

Pantomime: Motion Data Anonymization Using Foundation Motion Models

Simon Hanisch
Centre for Tactile Internet (CeTI)
Technical University Dresden
Dresden, Germany
simon.hanisch@tu-dresden.de

Julian Todt
KASTEL Security Research Labs
Karlsruhe Institute of Technology
Karlsruhe, Germany
julian.todt@kit.edu

Thorsten Strufe
KASTEL Security Research Labs
Karlsruhe Institute of Technology
Karlsruhe, Germany
thorsten.strufe@kit.edu

Abstract

Human motion is a behavioral biometric trait that can be used to identify individuals and infer private attributes such as medical conditions. This poses a serious threat to privacy as motion extraction from video and motion capture are increasingly used for a variety of applications, including mixed reality, robotics, medicine, and the quantified self. In order to protect the privacy of the tracked individuals, anonymization techniques that preserve the utility of the data are required. However, anonymizing motion data is a challenging task because there are many dependencies in motion sequences (such as physiological constraints) that, if ignored, make the anonymized motion sequence appear unnatural. In this paper, we propose *Pantomime*, a full-body anonymization technique for motion data, which uses foundation motion models to generate motion sequences that adhere to the dependencies in the data, thus keeping the utility of the anonymized data high. Our results show that *Pantomime* can maintain the naturalness of the motion sequences while reducing the identification accuracy to 10%.

Keywords

anonymization, privacy, motion data, mixed reality

1 Introduction

Human motion data is a rich source of information with various applications in medicine, *Mixed Reality* (MR), or robotics. As motion tracking systems become cheaper and easier to use, we are now seeing more widespread adoption of motion tracking in everyday life. Examples of this are full-body motion tracking *Inertial-Measurement-Unit* (IMU)-suits and similar IMU-trackers for MR applications. Another example is the hand motion tracking most MR headsets can now perform using video cameras. The captured motion data is used for applications such as the animation of digital avatars in social MR scenarios [22], the remote control [5] and training [40] of robots, as input modalities for MR devices in general [7], or for the monitoring of rehabilitation efforts in medicine [35].

While these applications demonstrate how valuable motion data can be, it also carries risks, as it is behavioral biometric data that contains inherently sensitive information about the recorded person. Identifying people through their gait (pattern of walking) has long been a research topic for surveillance systems as it can be

performed even on low-quality video data [44]. In recent years, 3D motion capture data has also been investigated for gait recognition purposes, and it has been shown that the identification of individuals is simple and robust [8, 14]. Further, it has been shown that diseases like Parkinson's can be inferred from gait data [1].

Therefore, to prevent the user identification from motion data, while preserving its utility, we require anonymization techniques for full-body motion data. However, as prior work [8] has shown motion data is difficult to anonymize while maintaining high utility, because it contains a large number of correlations between data points, which an anonymization must consider. Otherwise, it is possible to break the protection or even reconstruct the original data from the anonymized data [10, 43].

Therefore, we propose *Pantomime*, the first anonymization for full-body motion sequences that is robust against re-identification while enabling high utility. *Pantomime* uses foundation motion models to hide the identity of people captured in motion data, which makes it applicable to different motion data formats and eliminates the requirement to train on the data it should anonymize. The use of foundation motion models allows *Pantomime* to encode the motion data into the motion space. *Pantomime* then anonymizes the motion data in the motion space by adding random noise to it before decoding the motion data back into its original space. The advantage of this approach is that by adding noise to the motion space, we create a new plausible motion that is similar to the original. By increasing the noise, the new motion can be further away from the original motion, and thus anonymization can be increased at the expense of motion utility, allowing us to configure the privacy-utility tradeoff of *Pantomime*.

Furthermore, using foundation motion models has the advantage that they are trained on general motion data and therefore generalize well to different motions [36]. Because of this, *Pantomime* does not require the enrollment of specific users or the tuning of the model to specific datasets, as in previous work [29]. Moreover, *Pantomime* is applied to the motion data time-step by time-step, making it flexible in its application to motion data (e.g., when animating an avatar in a MR chat application).

We investigate the privacy-utility tradeoff of *Pantomime* by measuring the utility via naturalness and by comparing the similarity of the anonymized sequence to the original using a user study. To measure the privacy protection we measure the person identification accuracy with a state-of-the-art biometric recognition system [14].

Furthermore, we investigate how much of the individual components (e.g., body shape, joint rotations, etc.) of the motion data contribute to person identification. To better understand which component requires anonymization. The contributions of this work are as follows:

This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license visit <https://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.



Proceedings on Privacy Enhancing Technologies 2026(2), 384–396
© 2026 Copyright held by the owner/author(s).
<https://doi.org/10.56553/popets-2026-0053>

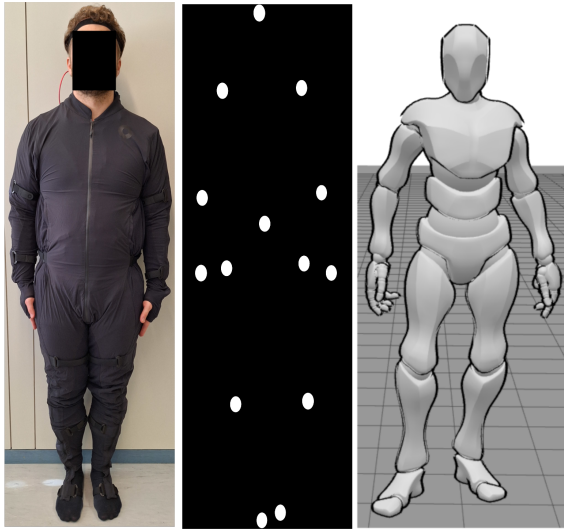


Figure 1: Left a person in an IMU-suit. Middle a point light display representation of motion data. Right a digital avatar.

- We propose Pantomime, the first general technique for anonymizing motion data that does not require training on the dataset to be anonymized and can configure its privacy-utility tradeoff.
- We evaluate Pantomime’s privacy protection using three full-body motion capture datasets.
- We conduct a user study to investigate the naturalness and action similarity of the anonymized motion sequences.

The paper is organized as follows. First, we introduce the background in Section 2. We then examine the relevant related work in Section 3 before describing Pantomime in Section 4. We then describe how we evaluated Pantomime in Section 5 and present the results in Section 6. The paper ends with a discussion in Section 7 and a short conclusion in Section 8.

2 Terminology & Background

In the following we introduce the terminology, autoencoder background, the SMPL body model, and foundation models used in this paper.

2.1 Terminology

Here we define the terminology for motion data and gait that we use in this paper.

A *foundation model* is, as defined by Bommasani et al. [2]: "... any model that is trained on broad data (generally using self-supervision at scale) that can be adapted (e.g., fine-tuned) to a wide range of downstream tasks". In this paper, we adapt motion models trained on a large motion corpus for motion sequence encoding and decoding to the task of anonymization.

Throughout this paper we will be working with *motion sequences*. Motion sequences are time series of body poses. A *body pose* consists of the position and orientation of all body parts of the captured person at a given point in time. In this paper we will use two

representations of body poses. The first is the representation as a set of 3D points located at fixed body landmarks (e.g., shoulders) of the person. The second is the representation as a body shape plus the joint rotations of the individual body parts (e.g. arms, legs, ..).

Walking is a cyclic motion. For this work, we use the definition of a *gait cycle* from Perry et al.’ [34]. There, a full gait cycle is completed when the feet are in the same relative position to each other as they were at the beginning of the cycle.

For our user study we use *point-light displays* [6]. Point-light displays (see Figure 1) are a well-established method from psychology for representing motion data. Instead of displaying the whole body, only certain landmarks of the body are displayed as white dots on a black background. The advantage of point-light displays is that they allow viewers to judge the execution of a motion without being influenced by the appearance of the person performing the motion.

2.2 Autoencoders

Pantomime uses foundation models that use variants of *Autoencoder* (AE) architectures, and relies on some of the features of these architectures. This is because anonymization is performed in their latent space. Here we focus on the features of AE and its variants that are important for Pantomime.

An AE [17] is a machine learning architecture consisting of an encoder model and a decoder model. The encoder’s goal is to translate the data into a much smaller latent space, which the decoder then translates back into the original data space. In other words, the encoder performs a compression of the data and the decoder decompresses it back to its original form. The overall goal of the model is to learn an efficient coding of the input data. Since the output of the model should be the same as the input, it can be trained unsupervised by using a reconstruction loss (e.g., *Mean Squared Error* (MSE)) between the encoder input and the decoder output. *Variational Autoencoder* (VAE) [16] is a type of AE that, instead of learning a discrete latent code for a given input, maps the input to the parameters of a probability distribution. A discrete latent code is then drawn from this probability distribution and decoded into the input space by the decoder. When training a VAE, the shape of the learned latent distribution is often regularized by using a Kullback-Leibler divergence [18] to resemble a normal distribution. In this case, the learned latent space can be interpreted as a mixture of normal distributions.

