e_1), \dots, (s_m, e_m))$ is an admissible feature vector corresponding to the feature signature (m, r), if

1. $s_1 < s_2 < \cdots < s_m$ and $s_1 = 0$ and $s_m = 100$
2. for $r = 1$ and m even: $e_1 \le e_2, e_2 \ge e_3, e_3 \le e_4, \dots, e_{m-2} \ge e_{m-1}, e_{m-1} \le e_m$
3. for $r = 1$ and m odd: $e_1 \le e_2, e_2 \ge e_3, e_3 \le e_4, \dots, e_{m-2} \le e_{m-1}, e_{m-1} \ge e_m$
4. for $r = 2$ and m even: $e_1 \ge e_2, e_2 \le e_3, e_3 \ge e_4, \dots, e_{m-2} \le e_{m-1}, e_{m-1} \ge e_m$
5. for $r = 2$ and m odd: $e_1 \ge e_2, e_2 \le e_3, e_3 \ge e_4, \dots, e_{m-2} \ge e_{m-1}, e_{m-1} \le e_m$.

For a given admissible feature vector F corresponding to the feature signature (m, r) we compute a function $T_\theta^F: [0,100] \to \boldsymbol{R}$ such that

1. $T_\theta^F(s_j) = e_j$ for $j = 1, \dots, \mathrm{m}$.
2. $s_2, \dots, s_{m-1}$ are the only extremal points of T_θ^F in $]0,100[$.

We compute T_θ^F by scaling the segments p_j of T_θ^c such that condition 1 is satisfied. In other words, T_θ^F consists of m quintic polynomials q_j that are scaled versions of the polynomials p_j of T_θ^c, i.e.

$$T_\theta^F(\mathrm{s}) = q_j(s) \quad \text{for } s \in [s_j, s_{j+1}], \quad j = 0, \dots, m-1$$

$$q_j(s) = b_{0,j}^* B_{0,j}^*(s) + \cdots + b_{5,j}^* B_{5,j}^*(s), \; s \in [s_j, s_{j+1}]$$

with Bezier points $b_{k,j}^* = \alpha_j b_{k,j} + \beta_j, \quad k = 0, \dots 5$

where $\alpha_j = \frac{e_j - e_{j+1}}{b_{0,j} - b_{5,j}}, \qquad \beta_j = e_{j+1} - \frac{e_j - e_{j+1}}{b_{0,j} - b_{5,j}} b_{0,j}$

and $B_{k,j}^*$ denote the quintic Bernstein polynomials on $[s_j, s_{j+1}]$.

It can be easily verified, that this definition specifies T_θ^F as a C^1 quintic spline $T_\theta^F: [0,100] \to \boldsymbol{R}$ that confirms with 1. Condition 2 is also satisfied because scaling preserves monotonicity and thus no additional extremal points are introduced. For the motion generation, T_θ^F is discretized by sampling it at the integers $i = 0, \dots, 100$. Note, that in general q_j is represented by a different number of sample points than p_j because $[s_j, s_{j+1}]$ in general represents a different portion of $[0,100]$ than $[t_j, t_{j+1}]$.

6.2 Computing the poses

For the motion generation, we recall from section 3.1 that a motion is a sequence of poses of the form $\boldsymbol{P}^t = (\boldsymbol{p}_0^t, \boldsymbol{q}_0^t, \boldsymbol{q}_1^t, \boldsymbol{q}_2^t, \dots, \boldsymbol{q}_{nJ}^t)$ where $\boldsymbol{p}_0^t$ specifies the global position of the skeleton, $\boldsymbol{q}_0^t$ its global orientation and $\boldsymbol{q}_i^t$ is the 3d rotation that brings the bone $\boldsymbol{T}_i$ into the desired orientation specified by the angular values $\alpha_i(t), \beta_i(t)$ and $\gamma_i(t)$ $(i = 1, \dots, nJ)$. The anthropometry is specified by the lengths $|\boldsymbol{T}_i|$ of the bones in our skeleton.

As output of the second step in the motion generation described in section 6.1, we obtained a progression function T_θ^F for each joint angle θ of the skeleton. Evaluating these functions at t we obtain the rotational values $\alpha_i(t), \beta_i(t)$ and $\gamma_i(t)$ of each joint J_i, i.e. $\boldsymbol{q}_\mathrm{i}^t$ for $i = 1, \dots, nJ$.

The rotation $\boldsymbol{q}_0^t$ does not correspond to joint angles of the skeleton, but represents the global orientation of the skeleton in its root node. While $\boldsymbol{q}_0^0$ is an input to the motion generation (the initial orientation of the skeleton) the differences $\boldsymbol{o}^t = \boldsymbol{q}_0^t - \boldsymbol{q}_0^0$ represent an individual motion characteristic (the orientational fluctuation around the motion direction). Therefore, we process the elementary rotations α, β and γ associated with $\boldsymbol{o}^t$ in the same way as being done for the joint angles of the skeleton. This means that our motion model includes 3 template functions for the elementary rotations of $\boldsymbol{o}^t$ which are adapted to match the feature information provided by the feature functions. Then, $\boldsymbol{o}^t$ is computed by evaluating the 3 stride length adapted templates at t and $\boldsymbol{q}_0^t$ is obtained as $\boldsymbol{q}_0^t = \boldsymbol{o}^t + \boldsymbol{q}_0^0$.