Several VAE variants exist. β -VAEs [11] use a weighted Kullback-Leibler divergence to increase its influence in the loss and thereby force a disentanglement of the dimensions of the latent code [3, 39]. Another type of VAE is a *Conditional Variational Autoencoder* (CVAE) [39] that uses an additional label to constrain the latent code of an input to be deterministic. In this way, only the label can be fed to the decoder to generate a sample that belongs to the class of the input label.

2.3 SMPL Body Model

The *Skinned Multi-Person Linear Model* (SMPL) [21] model is a body shape model that uses blend shapes to represent the human body shape for different poses. The SMPL decomposes the body shape into a fixed identity-based body shape and a variable shape that

depends on the body pose, which is represented as joint rotations. Due to this decomposition into static and dynamic body pose, the SMPL model respects the body shape deformation that occurs in different poses, i.e. soft tissue deformations when a person is moving compared to when the person is standing still. The SMPL model has been trained to minimize the reconstruction error of high-resolution 3D body scans. For the remainder of the paper, the SMPL model is defined as a differentiable function $M(r, \Phi, \Theta, \beta)$ mapping the root translation $r \in \mathbb{R}^3$, root rotation $\Phi \in \mathbb{R}^3$, body pose as joint angles $\Theta \in \mathbb{R}^{3 \times 21}$, and identity-based body shape parameters $\beta \in \mathbb{R}^{16}$ to the vertices $V \in \mathbb{R}^{3 \times 6890}$. The joint positions $J \in \mathbb{R}^{3 \times 22}$ can be calculated from the vertices using a regressor matrix.

To find the SMPL model representation of a given body pose, we perform *3D body fitting*. The goal of the fitting is to find the best parameters for the SMPL model to fit the given data of the body pose (e.g. all limb positions of a person).

2.4 Foundation Models

In the following, we will introduce the two foundation motion models VPoser and HuMoR that we use for Pantomime.

2.4.1 VPoser. VPoser [32] is a VAE that has learned the probability distribution of plausible human body poses. It takes a single body pose as input and tries to output the same body pose. Since it models plausible poses, it can be used to judge whether a given pose is plausible or not. This is used to fit the SMPL body model to given motion data by using the model as a loss.

VPoser uses the normal distribution $\mathcal{N}(\mu, \sigma)$ for its latent space. The VPoser encoder takes body poses as joint angles Θ as defined by the SMPL body model and outputs a latent code z given the learned parameters ϕ and θ . The encoder can be formulated as $q_\phi(z_t|\Theta_t) = \mathcal{N}(z_t; \mu_\phi(\Theta_t), \sigma_\phi(\Theta_t))$. The latent code is then mapped back into the original pose space by the decoder $p_\theta(\Theta_t|z_t)$ to get the reconstructed pose $\hat{\Theta}_t$.

2.4.2 HuMoR. HuMoR [36] is a CVAE that has learned the transition from one pose to the next. Unlike VPoser, it uses two successive poses as input to its encoder and then tries to output the transition from the first to the second pose with its decoder. With its focus on pose transitions, HuMoR effectively models the distribution of human motion. Like VPoser, HuMoR can be used to fit body models to motion sequences.

HuMoR represents the state of a moving person as a matrix $x \in \mathbb{R}^{69}$ consisting of root translation, root orientation, body joint angles, body joint positions, and the velocities of root translation, root orientation, and joint positions. The encoder of HuMoR takes two sequential states x_t and x_{t-1} as input and generates a latent code z_t via the parameterization of a normal distribution $q_\phi(z_t|x_t, x_{t-1}) = \mathcal{N}(z_t; \mu_\phi(x_t, x_{t-1}), \sigma_\phi(x_t, x_{t-1}))$ with the learned parameters ϕ . The latent code z_t is then decoded by the decoder $p_\theta(x_t|z_t, x_{t-1})$ into the state change Δ_θ and the ground contacts c_t with the learned parameters θ . The reconstructed motion state \hat{x}_t is then $\hat{x}_t = x_{t-1} + \Delta_\theta(z_t, x_{t-1})$.

3 Related Work

Below is an overview of research in the emerging field of motion anonymization that attempts to prevent identification. We categorize the works according to the type of motion data they anonymize.

3.1 Body Shape Anonymization

Sattar et al. [38] investigated the privacy of body shapes extracted from single images. They show that a person’s body shape is considered private information by performing a small user study, and propose an adversarial perturbation to prevent the automatic extraction of shape information from images. While this work is not motion anonymization, it highlights the need for anonymizing the body shape information that is implicitly contained in motion data.

3.2 Video Anonymizations

After motion data was identified as private information, some early anonymization approaches were proposed to protect it from inference. An early approach to anonymize human gait in videos was proposed by Tieu et al. [42], where the silhouette and a noise silhouette of a walking person are mixed using a *Convolutional Neural Network* (CNN) trained similarly to a AE with the weighted MSE of the original and noise silhouettes as the reconstruction loss. Hirose et al. [12, 13] investigated how gait videos can be anonymized by deforming the silhouette of the walking person and then computing a new texture for it to preserve utility. The approach uses a AE that learns a code for the silhouette of the walking person. It also estimates the phase of the gait. Both the phase and the shape coding are then anonymized before being fed to the decoder to generate a new gait silhouette. The anonymization of the shape code is done using a k -anonymity approach, where the k -nearest shape codes are mixed and averaged with the encoded one. The last step is to compute a new texture for the silhouette. Romero et al. propose GaitGuard [37], a real-time anonymization for gait in the camera view of MR devices, where the anonymization is offloaded to a local server. For the anonymization itself, different approaches such as applying Gaussian blur have been tested. The goal of GaitGuard is to prevent pose extraction.

The anonymizations described above only work for 2D gait videos and are therefore limited, as they cannot be applied to 3D motion captures and do not work for motion data in general. However, the use of a AE for silhouette anonymization can be adapted for 3D motion data.

3.3 Egocentric Video Anonymization

Thapar and Arora [41] propose an anonymization technique for egocentric gait videos. In egocentric videos, the person using the camera is never seen directly, as the camera captures the person’s field of view. However, it is still possible to infer the identity of the person recording the video. The anonymization technique uses machine learning to estimate the rotation of the camera and then adds the rotations of another video to anonymize it. Egocentric video anonymization can be considered a special case of 3D motion anonymization, since only the 3D rotation of the camera is anonymized. Full body motion data consists of a set of 3D positions and rotations, so this approach is not applicable.

3.4 Motion Capture Anonymization

Malek-Pdjaski and Deligianni [24] developed an anonymization technique for 3D motion capture that extracts features that do not allow identification but can still be used for affect recognition. They attempt to separate the information needed for affect recognition from the information used for identification by using two AE. One AE is trained to be subject-specific and one AE is trained to be affect-specific. The disadvantage of this approach is that the AEs have to be trained on the dataset to be anonymized. Moon et al. [27] proposed an adversarial anonymization scheme for 3D motion capture data in which a machine learning model is trained to minimize the identity recognition and maximize the action recognition. Both approaches are not suitable for preserving the naturalness of the motion data (as determined by a user study), since the benefit must be quantifiable so that it can be used as a loss in the training of these approaches.

Simple Anonymizations. Moore et al. [28] suggest using only the velocities of the motion sequences and Miller et al. [26] suggest using only the joint rotations to reduce identifiability. Meng et al. [25] additionally propose adding noise to the joint rotations. Hanisch et al. [8] investigated the effect of various simple anonymization techniques such as noise injection or removal of specific body parts from 3D motion capture data. All of these simple anonymization techniques fail to achieve a good privacy-utility tradeoff and show that identity recognition is possible even on heavily corrupted data.

Mixed Reality Anonymizations. Nair et al. investigate how the (motion) data collected by *Virtual Reality* (VR) headsets can be anonymized. They proposed a framework called MetaGuard [30] which claims to protect various attributes collected by VR headsets. Their second proposal *Deep Motion Masking* (DMM) [29] is a machine learning approach which reduces identity similarity while maintaining action similarity. The evaluation of the approaches performed in [29] shows that DMM is effective and can anonymize motion sequences of different datasets against different attackers, but MetaGuard is not effective and does not provide sufficient protection. The drawback of DMM is that it requires a large training dataset and is application specific for the data it was trained on.

3.5 Summary

In summary, previous work is limited in several ways. They either require large amounts of training data, are action specific (gait only), cannot be applied to full body motion data, do not preserve the utility of the data, or are simply not effective.

4 Methodology

As previous work [8, 29] has shown, anonymizing motion data is a challenging task. The main problem is that motion data contains a large number of dependencies between the individual tracked points, and constraints such as the maximum degree of flexion of certain joints. In addition, the physiology of the person performing the motion is important, as it strongly influences how motions are executed to perform the same action. For example, a tall person will bend their shoulder joint differently to grab an object from a table than a shorter person in the same situation. Because of all these dependencies, directly modifying the data is either not effective because the dependencies can be used to reconstruct the original data,

or the modification has to be very strong, which greatly reduces the utility of the motion data. Another interpretation of the dependencies is that they represent redundancy in the data, since the true dimension of the motion data, which can be changed independently, is much smaller than the recorded positions.

The main idea of Pantomime is to remove as many of the dependencies described above as possible before performing anonymization, and then reintroduce the dependencies after anonymization to generate a new sequence of motion data. We perform the dependency removal by mapping a motion sequence into the motion space of a foundation motion model. We then anonymize the motion sequence in the motion space by adding noise to it. Finally, we map the anonymized motion sequence back to its original position space.

4.1 Requirements

The two main goals of motion anonymization are to prevent the identification of an individual from their respective motion sequences, and to preserve the utility of the motion sequences for the application for which they are intended. For Pantomime, the utility goals are naturalness and action similarity. Naturalness means that the motion sequence appears as a genuine motion sequence to a human observer. And action similarity means that the action in the anonymized sequence should be as similar as possible to the original one. We chose naturalness because for many applications of full-body motion data, such as social interactions in MR, it is important that the motions appear believable and realistic. The action similarity should prevent our anonymized motion sequences from deviating too much from the original ones, otherwise we could just generate random motion sequences to satisfy the naturalness goal. In addition, it would be beneficial to meet the following requirements derived from common applications of motion data. The first application requirement is the applicability of anonymization to full-body tracking, since even from sparse input data, such as tracking from MR devices, the full-body pose can be estimated. The second is that the anonymization should be general with respect to the format in which the motion data is captured, since motion capture systems vary in the number of points tracked and the specific body landmarks that are captured.

4.2 Adversary Model

The main goal of the adversary is to identify which person belongs to which motion sequence. We assume a strong adversary as described by Hanisch et al. [8] in that the adversary is aware of the existing anonymization and can adapt their attack to it. The adversary becomes stronger by adapting their attack, resulting in more rigorous evaluation results. To achieve their goal, the adversary uses a biometric recognition system to perform identification. To train the recognition system and learn templates for each person under attack, the adversary has access to clear motion data of these people. Since the adversary knows the anonymization used and its parameters, they can anonymize the clear data. By training the recognition system on anonymized data, the attacker can adapt the recognition system to the changes that the anonymization makes to the original data, and thus adapt their identification attack to the

anonymization, thus posing a greater challenge to the anonymization.

4.3 Anonymizing Human Motion

We will now explain how Pantomime works. As a first step, we unify different motion capture formats into the format of the SMPL model by performing a fitting step. This step is necessary because we want Pantomime to be general enough to work with different formats of motion data. After transforming the original data into SMPL data, we map the data into the motion space of a foundation motion model and then perform anonymization by adding noise to the data. The intuition here is that since the motion space encodes plausible motions, by modifying the input data in this space, we end up with a plausible motion for the output of the anonymization. In contrast, performing anonymization by adding noise to the original position data quickly leads to implausible motions because the individual points are modified without adhering to the given physiology of the body or physics. The final step is to decode the anonymized motion space data back into the SMPL model format and then back into the original motion data format.

In the following, we describe the different steps of our anonymization pipeline for Pantomime, an overview of the whole process can be seen in Figure 2.

4.3.1 SMPL Model Fitting. To unify the motion data formats, we perform a fitting of our original motion sequence to the SMPL body model using the HuMoR [36] fitting method. This way all of our motion sequences have the same format and can be used with the same foundation motion models. The fitting process is a function optimization that is performed for each pose of a motion sequence to find the best SMPL parameters to represent that pose.

Given an observed motion sequence, we try to find the parameters of the SMPL body model $M(r, \Phi, \Theta, \beta)$ that best describe the sequence. To perform this function optimization, three loss terms must be minimized. The first is the plausibility of the motion. Here we use either VPoser or HuMoR as a prior to measure the plausibility of a given pose (VPoser) or the plausibility of a given pose transition (HuMoR). The second loss term is a reconstruction loss comparing the original with the found SMPL joint positions using MSE. Since the skeletons between motion data representations may have different kinematic structures, we define a mapping between the SMPL body model skeleton and the dataset we are trying to fit to compute this reconstruction loss. The third loss is a regularization term that enforces consistency of bone length, ground contact, ground position, and body shape across the entire motion sequence (i.e., bone length and body shape should not change much during a single motion sequence). The weighting of the three loss parts is dataset specific and is determined by performing a hyperparameter optimization to find the optimal values. A more detailed description of the loss terms can be found in [36].

This first step can also be seen as a decoupling of the dependencies, since the original position data is split into the body shape β and the joint angles Θ , effectively decoupling these two aspects of the data. While this has a positive effect on the anonymization, the downside is that the fitting process decreases the quality of the original data.

4.3.2 Encoding the Motion Sequences. Next, we encode the motion sequences into the latent space of a foundation motion model. The foundation model used for this step is interchangeable, as Pantomime only requires that it be a VAE that encodes from the SMPL parameters to a latent space of plausible poses or pose transitions and back to the SMPL parameters. The rationale behind this requirement is that VAEs compress the original data by removing dependencies [3, 39] (i.e., correlations) between data points, and thus our anonymization no longer needs to adhere to these dependencies for effective anonymization. In other words, by anonymizing in a space of plausible motions, where similar motions are close to each other, the resulting motion is plausible and similar to the original, thus preserving the data’s utility. We will later test this assumption through a user study. In this paper, we use VPoser and HuMoR as two possible foundation motion models that satisfy this requirement. However, it is important to note that VPoser only takes the body pose (joint angles Θ) as input, while HuMoR also requires the root orientation, root translation, and body shape.

4.3.3 Anonymizing the Latent Code. Now we anonymize the latent code. To do this, we draw a noise vector p from a normal distribution $\mathcal{N}(\mu, \sigma)$ with $\mu = 0$ and $\sigma = 1$ and add it to the latent code z_t at time step t . We chose normal distributed noise because VAEs regularize their latent space to be normal distributed and adding two normal distributed random variables results in a normal distributed sum [19]. Since the decoder expects a normally distributed random variable as input, this should result in the least utility loss. We chose the mean μ to be 0 and the σ to be 1 because we do not want to introduce a bias into the anonymized motion sequences. With a mean of 0, the resulting distribution of anonymized motion sequences will cluster around the original motion sequence. The noise vector p is scaled by the scalar γ to make the anonymization configurable by increasing the distance in motion space to the original sequence. There are two different modes for adding the noise. In *variable* we draw a new noise vector p_t for each time-step z_t , giving us $a_t = \gamma p_t + z_t$ for the anonymized latent code a_t , while in *static* we add the same noise vector p to each z_t , giving us $a_t = \gamma p + z_t$.

4.3.4 Decoding the Motion Sequence. The last step is to decode the anonymized latent code a_t into the parameters of the SMPL body model using the motion model, and from there back into the original position format. Note that for Pantomime, we only focus on the anonymization of the body pose, the Θ parameter of the SMPL body model. We do not consider the anonymization of the body shape, the root translation, and the root orientation and remove them by setting their respective parameters to zero. We do this because we assume that a person who wants to be anonymous will choose a digital body that does not resemble them, including the body shape. The root translation and orientation can be estimated from the resulting motion sequence.

5 Evaluation

We now evaluate the privacy-utility tradeoff of Pantomime to understand how much noise must be added for effective anonymization and how much utility is retained. We also test the assumptions we made when designing Pantomime.

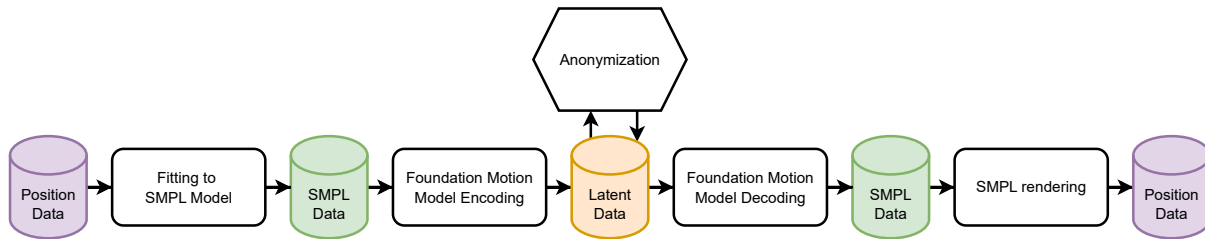


Figure 2: The Pantomime anonymization pipeline.

5.1 Datasets

We select our datasets to include a large number of full body motion capture sequences with a preference for the gait task, as this has been shown to be highly identifiable. For our evaluation datasets, we do not want overlap with the datasets used to train the foundation models that we use for anonymization. Therefore, the foundation motion models must be able to generalize to new datasets that they did not encounter during training. Thus, we do not use the AMASS [23] dataset, or any of the datasets included in it, because AMASS was used to train VPoser and HuMoR.

For our evaluation, we use the *CeTI-Locomotion* [9], Horst-DB [14], and HuMMan [4] datasets. *CeTI-Locomotion* includes a variety of walking modalities as well as sit-to-stand exercises in which participants stand up and then sit down as quickly as possible. The dataset was recorded from 50 healthy participants using an IMU suit that captures the relative motion of each body part. In combination with the anthropometric measurements of the participants, the 17 body segment positions are calculated. Horst-DB contains only one walking modality and was recorded using optical motion tracking with 54 reflective markers attached to body landmarks such as joints or the iliac crest. The resulting data are scalar values for the x,y,z positions of each of the markers. The HuMMan dataset was recorded for action recognition. We specifically use the motion generation subset which contains actions performed by participants recorded by a camera dome. We use the already fitted SMPL parameters for our experiments. As there are not enough samples for each participant and each action we had to filter the HuMMan dataset (see Section 5.2.1 for details).