The global body position $\boldsymbol{p}_0^t$ of each pose corresponds to the location of the pelvis node. To compute $\boldsymbol{p}_0^t$, we execute a forward kinematics of the skeleton starting from the position of the supporting foot. For this, it is necessary to know the supporting foot in each moment of the motion. More precisely, we need to know whether the supporting foot touches the ground with its sole or only with the toes. In the first case, the position of the ankle joint of the supporting foot is used as the starting point of the forward kinematics while in the second case, the position of the toe joint is used. Therefore, we have to divide the ground contact of the walking motion in four different phases: sole contact left, toe contact left, sole contact right and toe contact right. When executing the forward kinematics, we need to know to which of these phases the current pose belongs. This distinction is made upon the occurrence of particular configurations in the progression functions that indicate the change of the support phase.

The analysis of captured data showed that the change of a support phase can be determined based on the α-progressions of the toe, ankle and knee joints (Fig. 11). We determine the phase change from left (right) toe to right (left) foot based on the occurrence of a global minimum of the flexion of the left (right) toe. To determine the support change from the ankle joint to the toe joint of the same foot, different methods could be used, e.g. thresholding the flexion of the toe. However, in our experiments we found it more reliable to detect this support change using the occurrence of a global

minimum of the flexion of the knee of the other leg. Note, that as a result of 6.1 the progression functions are available to perform the computation of the minima.

By default, the left ankle is the support joint in the first pose for which the position of the root node is given as an input parameter $\boldsymbol{p}_0^0$ of the motion generation. To compute the position of the left ankle, we execute a forward kinematics of the skeleton starting from the pelvis position downwards to the positions of the joints of the left foot. For this, the joint angles synthesized for the first pose are used. In the next step, we execute the forward kinematics starting from the position of the left ankle up to the position of the root node. This time, using the joint angles synthesized for the second pose. In this way, we continue alternating between a forward kinematics starting at the pelvis or the position of the supporting joint. The switch of the supporting phase from one leg to the other takes place in a double support phase where both feet touch the ground. Therefore, it does not happen that the position of a foot is taken as the starting point of the computation when the foot is not grounded.

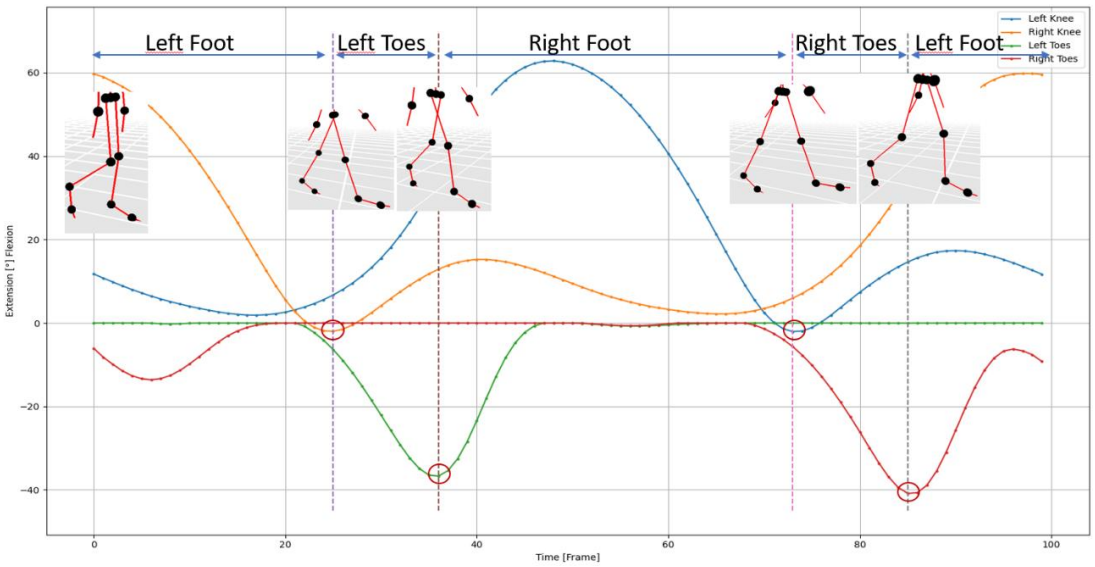

Figure 11: Changes of the support phases approximately coincide with minima of certain joint angle progressions.

Finally, we apply a smoothing filter to the root joint trajectory $\boldsymbol{p}_0^t$ to suppress variations due to the iterative nature of its computation.

The computed joint angle progressions still have the normalized duration of τ frames, which means that all generated motions have the same speed when they are visualized. To change the speed of the motion, we need to change the duration of the joint angle curves. If the desired walking speed is v, all joint angle progressions are resampled with $d = \omega \,.\, l \,/\, v$ frames, where l is the stride length and ω the visualization frequency. The walking speed is an optional input parameter. Thus, if no speed is given, the generator

will compute a suitable one using a linear function in stride length. This function is determined during the modeling process and it can vary slightly from one subject to another as we can see in Fig. 12.