The main difference between the three datasets is that *CeTI-Locomotion* includes several different walking modalities (normal, fast, carrying a backpack, and carrying a bottle crate) plus an additional sit-to-stand exercise. Further, the HuMMan dataset was designed for action recognition containing a high number of similar actions and does not require the SMPL fitting as the data is already in the SMPL parameter format.

Also, the HuMMan camera tracking and the IMU tracking are less accurate than optical marker tracking, which is considered the gold standard of motion tracking. See Table 1 for a comparison of the three datasets.

5.2 Implementation

Here we describe the implementation details for both Pantomime and the biometric recognition systems we will use for our experiments.

Table 1: Overview numbers of the used evaluation dataset

Name	Points	Participants	Tasks	Samples
CeTI-Locomotion	17	50	5	4672
Horst-DB	54	57	1	1140
HuMMan (filtered)	17	55	117	2294

5.2.1 Data Preparation. We preprocess all of our evaluation datasets to have a frame rate of 30 Hz, the same as in the previous work [36]. For the Horst DB dataset, we additionally trim the samples to exactly one gait cycle (see Section 2.1) using the additional force plate data and a threshold to identify the first and last pose of the cycle, as described by Horst et al. [14]. For the HuMMan dataset we have to filter the participants and actions to have enough samples for each. We first select only the participants which have at least 40 samples and then from these samples filter all actions that have less than 10 samples. This results in 55 participants, and 117 actions, with a total of 2294 samples.

5.2.2 SMPL Parameter Fitting. Overall, the goal of the fitting is to find the parameters of the SMPL model ($M(r, \Phi, \Theta, \beta)$) that match the positions of the input data as closely as possible, as well as to achieve a high plausibility with the used foundation motion model (here HuMoR or VPoser). The foundation motion model is used here to enforce the generation of only plausible motions for the SMPL poses. To fit our two evaluation datasets we use the code of HuMoR [36], which implements the whole process in three steps. In the first stage, the root translation and rotation are optimized using VPoser as a prior, in the second stage, the entire SMPL parameters are optimized using VPoser as a prior, and in the third stage, HuMoR is used as a prior for the optimization. Since the code performs the optimization first with VPoser and then with HuMoR, we use it to generate both fits.

To perform the fitting for the *CeTI-Locomotion* and Horst-DB dataset (the HuMMan dataset already is in the SMPL format), we need a mapping from our motion data joint positions to the SMPL joint positions (i.e., which positions in our data correspond to which position in the SMPL joints). We created the mapping manually as follows. For *CeTI-Locomotion* this mapping is undercomplete because *CeTI-Locomotion* has only 17 joints while the SMPL model has 21. For the Horst DB, it is the other way around because the datasets tracked 54 points, some of which are ignored while others are combined to better match the joint positions of the SMPL body model. Joint positions for which there is no matching position in

the original data are set to infinity. To obtain good fitting results, we perform hyperparameter optimization on both datasets to find the weights for the fitting losses described in section 4.3.1. Due to the length of the fitting process, we perform the hyperparameter optimization on 10 random motion sequences of each dataset and then use the found parameters for all of them. We excluded some motion sequences from both datasets because the SMPL fitting process failed for them (producing NaN values at some stage of the process), and we were unable to obtain an SMPL body model representation. In both datasets, this accounted for less than 1% of the total motion sequences.

Although the SMPL model fitting allows Pantomime to be generalized by unifying and anonymizing other motion data formats, the quality of the motion data is reduced by the fitting and the fitting process itself is slow and resource-intensive.

As Pantomime removes the root translation from the sequences, we estimate a new root translation for the anonymized sequences by using the absolute trajectory of the right foot as the root translation.

5.2.3 Biometric Recognition System. For the biometric recognition system, we adapt a state-of-the-art method used in previous work [8, 14], in which the motion sequence is resampled to a fixed 100 frames and then flattened into a single vector. Note that fixed-length sequences are only used for evaluation. Anonymization can be performed on sequences of arbitrary length. The samples are then divided into 80% training data and 20% test data in a stratified fashion, while keeping the same percentage of samples per identity class. To account for the randomization during the data splitting we run every experiment 10 times and report the average value, as well as the 95% confidence interval. We use the same system for action recognition, but there we split the data so that a single participant’s samples are either in the test data or in the training data. This should prevent the system from learning actions for specific people and test generalization. The datasets are then processed independently. Each dimension of the feature vector is min-max normalized before its dimensionality is reduced using a *principal component analysis* (PCA). As a classifier, we use a *Support Vector Machine* (SVM) with a *radial basis function* (RBF) kernel, which is trained in a 5-fold stratified cross-validation procedure using balanced accuracy as a metric. The final result is obtained by running the SVM on the test data. The recognition system is always trained on the anonymized data of the anonymization we are testing, as defined by our adversary model. In addition to this simple SVM system, we use a variation of EKYT [20], a deep learning recognition system that uses a DenseNet architecture. This architecture works well with medium-size datasets, such as the ones we use. We use the same training-to-test data ratio as we do for the SVM system. We train for 200 epochs, with a batch size of 1280 using Adam as the optimizer and a learning rate of 0.001 and a learning step size of 50.

5.2.4 Code Availability. We implemented the biometric recognition system using Python, scikit-learn [33] and PyTorch [31]. Pantomime itself is implemented in Python on top of the existing HuMoR code [36]. The anonymization and evaluation code for Pantomime can be found here: <https://github.com/kit-ps/Pantomime-Motion-Data-Anonymization-using-Foundation-Motion-Models/blob/main/>.

5.3 Experiments

Here we detail the experiments we performed to investigate our underlying assumptions in designing Pantomime, and then to evaluate Pantomime’s privacy-utility tradeoff.

5.3.1 Assumptions & Baseline.

Baseline Identification & Action Recognition. First, we establish an identification and action recognition baseline on the original position data, against which we will later compare the anonymization results of Pantomime. This will allow us to evaluate how good Pantomime’s privacy protection is and how much utility we lose as a result. In our baseline identification experiment **E1**, we train and test the biometric recognition system on the dataset to perform person identification. To do this, we split the dataset into a test and a training part of the dataset, with each person having different samples in both partitions. We then train the biometric recognition system on the training dataset in a supervised manner. We then determine the identification performance on the test set. In the action recognition baseline experiment **E2**, we determine the baseline action recognition performance. We now train and test an action recognition system. Unlike for person identification, we split the dataset so that a person is either in the test or training dataset to avoid the system learning the unique action performance of a person and to better generalize. We then measure how well the system can identify the action performed in the sample.

Identification Potential SMPL Parameters. The unification of the motion data in the SMPL body format splits the data into motion data (poses as joint angles Θ), body shape β , root translation r , and root orientation Φ . Pantomime focuses only on anonymizing the motion data of the poses. For the remaining SMPL parameters, we test how much identification potential they have on their own.

For **E3** we use the SMPL body model fits from both VPoser and HuMoR for our motion data where specific components are removed by setting them to zero. We then generate new position data for the classification. We expect the body shape and poses alone to carry a high identification potential, as has been shown in previous studies [8, 38]. For the root translation and especially for the root rotation we expect a lower identification potential, because these are single vectors, which should carry less information than the body poses.

Noise Mode Comparison. We assume that adding the same noise vector to all poses of a motion sequence (static noise mode) will perform better than adding a new noise vector to each pose (variable noise mode). In our noise mode comparison experiment **E4**, we run both modes with different values of the noise scaling γ on the original motion data to test this assumption. We then measure privacy by performing identification with our biometric recognition system. We expect that the static noise mode will always outperform the variable noise mode. Since a motion sequence is a time series of poses, two consecutive poses will be very similar to each other because a person cannot move much in a single time step. Adding different noise vectors to two similar poses makes it easier to separate the noise from the underlying data, since much of the difference between the two poses after the noise is added is the noise. If we add the same noise vector to both poses, then the

Table 2: The combinations of motion representations, anonymizations, and protection targets which we use in our privacy-utility tradeoff.

motion rep.	anonymization	protection targets
original	direct	10%, 20%
SMPL (VPoser)	direct, VPoser, HuMoR	10%, 20%
SMPL (HuMoR)	direct, VPoser, HuMoR	10%, 20%

difference between the two poses is still the same as it was before anonymization, and we cannot distinguish noise from data.