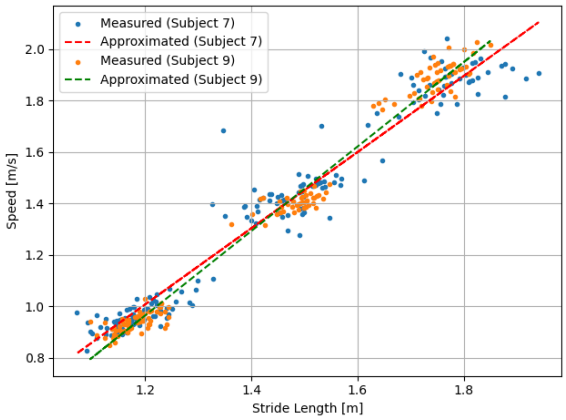

Figure 12: The relationship between stride length and walking speed in the recorded data of subject 7 and subject 9.

The output of the walking generator is a BVH file [32] consisting of a human skeletal model and a sequence of poses which represent a walking motion of the given stride length and the given walking speed.

7 Experimental Results

In this section, we evaluate the quality of the motions generated with our method. Since NaturalWalk represents a non-interpolatory scheme, a numerical evaluation is possible by comparing a recorded motion and a synthesized motion of same stride lengths and speeds. While such a comparison can be used to evaluate the mathematical similarity of two motions, it fails to answer the question how synthesized motions are perceived by observers. Therefore, we include a user study on the perceived naturalness of the synthesized motions in our evaluation.

7.1 Numerical Evaluation

The evaluation is based on the strides provided by the segmentation except those strides that are performed beginning or ending with a standing pose. The smallest number of such strides available for each of the 12 recorded subjects is 140. To include all subjects with the same weight, the evaluation is based on 140 * 12 = 1680 strides with lengths varying between 105cm and 198cm. For each of the 140 strides of a subject S, the stride length l and duration d was extracted and the motion generator was used to compute a

synthetic stride using the same parameters l and d, the feature functions of S and the anthropometry of S. In this way, a synthetic twin is computed for each of the 1680 motions.

Let M, M' denote a captured motion and its synthetic twin, respectively. For the evaluation, both strides are normalized to a duration of $\tau = 101$ frames. Let $\boldsymbol{p}_i^t$ and $\hat{\boldsymbol{p}}_i^t$ denote the 3D global positions of the joint J_i $(i = 1, \dots, nJ)$ at frame t of M and M' respectively. Then

$$MD(M, M') = \frac{1}{\tau * nJ} \sum_{t=1}^{\tau} \sum_{i=1}^{nJ} |\boldsymbol{p}_i^t - \hat{\boldsymbol{p}}_i^t|$$

provides the mean joint displacement, i.e., the average value over all frames and joints of the distances between corresponding joints of M and M' at the same frames, while

$$\mathrm{Max}(M, M') = \max_{t=1,\tau} (\max_{i=1,nJ} (|\boldsymbol{p}_i^t - \hat{\boldsymbol{p}}_i^t|))$$

provides the maximum joint displacement, i.e., the largest distance over all frames between corresponding joints of M and M' at the same frames. In addition to these metrics that measure the spatial similarity between two motions, we consider the following metric adapted from [37] to measure the temporal coherence between a captured and a synthesized motion:

$$\mathrm{C}(M, M') = \frac{1}{(\tau - 1) * nJ} \sum_{t=1}^{\tau-1} \sum_{i=1}^{nJ} |c_i^t|$$

$$c_i^t = \frac{\left|\boldsymbol{p}_i^t - \boldsymbol{p}_i^{t+1}\right| - \left|\hat{\boldsymbol{p}}_i^t - \hat{\boldsymbol{p}}_i^{t+1}\right|}{v_i^t}$$

Here, v_i^t is an approximation of the speed of the joint Ji at frame t in the captured stride M computed as average of $\left|\boldsymbol{p}_i^t - \boldsymbol{p}_i^{t+1}\right|$ over 6 frames around t. (If v_j^t=0, the average is computed over more frames.)

The temporal coherence metric C measures how similar the captured and synthesized motion evolve over time. Since v_i^t has the dimension [cm/frame], C has the dimension [frame], i.e. a value of $\mathrm{C}(M, M') = \mathrm{nF}$ means that a particular pose of M appears nF frames earlier or later in M'.

For each of these metrics, we computed the mean (MEAN), standard deviation (SD), minimum (Min) and maximum (Max) value for all 1680 evaluation pairs. The results listed in Table 2 show a very high similarity between synthesized and captured motions. In particular, from the fact that the mean joint displacement MD varies between 0.5 and 3.6cm we conclude that most of the corresponding poses in synthesized and captured motions are visually indistinguishable. Furthermore, the temporal coherence between similar poses is extremely high with an average deviation of 1.6 frames and a maximum deviation of 4.8 frames.

Larger displacements of certain joints are mostly due to a correction of the ground contact in the motion generation. Due to inaccuracies in the motion capturing, a foot in the stance phase is usually not strictly fixed to the ground in the recorded data. In contrast, the motion generator enforces the ground contact of the foot during the stance phase and thus introduces differences to the captured data. These differences accumulate because of the iterative procedure of motion generation.

	MD [cm]	Max [cm]	C [frame]
Mean	1.2	7.6	1.6
SD	0.4	2.0	0.7
Min	0.5	2.5	0.8
Max	3.6	11.9	4.8

Table 2: Spatial and temporal differences for 1680 motion pairs of 12 subjects.