5.3.2 Privacy-Utility Evaluation. In our privacy-utility experiment E5 we evaluate the privacy and utility of Pantomime. For a better comparison, we not only study how noise injection into the latent space affects the privacy-utility tradeoff, but also test applying noise directly to the original data and the fitted SMPL representation. Due to the different representation of the motion sequence (original, SMPL fit, latent code), the noise parameters of our different anonymization techniques are not directly comparable. For example, adding the same amount of noise to the position of a joint will have a different effect than adding noise to the joint rotations of an SMPL fit. In order to achieve comparability between where we apply noise, we define protection targets. A protection target is a given value of recognition accuracy, for example 20%. We then tune the noise parameters of our anonymization to achieve the given target ($\pm 2\%$). In this way, the anonymization performance of our techniques is the same, and we can directly compare the utility of our approaches to judge which anonymization has the better tradeoff. An overview of all combinations of motion representation, anonymization, and protection target is given in Table 2.

We conduct a more thorough test of identification on recognition target motion sequences in E6. Here, we use EKYT biometric recognition to conduct another identification experiment. The rationale is to use a different system than the one used to select the recognition targets to determine whether other biometric systems might perform better than the simple SVM system employed in the other experiments. Additionally, we perform majority voting on all test motion sequences of a given participant to determine whether this improves identification.

We investigate utility by performing action recognition on the CeTI-Locomotion and HuMMan datasets and by conducting a user study on the CeTI-Locomotion and Horst-Study datasets. We investigate two utility goals in our user study, the first is the naturalness of an anonymized motion sequence, and the second is the motion similarity between an original and corresponding anonymized motion sequence.

VPoser only works on a single pose at a time, while HuMoR works on the pose transition and thus on pose pairs. Due to this, and the better overall performance for pose fitting reported by HuMoR [36], we expect HuMoR to do a better job of removing the dependencies, and therefore Pantomime to achieve better utility. Furthermore, we expect Pantomime to achieve better utility than adding noise to the original pose data, regardless of the foundation motion model chosen.

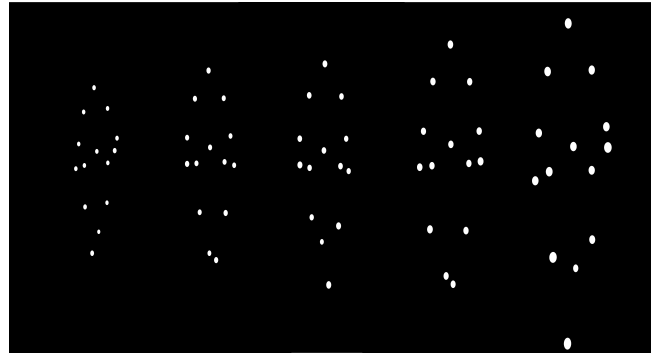


Figure 3: Example rendering of a motion sequence used during the user study. Link to animated examples: <https://github.com/kit-ps/Pantomime-Motion-Data-Anonymization-using-Foundation-Motion-Models/blob/main/Pantomime-Examples/>.

5.4 Utility / User Study

For utility, we investigate how natural the anonymized motion sequences appear and how similar the actions are compared to the original motion sequence. We do this by conducting a user study similar to prior work [43] on the evaluation of biometric anonymizations. We use two tasks to evaluate our objectives. In the first task, we show participants a single motion sequence. The participants then rate how natural the sequences appear by answering the question "Is this a natural human motion?" The rating is done on a 5-point Likert scale from "very unnatural" (1) to "very natural" (5). In the second task, we show the original sequence next to an anonymized version of the sequence. Participants then answer the question "How similar are the motions performed in the two videos?" by rating on a 5-point Likert scale from "very dissimilar" (1) to "very similar" (5).

We conducted an online survey in which participants were shown different motion sequences from the CeTI-Locomotion and Horst-DB datasets as short video sequences (see Figure 3 for an example) rendered at 20 frames per second. We reduced the rendering from 30 to 20 frames per second to allow users to better judge the execution of the motion. The motion is rendered as point-light displays to reduce the influence of appearance and to focus on the motion. From the Horst DB, we randomly select four motion sequences, two from male participants and two from female participants. From CeTI-Locomotion, we randomly select two male and two female motion sequences for the modalities *gait-normal*, *gait-fast*, *gait-bottle-crate*, and *sit-to-stand*. These 20 motion sequences were anonymized with a subset (original+direct, HuMoR+direct, HuMoR+HuMoR, VPoser+direct, and VPoser+VPoser) of the combinations described above (see Table 2) for the protection targets of 10% and 20%. We publish an overview¹ of the motion sequences used and their anonymized variants alongside this paper. Since we perform two tasks (naturalness and similarity see above) and add the original sequences, this results in 440 unique questions. From this question pool, each participant answers 40 random questions.

¹<https://github.com/kit-ps/Pantomime-Motion-Data-Anonymization-using-Foundation-Motion-Models/blob/main/Pantomime-Examples/>

Table 3: The influence of the individual parameters of the VPoser and Humor SMPL fits on the identification given as accuracy (95% confidence interval)

SMPL para.		datasets	CeTI-Loconotion	Horst-DB
VPoser	shape (β)		0.6 (0.59, 0.61)	0.46 (0.43, 0.49)
	joint poses (Θ)		0.26 (0.25, 0.27)	1.0 (1.0, 1.0)
	root trans. (r)		0.4 (0.4, 0.4)	0.51 (0.5, 0.52)
	root orient. (Φ)		0.19 (0.18, 0.2)	0.94 (0.93, 0.95)
HuMoR	shape (β)		0.1 (0.1, 0.1)	0.1 (0.09, 0.11)
	joint poses (Θ)		0.18 (0.17, 0.19)	0.47 (0.45, 0.49)
	root trans. (r)		0.21 (0.2, 0.22)	0.38 (0.37, 0.39)
	root orient. (Φ)		0.06 (0.06, 0.06)	0.18 (0.17, 0.19)

5.5 Ethical Considerations

The user study data collection was approved by the ethics commission of the Karlsruhe Institute of Technology (research project "Utility of Anonymized Motion Sequences") and was conducted in accordance with the Declaration of Helsinki. All data was collected in an anonymous online survey in December 2024 using an online recruitment platform² to recruit 224 participants (112 male, 112 female; mean age 30.8, std 9.57). Participation took a median of 7:47 minutes and participants were paid an average of 10.71€ per hour.

The CeTI-Loconotion and Horst-DB datasets used in this study were both approved by their respective ethics committees and their participants gave written informed consent to participate in the data collection.

6 Results

Here, we describe the results of our experiments, similar to the evaluation section, where we start with the results of testing the underlying assumptions for Pantomime before describing the results of the privacy-utility evaluation.

6.1 Assumptions & Baseline

We start with the baseline identification experiment **E1** and the action recognition experiment **E2**. For both CeTI-Loconotion and Horst-DB we observe that the identification recognition accuracy for the original position data is high with 83% and 96%, respectively, however, for HuMMan we only achieve 26% recognition accuracy. For the CeTI-Loconotion and HuMMan datasets, we also perform action recognition to classify which of the actions was performed in a motion sequence, achieving an accuracy of 80% for CeTI-Loconotion and 47% for HuMMan. Overall, these results are in line with prior work and show that the recognition systems used for identification and action recognition work, however, identification works significantly worse for the HuMMan dataset as this dataset was not designed for identification.

For **E3** we look at the difference in identification accuracy when using only certain parts of the data in their SMPL representation using the CeTI-Loconotion and Horst-Study dataset. In Table 3 we

²<https://prolific.com>

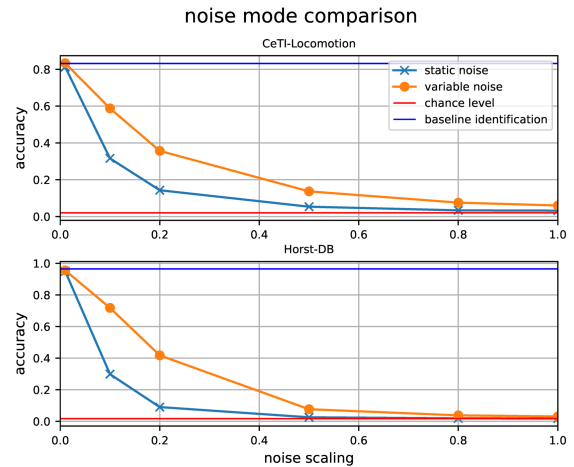


Figure 4: Comparison of the static noise vs. variable noise mode of directly applying noise to the original data of both CeTI-Loconotion and Horst-DB over different γ noise scaling choices.

report the identification accuracy for the data when using positions generated from the SMPL representation with only the specific parameter intact, while the rest of the parameters are set to zero. We find that most parameters have a high identification potential (greater than 50%) on their own. The lowest identification potential is observed for the SMPL fit of the CeTI-Loconotion data using HuMoR. But even here, the individual parameters have significantly more identification potential than the chance level (2%) for CeTI-Loconotion. This leads us to conclude that anonymizing only the SMPL joint poses in the latent space is not sufficient, as the remaining components of the SMPL body model can be used for identification. This justifies our decision to set the remaining parameters to zero.

Next, we report the results for our noise mode experiment **E4**, see Figure 4. For both datasets, we can see that the variable noise mode is always outperformed by the static noise mode when using the same noise scaling γ . When visually inspecting the resulting motion sequences using a rendering of the sequence, it can also be seen that the variable approach leads to a very visible shaking of the joint points, the injected noise becomes visible in the motion execution.