7.2 User Study

We conducted a user study to evaluate the perceived naturalness of the motions synthesized with our method. Due to Corona restrictions, the experiment took place online with a total number of 72 participants. All participants were students enrolled in one of the programs of the Institute of Media Research at Chemnitz University of Technology. The age of the participants varied from 18 to 38 years (M = 23.4, SD = 4.3). Fifty-seven participants specified their gender as female, twelve as male. Three participants declined to answer ("I don't want to answer the question."). Before the start of the study, the participants were informed about the actual purpose of it.

The experiment consisted of two phases. For each phase we included a test run to familiarize the participants with the look of the virtual characters and their task in the experiment. Scores achieved in the test runs were not counted in the evaluation. We reserved the motions of two recorded subject (one male, one female) exclusively for the test runs. Therefore, the recorded motions of ten subjects (5 male and 5 female) were used in the experiments. To visualize the motions, we used two virtual characters which are clearly distinguishable as male or female gender (see Fig. 13). Motions captured from men (women) were visualized with the male (female) virtual character. For the experiments we used the recorded walking motions of medium speed. From the pool of these recorded motions, we selected one recording for each subject such that all chosen motions have similar walking speed and stride length. For each selected recording, we synthesized a walking motion using the average walking speed and average stride length of the recording as input to the motion generator. For each of the resulting 20 motions, we created a video clip (60fps) showing four complete strides of the virtual character from the same perspective.

Figure 13: Male and female model used to visualize the motions.

In the first phase of the experiment, we explicitly addressed the perceived naturalness of the motions. For this, we used an experimental design where each participant watches all 20 videos presented in a fully randomized order. Each video was played 5 times in succession and then rated by the participant on an end-labeled bipolar 5-point Likert scale, ranging from artificial to natural (see Fig. 14).

After scaling the scores to the interval [0,1] we obtained:

$$M_r = .506,\ SD_r = .292 \text{ and } M_s = .513,\ SD_s = .296,$$

where M_r and M_s denote the average score of all videos showing recorded movements and synthesized movements, respectively. Table 3 provides a more detailed view by listing the ratings of captured versus synthesized motions of each subject. In summary, based on our measurements we found (nearly) no difference in the average rating of naturalness of recorded and synthesized motions. Furthermore, differences in the ratings of subjects (interpersonal differences) are much larger than differences in the ratings of captured and synthesized motions (methodical differences). Thus, our measurements indicate that the presented motion generator is able to reproduce the perceived naturalness of a human walking motion. In the following, we will evaluate the statistical significance of this result.

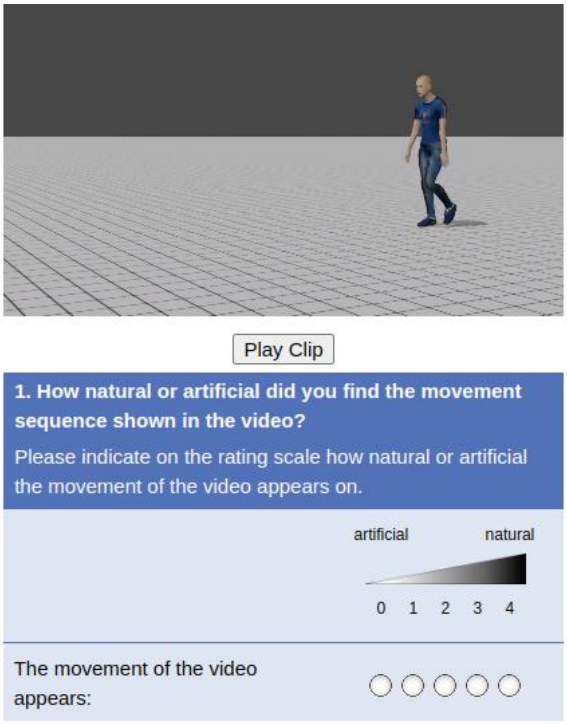

Figure 14: Screen shot of user interface of the online study. The texts shown here are translations of the German texts used in the experiment.

For the statistical analysis, average values of ratings of each participant in the two groups of real and artificial motions are considered. Fig. 15 shows box plots of this data. Since we have paired samples, we apply a dependent samples *t*-Test and use the common significance level $\alpha = .05$ to evaluate the statistical significance. For the *t*-Test we use the pair of hypotheses:

H_0: Participants do not perceive differences of naturalness between recorded and synthesized motions,

H_1: Participants do perceive differences of naturalness between recorded and synthesized motions.

Subject	Gender	Mean Cap	Mean Syn	SD Cap	SD Syn
1	m	.73	.66	.22	.26
3	m	.30	.24	.23	.24
4	m	.40	.38	.27	.29
7	m	.45	.44	.27	.26
9	m	.65	.73	.26	.23
2	f	.36	.44	.30	.26
5	f	.41	.47	.27	.26
6	f	.56	.58	.23	.27
10	f	.55	.50	.27	.24
12	f	.64	.68	.26	.28
Mean		.51	.51	.26	.26

Table 3: Mean values of naturalness per subject.

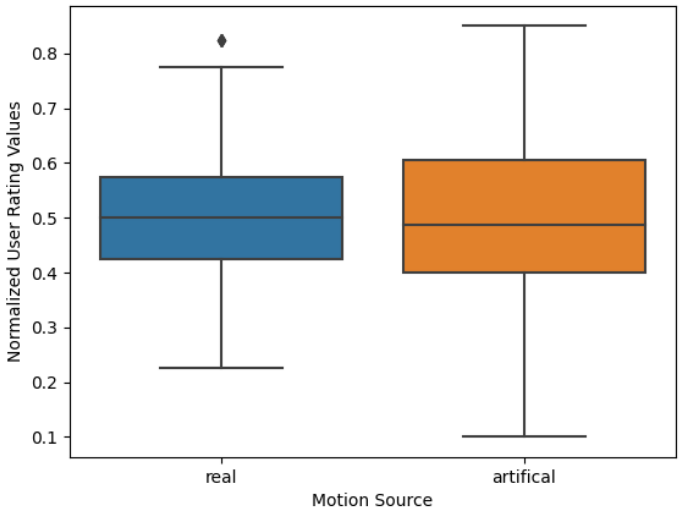

Fig. 15: box plots of the ratings of real and synthesized motions.

For the t-Test, we obtained t(71)= -.63, p= .53. Since the p value is much larger than the significance level α we find no statistical indication to reject the null hypothesis H_0 (in other words we find no indication to assume the alternative hypothesis H_1). An even stronger statement can be achieved using the two one-sided tests (TOST) procedure to test for equivalence ([38]) and reject the presence of the smallest effect size of interest (SESOI). For this, we formulate the hypotheses:
H_0: Participants do perceive differences in naturalness between recorded and synthesized motions that (in absolute value) are larger than Δ,
H_1: Participants do not perceive differences in naturalness between recorded and synthesized motions that (in absolute value) are larger than Δ.

For Δ = .027 we obtained t(71) = 1.694, p= .047 and thus conclude statistical equivalence for the groups of recorded and synthesized motions for this value of Δ.

On the basis of this evaluation, we conclude that there is no relevant difference in the perceived naturalness between recorded motions and those computed with our motion generator.

In the second phase of the experiment, we investigated the perceived quality of the motions generated with our method based on an assignment task. Three of the original 72 participants did not complete this part of the experiment what resulted in a gender distribution of 55 women, 11 men and 3 diverse. Each participant completed twenty assignment tasks. In each of these, one of the 20 videos was randomly chosen and presented to the participant as a reference R followed by three other videos (the selection set). The selection set for a particular video R consists of R itself, the twin T of R and a distractor D. Two videos A and B are called twins, if either B shows the synthesized version of the recorded motion displayed in A or A shows the synthesized version of the

recorded motion displayed in B. The distractor D of R is randomly chosen among all videos different from R and T but showing a motion of the same gender as R. The videos R, T and D of the selection set appear in one row below the cue video. The order in which these videos appear in this row is chosen randomly (see Fig. 16). Each video of the selection set was started at the same time after the reference video has been shown the first time. Then, all videos including the reference were played five times in a row simultaneously.

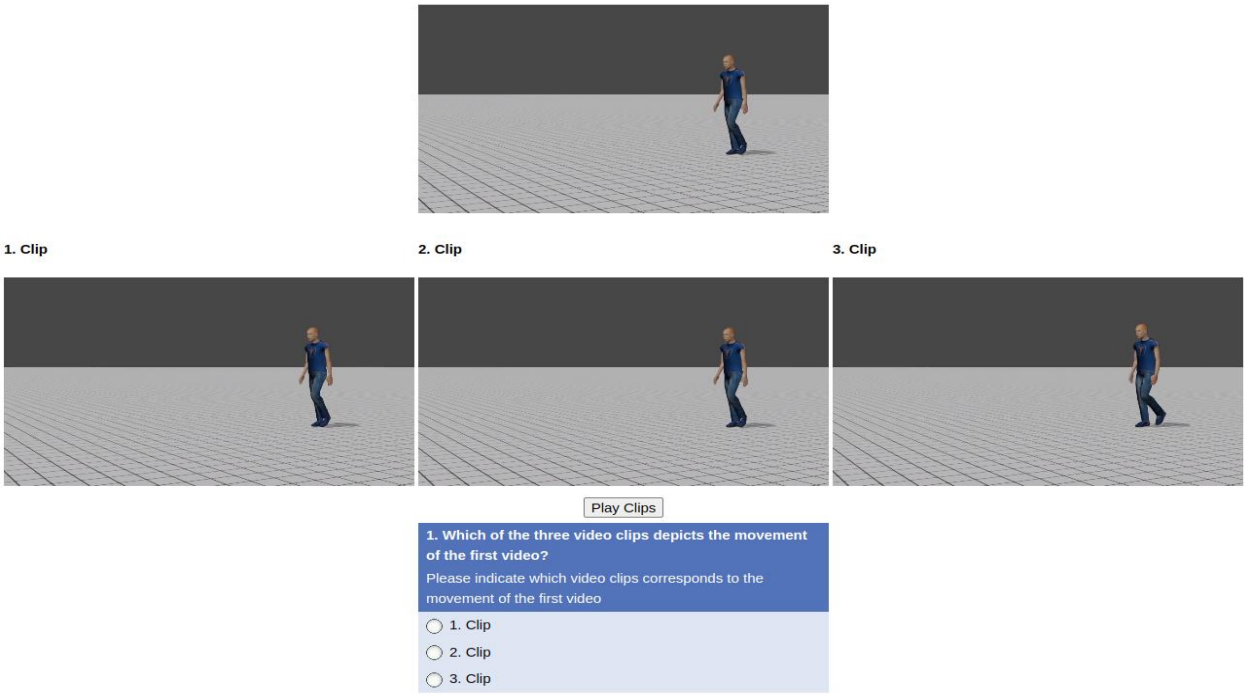

Figure 16: Screenshot of user interface in the second phase of the online study.