Since two consecutive poses are very similar to each other, changing the noise vector for each pose makes it easier to distinguish what is the real data and what is the noise addition. Our conclusion from this experiment is that the static noise mode is the better mode to add noise to the data because it always outperforms the variable noise for identification reduction.

6.2 Privacy-Utility Results

To investigate the privacy-utility tradeoff **E5**, we first report the action recognition results for the CeTI-Loconotion and HuMMan datasets, before doing the main comparison using our user study. Comparing the direct anonymization on the original data with all

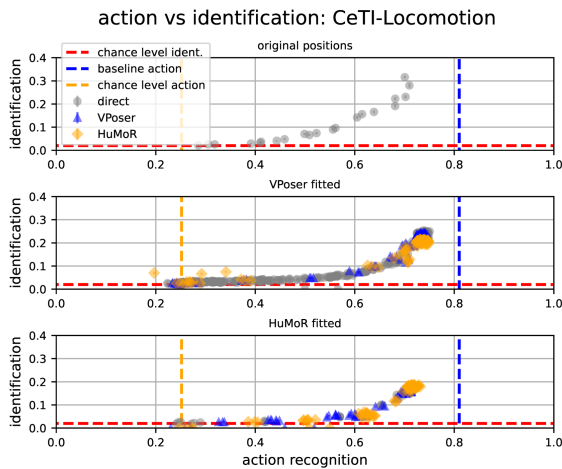


Figure 5: The action recognition vs identity recognition accuracy for CeTI-LoComotion. The error bars represent the 95% confidence interval.

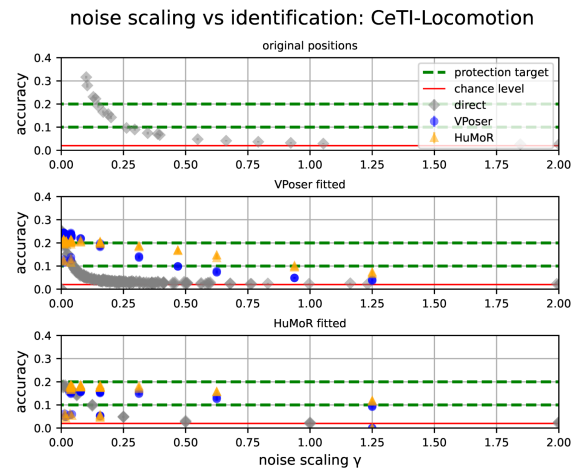


Figure 7: The scaling of the noise parameter γ in comparison to the achieved recognition accuracy of different anonymization combinations for CeTI-LoComotion. The error bars represent the 95% confidence interval.

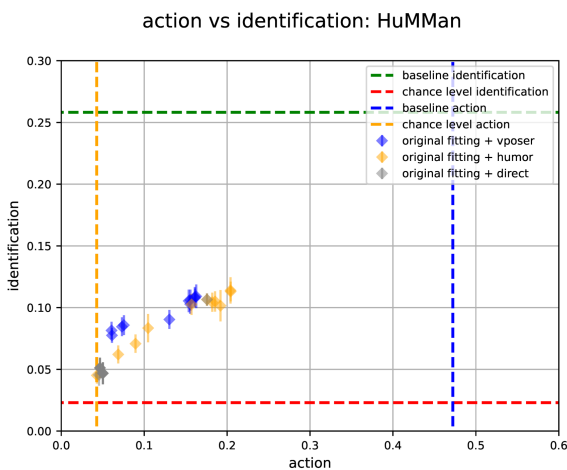


Figure 6: The action recognition vs identity recognition accuracy for HuMMan. The error bars represent the 95% confidence interval.

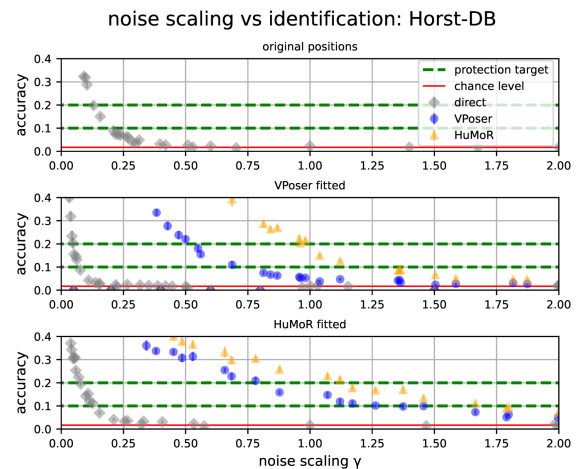


Figure 8: The scaling of the noise parameter γ in comparison to the achieved recognition accuracy of different anonymization combinations for Horst-DB. The error bars represent the 95% confidence interval.

the anonymizations on the adjusted data, we see that the adjusted data drops faster in person identification than in action recognition. For example (see Figure 5), at 60% action recognition accuracy, the original data is still at about 15% identification accuracy, while the fitted data is much lower and close to 5% accuracy. Surprisingly, there does not seem to be a difference if the anonymization on the fitted data is performed directly on the positions, the SMPL joints, or the latent encoding, as all these anonymizations perform similarly.

For the HuMMan action vs. identification see Figure 6. Compared to the CeTI-LoComotion dataset, the HuMMan dataset has much lower identification and action recognition rates. Since the baseline action recognition rate of 47% is below the 74% reported for RGB

videos [4] and the identification rate is only 26%, we conclude that the SMPL data fitted to the RGB videos is not optimal. Using HuMoR as the foundation motion model for anonymization further decreases identification to 11% and action recognition to 20% – even without applying any noise. This shows that preserving privacy halves the utility of data. At the same time, action recognition is still significantly above chance level (4%) showing the general viability of Pantomime’s design.

For the main privacy-utility tradeoff, we first report the scaling of the noise parameter versus the achieved recognition accuracy, see Figure 7 and Figure 8. We use these results to pick the noise

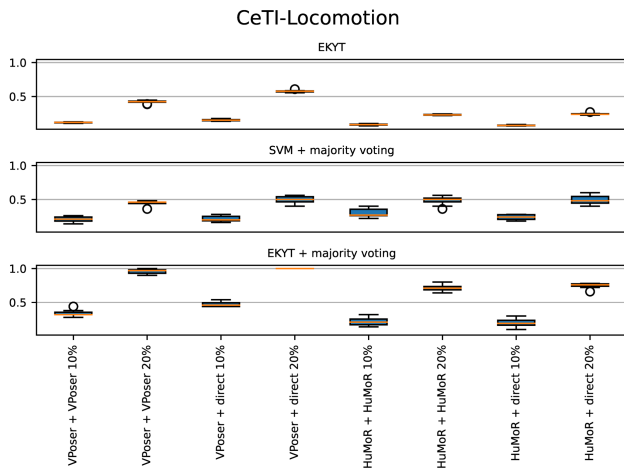


Figure 9: Average identification accuracy for the EKYT model and majority voting for SVM and EKYT for CeTI-Loconation.

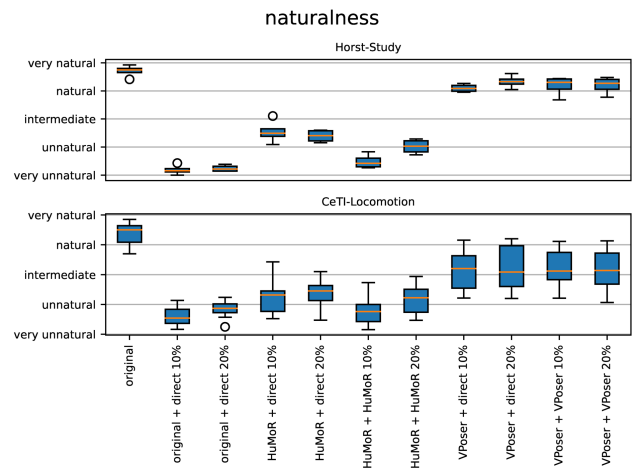


Figure 11: Average naturalness rating per question of the user study as a box plot per anonymization.

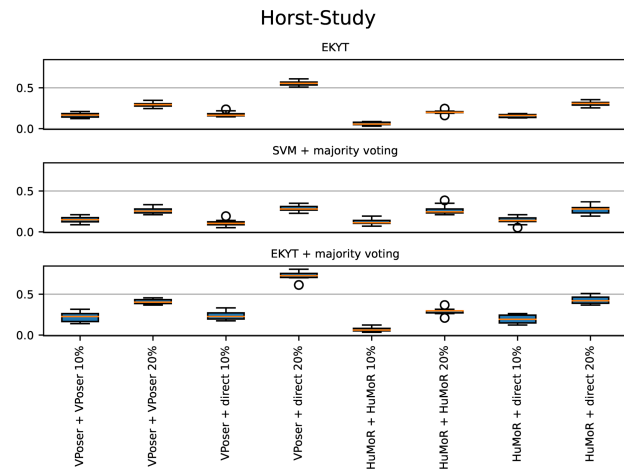


Figure 10: Average identification accuracy for the EKYT model and majority voting for SVM and EKYT for Horst-Study.

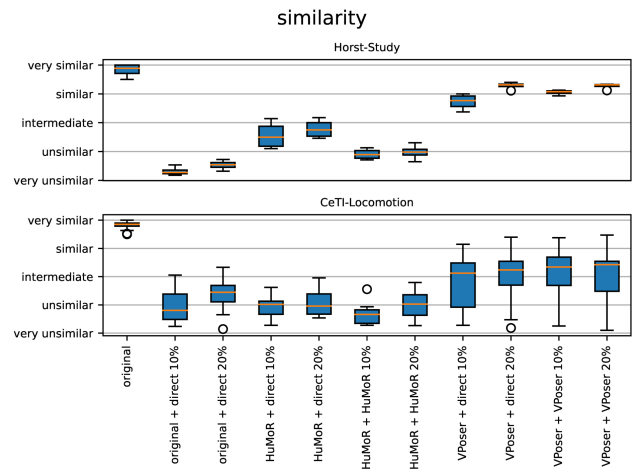


Figure 12: Average similarity rating per question of the user study as a box plot per anonymization.

scaling values which fulfill the specific protection targets (shown as green line). We find that the direct application of noise to the position data requires less noise scaling than when we apply the noise in the latent space or directly to the joint rotations of the SMPL model. For both datasets, we choose the anonymizations that achieve 10% or 20% ($\pm 2\%$) for the direct comparison of utility in the user study.

We now present the results of Experiment E6 (see Figure 9 and 10). Using EKYT as an extra recognition model on the anonymized data for both the CeTI-Loconation and the Horst Study shows that, for some of the 20% targets, identification increases. The most drastic case is VPoser + Direct, which climbs to over 50% recognition accuracy. These increases are exacerbated in majority voting, where all test samples of a subject are used to determine the most common

subject label. When the base accuracy is high, the majority vote can reach 100% accuracy in classifying the correct subject. However, when the recognition target is selected at 10%, this effect can be mitigated. For example, VPoser + VPoser and HuMoR + HuMoR only slightly increase their identification rates when majority voting is used. Overall the experiment shows that selecting a low enough recognition target is important, especially when an attacker that obtains a large set of motion sequences per person.

For our user study, we first calculate the average of the ratings for each question, and then use the averages to create a box plot for each anonymization technique. Comparing the naturalness results (see Figure 11) we see that the original motion sequences score the highest for both datasets, indicating that the participants consider the selected datasets to be representative of natural motion. For the direct addition of noise to the original position data, we see that both

datasets have very low naturalness ratings, close to very unnatural, showing that the naturalness of the motion data is destroyed. The anonymizations using HuMoR as a foundation model achieve more utility than the direct anonymization, but also all are rated close to unnatural. The best natural ratings we see for the anonymization using VPoser, for the CeTI-Locomotion dataset an intermediate rating is achieved, while for the Horst-DB the rating is even above natural. We conclude that using VPoser as a foundation model during anonymization helps to achieve natural motion sequences that are close to the naturalness of the original.

For the similarity results (see Figure 12), we generally see a similar pattern as for naturalness, with the original data rated as very similar to itself. Direct anonymization on the original data achieves almost no similarity, HuMoR achieves intermediate to unsimilar results, and VPoser again delivers the best similarity on the Horst-DB datasets. This again shows that using VPoser, Pantomime can successfully anonymize while keeping the motion sequence similar to the original, thus preserving utility.

7 Discussion, Limitations & Future Work

Our results demonstrate that Pantomime can successfully anonymize motion data by doing so in the latent space of a foundation motion model. Furthermore, we demonstrate that applying a fixed random vector to the entire motion sequence rather than varying it for each pose is a more effective method of anonymizing motion sequences, and that using a VPoser fitting with VPoser latent encoding achieves the best privacy-utility tradeoff.

In general, we find that Pantomime’s approach of using foundation motion models to first fit the position data to the SMPL body model and then anonymize the data in the latent space of the model is a viable approach to anonymize full-body motion data in a plausible manner. Because of the plausibility constraints added by the foundation models, the data can retain the utility of the motion data while performing effective anonymization.

In comparison with prior work, Pantomime has some key advantages. It anonymizes full-body motion capture data, it is not designed to work only with specific motions (such as gait), it does not require a large corpus of specific application data to train, and it is configurable via its noise scaling, allowing its privacy-utility tradeoff to be adjusted for specific applications. Furthermore, it is general as it performs a unification of motion data formats by fitting to the SMPL body model.

The same foundation models that we use to anonymize can also be used to generate synthetic motion data, by using the original data as an anchor in the latent space and then shifting it by adding noise, we essentially generate synthetic data that is similar to the original data. Thus, Pantomime can be considered as synthetic data generation. Similar to face anonymization [15], which generates new faces to anonymize facial images that have similar characteristics such as ethnicity or age.

Pantomime also has some limitations that need to be addressed. Its main drawback is the poor fitting quality for some of the motion sequences, especially when HuMoR is used for fitting. Another problem with fitting is that it does not work well for certain actions, such as the stand-to-sit tasks in CeTI-Locomotion. Furthermore, the current implementation of the fitting process is slow (it takes

about 1.5 weeks to process CeTI-Locomotion on a single GeForce RTX 3090). However, the anonymization process itself is much faster. It takes only about 16 minutes to anonymize the entire CeTI-Locomotion dataset, which contains about 1.5 hours of recorded sequences.

Pantomime’s approach to anonymization is very general and could be promising for other complex data types that are difficult to anonymize, such as trajectories or other biometric features. The only requirement would be that the foundation models use a VAE structure. Although our experimental evaluation shows that Pantomime offers a reasonable tradeoff between privacy and utility, future research could focus on formalizing this guarantee to understand the theoretical bounds that govern it. Furthermore, it should be investigated how the approach can be realized in a faster fashion to allow for applications like the real-time streaming of motion data.

8 Conclusion

We have presented Pantomime, a full-body motion anonymization that uses foundation motion models to anonymize in a plausible way. Our results show that using the VPoser model it is possible to achieve identification rates as low as 10% while keeping the anonymized motions natural and similar to the original ones. This is an important step towards a more privacy-preserving use of motion tracking in applications such as MR, robotics, or medicine.

Acknowledgments

The authors used generative AI-based tools to revise the text, improve flow and correct any typos, grammatical errors, and awkward phrasing. Funded by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) as part of Germany’s Excellence Strategy – EXC 2050/1 – Project ID 390696704 – Cluster of Excellence “Centre for Tactile Internet with Human-in-the-Loop” (CeTI) of Technische Universität Dresden; This work was funded by the Topic Engineering Secure Systems of the Helmholtz Association (HGF) and supported by KASTEL Security Research Labs, Karlsruhe.

References

- [1] Enas Abdulhay, N Arunkumar, Kumaravelu Narasimhan, Elamaran Vellaippan, and V Venkatraman. 2018. Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease. *Future Generation Computer Systems* 83 (2018), 366–373. <https://doi.org/10.1016/j.future.2018.02.009>
- [2] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. 2021. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258* (2021).
- [3] Christopher P Burgess, Irina Higgins, Arka Pal, Loic Matthey, Nick Watters, Guillaume Desjardins, and Alexander Lerchner. 2018. Understanding disentangling in β -VAE. *arXiv preprint arXiv:1804.03599* (2018).
- [4] Zhongang Cai, Daxuan Ren, Ailing Zeng, Zhengyu Lin, Tao Yu, Wenjia Wang, Xiangyu Fan, Yang Gao, Yifan Yu, Liang Pan, et al. 2022. Humman: Multi-modal 4d human dataset for versatile sensing and modeling. In *European Conference on Computer Vision*. Springer, 557–577.
- [5] Kouros Darvish, Luigi Penco, Joao Ramos, Rafael Cisneros, Jerry Pratt, Eiichi Yoshida, Serena Ivaldi, and Daniele Pucci. 2023. Teleoperation of Humanoid Robots: A Survey. *IEEE Transactions on Robotics* 39, 3 (2023), 1706–1727. <https://doi.org/10.1109/TRO.2023.3236952>
- [6] Marc H. E. de Lussanet. 2018. *Point Light Displays*. Springer International Publishing, Cham, 1–4. https://doi.org/10.1007/978-3-319-47829-6_663-1
- [7] Shangchen Han, Beibei Liu, Randi Cabezas, Christopher D Twigg, Peizhao Zhang, Jeff Petkau, Tsz-Ho Yu, Chun-Jung Tai, Muzaffer Akbay, Zheng Wang, et al. 2020. MEGATrack: monochrome egocentric articulated hand-tracking for virtual reality.