The rationale behind this experimental design is as follows: humans have a great ability to recognize even small differences in movements ([39,40]). Since the synthesis is not identical to the recording, observers should be able to pick the reference motion within the selection set. However, each false assignment to the twin of the reference is an indication that - despite the existing deviations - the synthesized motion looks so similar to the reference motion that it was possible to deceive the observer.

For an experiment with n assignments where an identical copy of the reference R instead of a synthesis is used for the twin T, the expected values for the assignments to R and T are equal: $\mu = (n - m)/2$, where m denotes the assignments to the distractor. Assignments to the distractor are not considered because these seem to indicate that the participant was unfocused or has a general problem with regard to judging movements.

Table 4 shows the results of the assignment experiment. In total, the participants completed 1380 assignments. In 92 cases the distractor was selected which corresponds to a share of 7 percent. The remaining 1288 assignments (Sum_RT) are divided into 749 assignments to the reference (Sum_Ref) and 539 assignments to the twin (Sum_Twin). A χ^2-test shows the expected statistical significance for the difference of Sum_Ref and Sum_Twin with a low effect size of 0.027. However, the strikingly high rate of 539 assignments to the twin corresponds to a value of 84% of the expected value of assignments μ in an experiment with R= T, further confirming the high quality of the movements generated by our method.

Sum_Ref	Sum_Twin	Sum_RT	Orig correct	Orig to Synth	Orig to Dis	Synth Correct	Synth to Orig	Synth to Dis	Sum
749	539	1288	382	263	45	367	276	47	1380

Table 4: Results of the assignment task.

Acknowledgement

This work is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 416228727 – SFB 1410.

References

1. Cai, S., Liu, Q. and Li, L.: A Case for Web-Based Interactive 3D Game Using Motion Capture Data. Edutainment 2010, LNCS 6249, pp. 540–548. 2010.

2. Magnenat-Thalmann N., Thalmann, D.: Handbook of Virtual Humans, Wiley, ISBN/EAN: 9780470023174. 2005.

3. Puthenveetil, S., C., Daphalapurkar, C., P., Zhu, W., Leu, M., C., Liu, X., F., Gilpin-Mcminn, J., K. and Snodgrass, S., D.: Computer-automated ergonomic analysis based on motion capture and assembly simulation. Virtual Reality 19, 119–128, DOI 10.1007/s10055-015-0261-9. 2015.

4. Zheng, X., Yi, P. and Zhang, Q.: Data-Driven Complex Motion Design for Humanoid Robots. Recent Developments in Intelligent Information and Database Systems, Studies in Computational Intelligence 642, DOI 10.1007/978-3-319-31277-4_31. 2016.

5. Collingwoode-Williams T, O'Shea Z, Gillies M. and Pan X.: The Impact of Self-Representation and Consistency in Collaborative Virtual Environments. Front. Virtual Real. 2:648601. DOI 10.3389/frvir.2021.64860, 2021.

6. Vaughan, C., Davis, B. and Connor, J.: Dynamics of Human Gait 2nd Edition. Howard Place, Western Cape 7450. 1999.

7. Borghese, N., A., Bianchi, L. and Lacquaniti, F.: Kinematic Determinants of Human Locomotion. Journal of Physiology, 494.3, pages 863–879. 1996.

8. Frigo, C., Crenna, P. and Jensen, L. M.: Moment-Angle Relationship at Lower Limb Joints during Human Walking at Different Velocities. Journal of Electromyography and Kinesiology, Vol. 6. No. 3, pages 177–190. 1996.

9. Sekiya, N. and Nagasaki, H.: Reproducibility of the Walking Patterns of Normal Young Adults: test-retest Reliability of the Walk Ratio (step-length: step-rate). Gait and Posture 7, pages 225–227. 1998.

10. Multon, F. and Anne-Hélène, O.: Biomechanics of Walking in Real World: Naturalness We Wish to Reach in Virtual Reality. Springer New York, Human Walking in Virtual Environments: Perception, Technology, and Applications, pages 55–77. 2013.

11. Du, H., Mannsb, M., Herrmann, E. and Fischer, K.: Joint Angle Data Representation for Data Driven Human Motion Synthesis. Proceedings of the 48th CIRP Conference on Manufacturing Systems, pages 746–75. 2015.

12. Bruderlin, A. and Calvert, T. W.: Goal-directed, dynamic animation of human walking. Proc. SIGGRAPH SIGGRAPH ’89, pages 233–242. 1989.

13. Boulic, R., Magnenat-Thalmann, N. and Thalmann, D.: A global human walking model with real-time kinematic personification. The Visual Computer 6. pages 344–358. 1990.

14. Ko, H. and Badler, N. I.: Straight Line Walking Animation Based on Kinematic Generalization that Preserves the Original Characteristics. Proceedings Graphics Interface ’93, pages 9–16. 1993.

15. Bruderlin, A., Williams, L.: Motion signal processing. In: SIGGRAPH 1995, pages 97–104. ACM, New York. 1995.

16. Witkin, A. and Popovic, Z.: Motion Warping. Proc. SIGGRAPH ’95, pages 105–108. 1995.

17. Wiley, D.J., Hahn, J.K.: Interpolation synthesis of articulated figure motion. IEEE Computer Graphics and Applications 17(6), pages 39-45. 1997.

18. Park, S.I., Shin, H.J., Shin, S.Y.: On-line locomotion generation based on motion blending. In: Proceedings of the 2002 ACM Eurographics Symposium on Computer Animation, SCA 2002, pages 105–111. 2002.

19. Curtis, S., Lin, M. and Manocha, D.: Walk This Way: A Lightweight, Data-driven Walking Synthesis Algorithm. Motion in Games. MIG 2011. Lecture Notes in Computer Science, vol. 7060, pages 400–411. 2011.

20. Xiang, Y., Arora, J., and Abdel-Malek, K.: Physics-based modeling and simulation of human walking: A review of optimization-based and other approaches. Structural

and Multidisciplinary Optimization. 42. pages 1-23. DOI 10.1007/s00158-010-0496-8. 2010.

21. Lee, S., Park, M., Lee, K. and Lee, J.: Scalable muscleactuated human simulation and control. ACM Transactions On Graphics (TOG) 38, 4, pages 1–13. 2019.

22. Park, J., Min, S., Chang, P. S., Lee, J., Park, M. and Lee, J.: Generative GaitNet. SIGGRAPH22 ACM. 2022.

23. Rose, C., Bodenheimer, B., Cohen, M.F.: Verbs and adverbs: Multidimensional motion interpolation. IEEE Computer Graphics and Applications 18, pages 32–40. 1998.

24. Boulic, R., Thalmann, D. and Glardon, P.: PCA-Based Walking Engine Using Motion Capture Data. In Proceedings of Computer Graphics International, Crete, 2004 pages 292-298, DOI 10.1109/CGI.2004.1309224. 2004.

25. Ye, Z., Wu, H. and Jia, J.: Human motion modeling with deep learning: A survey, AI Open,Vol. 3, pages 35-39, DOI 10.1016/j.aiopen.2021.12.002. 2022.

26. Zhao, R., Su, H., and Ji, Q.: Bayesian Adversarial Human Motion Synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 6225-6234, DOI 10.1109/CVPR42600.2020.00626. 2020.

27. Cheng, Y., Zhang, C., Huang, S., Wang, Z., Cheng, X. and Lin, J.: Synthesizing 3D Gait Data with Personalized Walking Style and Appearance. Applied Sciences, 13(4), 2084. 2023.

28. Holden, D., Saito, J., and Komura, T.: A deep learning framework for character motion synthesis and editing. ACM Transactions on Graphics (TOG), 35(4). 2016.

29. Park, S., Jang, D. K., and Lee, S. H.: Diverse motion stylization for multiple style domains via spatial-temporal graph-based generative model. Proc. Computer Graphics and Interactive Techniques, 4(3). 2021.

30. Aberman, K., Weng, Y., Lischinski, D., Cohen-Or, D., & Chen, B.: Unpaired motion style transfer from video to animation. ACM Transactions on Graphics 39, 4. 2020.

31. Tilmanne, J., Moinet, A. and Dutoit, T.: Stylistic Gait Synthesis Based on Hidden Markov Models. EURASIP Journal on Advances in Signal Processing. 2012.

32. Meredith, M. and Maddock, S.: Motion Capture File Formats Explained. Technical Report 211, Department of Computer Science, University of Sheffield, 36 pages, 2001.

33. Kronfeld, T., Fankhaenel, J. and Brunnett, G.: Representation of Motion Spaces Using Spline Functions and Fourier Series. Springer Berlin Heidelberg, Mathematical Methods for Curves and Surfaces, pages 265–282. 2014.

34. Xsens, http://www.movella.com/products/xsens. Date of Access May 8, 2023.

35. Roetenberg, D., Luinge, H. and Slycke, P.: Xsens MVN: Full 6DOF Human Motion Tracking Using Miniature Inertial Sensors. XSENS TECHNOLOGIES - VERSION APRIL 3. 2013.

36. Ester, M., Kriegel, H., Sander, J. and Xu, X.: A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining. 1996.

37. Vasa, L. and Brunnett, G.: Rate-distortion Optimized Compression of Motion Capture Data. Computer Graphics Forum, Volume 33, Issue 2, pages 283–292. 2014.

38. Lakens, D.: Equivalence Tests: A Practical Primer for t Tests, Correlations, and Meta-Analyses. Social Psychological and Personality Science, 8(4), pages 355–362. 2017.

39. Loula, F., Prasad, S., Harber, K. and Shiffrar, M.: Recognizing People from Their Movement. Journal of Experimental Psychology: Human Perception and Performance. Vol. 31, No. 1, pages 210–220. 2005.

40. Ambady, N., Hallahan, M., and Conner, B.: Accuracy of judgments of sexual orientation from thin slides of behavior. Journal of Personality and Social Psychology, 77, pages 538 –547. 1999.