- ACM Transactions on Graphics (ToG)* 39, 4 (2020), 87–1. <https://doi.org/10.1145/3386569.3392452>
- [8] Simon Hanisch, Evelyn Muschter, Admantini Hatzipanayioti, Shu-Chen Li, and Thorsten Strufe. 2023. Understanding Person Identification Through Gait. *Proc. Priv. Enhancing Technol.* 2023, 1 (2023), 177–189. <https://doi.org/10.56553/POPETS-2023-0011>
- [9] Simon Hanisch, Loreen Pogrzeba, Evelyn Muschter, Shu-Chen Li, and Thorsten Strufe. 2024. A kinematic dataset of locomotion with gait and sit-to-stand movements of young adults. *Scientific Data* 11, 1 (2024), 1209. <https://doi.org/10.1038/s41597-024-04020-6>
- [10] Simon Hanisch, Julian Todt, Jose Patino, Nicholas Evans, and Thorsten Strufe. 2024. A False Sense of Privacy: Towards a Reliable Evaluation Methodology for the Anonymization of Biometric Data. *Proc. Priv. Enhancing Technol.* 2024, 1 (2024), 116–132. <https://doi.org/10.56553/popets-2024-0008>
- [11] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. 2017. beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. In *International Conference on Learning Representations*. <https://openreview.net/forum?id=Sy2fzU9gl>
- [12] Yuki Hirose, Kazuaki Nakamura, Naoko Nitta, and Noboru Babaguchi. 2019. Anonymization of Gait Silhouette Video by Perturbing Its Phase and Shape Components. In *2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. 1679–1685. <https://doi.org/10.1109/APSIPAASC47483.2019.9023196>
- [13] Yuki Hirose, Kazuaki Nakamura, Naoko Nitta, and Noboru Babaguchi. 2022. Anonymization of Human Gait in Video Based on Silhouette Deformation and Texture Transfer. *IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY* 17 (2022), 3375–3390. <https://doi.org/10.1109/tifs.2022.3206422>
- [14] Fabian Horst, Sebastian Lapuschkin, Wojciech Samek, Klaus-Robert Müller, and Wolfgang I. Schöllhorn. 2019. Explaining the unique nature of individual gait patterns with deep learning. *Scientific reports* 9, 1 (2019), 2391. <https://doi.org/10.1038/s41598-019-38748-8>
- [15] Haakon Hukkelaa, Rudolf Mester, and Frank Lindseth. 2019. Deepprivacy: A generative adversarial network for face anonymization. In *International symposium on visual computing*. Springer, 565–578.
- [16] Diederik P Kingma and Max Welling. 2022. Auto-Encoding Variational Bayes. arXiv:1312.6114 [stat.ML] <https://arxiv.org/abs/1312.6114>
- [17] Mark A Kramer. 1991. Nonlinear principal component analysis using autoassociative neural networks. *AICHe Journal* 37, 2 (1991), 233–243. <https://doi.org/10.1002/aic.690370209>
- [18] S. Kullback and R. A. Leibler. 1951. On Information and Sufficiency. *The Annals of Mathematical Statistics* 22, 1 (1951), 79–86. <https://doi.org/10.1214/aoms/1177729694>
- [19] Don S Lemons and Paul Langevin. 2002. *An introduction to stochastic processes in physics*. JHU Press.
- [20] Dillon Lohr and Oleg V. Komogortsev. 2022. Eye Know You Too: Toward Viable End-to-End Eye Movement Biometrics for User Authentication. *IEEE Transactions on Information Forensics and Security* 17 (2022), 3151–3164. <https://doi.org/10.1109/TIFS.2022.3201369>
- [21] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2015. *SMPL: A Skinned Multi-Person Linear Model*. Vol. 34. ACM, 851–866. <https://doi.org/10.1145/3596711.3596800>
- [22] Shugao Ma, Tomas Simon, Jason Saragih, Dawei Wang, Yuecheng Li, Fernando De La Torre, and Yaser Sheikh. 2021. Pixel codec avatars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 64–73.
- [23] Naureen Mahmood, Nima Ghorbani, Nikolaus F. Troje, Gerard Pons-Moll, and Michael J. Black. 2019. AMASS: Archive of Motion Capture as Surface Shapes. In *International Conference on Computer Vision*. 5442–5451. <https://doi.org/10.1109/iccv.2019.00554>
- [24] Matthew Malek-Podjaski and Fani Deligianni. 2021. Towards Explainable, Privacy-Preserved Human-Motion Affect Recognition. In *2021 IEEE Symposium Series on Computational Intelligence (SSCI) (2021-12)*. IEEE, 01–09. <https://doi.org/10.1109/SSCI50451.2021.9660129>
- [25] Yan Meng, Yuxia Zhan, Jiachun Li, Suguo Du, Haojin Zhu, and Xuemin Shen. 2024. De-Anonymizing Avatars in Virtual Reality: Attacks and Countermeasures. *IEEE Transactions on Mobile Computing* 23, 12 (2024), 13342–13357. <https://doi.org/10.1109/TMC.2024.3426046>
- [26] Mark Roman Miller, Fernanda Herrera, Hanseul Jun, James A Landay, and Jeremy N Bailenson. 2020. Personal identifiability of user tracking data during observation of 360-degree VR video. *Scientific Reports* 10, 1 (2020), 17404. <https://doi.org/10.1038/s41598-020-74486-y>
- [27] Saemi Moon, Myeonghyeon Kim, Zhenyue Qin, Yang Liu, and Dongwoo Kim. 2023. Anonymization for skeleton action recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 15028–15036.
- [28] Alec G. Moore, Ryan P. McMahan, Hailiang Dong, and Nicholas Ruozzi. 2021. Personal Identifiability and Obfuscation of User Tracking Data From VR Training Sessions. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 221–228. <https://doi.org/10.1109/ISMAR52148.2021.00037>
- [29] Vivek Nair, Wenbo Guo, James F. O'Brien, Louis Rosenberg, and Dawn Song. 2024. Deep Motion Masking for Secure, Usable, and Scalable Real-Time Anonymization of Ecological Virtual Reality Motion Data. In *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 493–500. <https://doi.org/10.1109/VRW62533.2024.00096>
- [30] Vivek C Nair, Gonzalo Munilla-Garrido, and Dawn Song. 2023. Going Incognito in the Metaverse: Achieving Theoretically Optimal Privacy-Usability Tradeoffs in VR. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (San Francisco, CA, USA) (UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 16, 16 pages. <https://doi.org/10.1145/3586183.3606754>
- [31] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 8024–8035. <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
- [32] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black. 2019. Expressive Body Capture: 3D Hands, Face, and Body from a Single Image. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*.
- [33] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, M. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
- [34] Jacquelin Perry and Judith M Burnfield. 2010. Gait analysis. Normal and pathological function 2nd ed. *California: Slack* (2010).
- [35] Robert Prill, Marina Walter, Aleksandra Królikowska, and Roland Becker. 2021. A Systematic Review of Diagnostic Accuracy and Clinical Applications of Wearable Movement Sensors for Knee Joint Rehabilitation. *Sensors* 21, 24 (Dec. 2021), 8221. <https://doi.org/10.3390/s21248221>
- [36] Davis Remppe, Tolga Birdal, Aaron Hertzmann, Jimei Yang, Srinath Sridhar, and Leonidas J. Guibas. 2021. HuMoR: 3D Human Motion Model for Robust Pose Estimation. In *International Conference on Computer Vision (ICCV)*.
- [37] Diana Romero, Ruchi Jagdish Patel, Athina Markopoulou, and Salma Elmalaki. 2024. GaitGuard: Towards Private Gait in Mixed Reality. arXiv:2312.04470 [cs.HC] <https://arxiv.org/abs/2312.04470>
- [38] Hosnieh Sattar, Katharina Krombholz, Gerard Pons-Moll, and Mario Fritz. 2020. Body shape privacy in images: understanding privacy and preventing automatic shape extraction. In *Computer Vision—ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*. Springer, 411–428.
- [39] Kihyuk Sohn, Honglak Lee, and Xinchen Yan. 2015. Learning structured output representation using deep conditional generative models. *Advances in neural information processing systems* 28 (2015).
- [40] Omer Terlemez, Stefan Ulbrich, Christian Mandery, Martin Do, Nikolaus Vahrenkamp, and Tamim Asfour. 2014. Master Motor Map (MMM)—Framework and toolkit for capturing, representing, and reproducing human motion on humanoid robots. In *2014 IEEE-RAS International Conference on Humanoid Robots*. IEEE, 894–901.
- [41] Daksh Thapar, Aditya Nigam, and Chetan Arora. 2021. Anonymizing Egocentric Videos. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 2320–2329.
- [42] Ngoc-Dung T. Tieu, Huy H. Nguyen, Hoang-Quoc Nguyen-Son, Junichi Yamagishi, and Isao Echizen. 2017. An approach for gait anonymization using deep learning. In *2017 IEEE Workshop on Information Forensics and Security (WIFS)*. 1–6. <https://doi.org/10.1109/WIFS.2017.8267657>
- [43] Julian Todt, Simon Hanisch, and Thorsten Strufe. 2024. Fantómas: Understanding Face Anonymization Reversibility. *Proc. Priv. Enhancing Technol.* 2024, 4 (2024), 24–43. <https://doi.org/10.56553/POPETS-2024-0105>
- [44] Changsheng Wan, Li Wang, and Vir V. Phoha. 2018. A Survey on Gait Recognition. *ACM Comput. Surv.* 51, 5 (2018), 1–35. <https://doi.org/10.1145/3230633>