Chemnitzer Informatik-Berichte

In der Reihe der Chemnitzer Informatik-Berichte sind folgende Berichte erschienen:

CSR-19-04 Johannes Götze, René Schmidt, Wolfram Hardt, Hardwarebeschleunigung von Matrixberechnungen auf Basis von GPU Verarbeitung, März 2019, Chemnitz

CSR-19-05 Vincent Kühn, Reda Harradi, Wolfram Hardt, Expert System for Adaptive Flight Missions, Juni 2019, Chemnitz

CSR-19-06 Samer Salamah, Guido Brunnett, Christian Mitschke, Tobias Hess, Syntheszing gait motions from spline-based progression functions of controlled shape, Juni 2019, Chemnitz

CSR-19-07 Martin Eisoldt, Carsten Neise, Andreas Müller, Analyse verschiedener Distanzmetriken zur Messung des Anonymisierungsgrades θ, Juni 2019, Chemnitz

CSR-19-08 André Langer, Valentin Siegert, Martin Gaedke, Informationsverwertung basierend auf qualitätsoptimierten semistrukturierten Datenbeständen im Wachstumskern "LEDS", Juli 2019

CSR-20-01 Danny Kowerko, Chemnitzer Linux-Tage 2019 - LocalizeIT Workshop, Januar 2020, Chemnitz

CSR-20-02 Robert Manthey, Tom Kretzschmar, Falk Schmidsberger, Hussein Hussein, René Erler, Tobias Schlosser, Frederik Beuth, Marcel Heinz, Thomas Kronfeld, Maximilian Eibl, Marc Ritter, Danny Kowerko, Schlussbericht zum InnoProfile-Transfer Begleitprojekt localizeI, Januar 2020, Chemnitz

CSR-20-03 Jörn Roth, Reda Harradi und Wolfram Hardt, Indoor Lokalisierung auf Basis von Ultra Wideband Modulen zur Emulation von GPS Positionen, Februar 2020, Chemnitz

CSR-20-04 Christian Graf, Reda Harradi, René Schmidt, Wolfram Hardt, Automatisierte Kameraausrichtung für Micro Air Vehicle basierte Inspektion, März 2020, Chemnitz

CSR-20-05 Julius Lochbaum, René Bergelt, Time Pech, Wolfram Hardt, Erzeugung von Testdaten für automatisiertes Fahren auf Basis eines Open Source Fahrsimulators, März 2020, Chemnitz

Chemnitzer Informatik-Berichte

CSR-20-06 Narankhuu Natsagdorj, Uranchimeg Tudevdagva, Jiantao Zhou, Logical Structure of Structure Oriented Evaluation for E-Learning, April 2020, Chemnitz

CSR-20-07 Batbayar Battseren, Reda Harradi, Fatih Kilic, Wolfram Hardt, Automated Power Line Inspection, September 2020, Chemnitz

CSR-21-01 Marco Stephan, Batbayar Battseren, Wolfram Hardt, UAV Flight using a Monocular Camera, März 2021, Chemnitz

CSR-21-02 Hasan Aljzaere, Owes Khan, Wolfram Hardt, Adaptive User Interface for Automotive Demonstrator, Juli 2021, Chemnitz

CSR-21-03 Chibundu Ogbonnia, René Bergelt, Wolfram Hardt, Embedded System Optimization of Radar Post-processing in an ARM CPU Core, Dezember 2021, Chemnitz

CSR-21-04 Julius Lochbaum, René Bergelt, Wolfram Hardt, Entwicklung und Bewertung von Algorithmen zur Umfeldmodellierung mithilfe von Radarsensoren im Automotive Umfeld, Dezember 2021, Chemnitz

CSR-22-01 Henrik Zant, Reda Harradi, Wolfram Hardt, Expert System-based Embedded Software Module and Ruleset for Adaptive Flight Missions, September 2022, Chemnitz

CSR-23-01 Stephan Lede, René Schmidt, Wolfram Hardt, Analyse des Ressourcenverbrauchs von Deep Learning Methoden zur Einschlagslokalisierung auf eingebetteten Systemen, Januar 2023, Chemnitz

CSR-23-02 André Böhle, René Schmidt, Wolfram Hardt, Schnittstelle zur Datenakquise von Daten des Lernmanagementsystems unter Berücksichtigung bestehender Datenschutzrichtlinien, Januar 2023, Chemnitz

CSR-23-03 Falk Zaumseil, Sabrina, Bräuer, Thomas L. Milani, Guido Brunnett, Gender Dissimilarities in Body Gait Kinematics at Different Speeds, März 2023, Chemnitz

CSR-23-04 Tom Uhlmann, Sabrina Bräuer, Falk Zaumseil, Guido Brunnett, A Novel Inexpensive Camera-based Photoelectric Barrier System for Accurate Flying Sprint Time Measurement, März 2023, Chemnitz

CSR-23-05 Samer Salamah, Guido Brunnett, Sabrina Bräuer, Tom Uhlmann, Oliver Rehren, Katharina Jahn, Thomas L. Milani, Güunter Daniel Rey, NaturalWalk: An Anatomy-based Synthesizer for Human Walking Motions, März 2023, Chemnitz

Chemnitzer Informatik-Berichte

ISSN 0947-5125

Herausgeber: Fakultät für Informatik, TU Chemnitz
Straße der Nationen 62, D-09111 Chemnitz