applied sciences

MDPI

Article

Synthesizing 3D Gait Data with Personalized Walking Style and Appearance

Yao Cheng [1], Guichao Zhang [1], Sifei Huang [1], Zexi Wang [2], Xuan Cheng [2] and Juncong Lin [2,*]

[1] China Mobile (Hangzhou) Information Technology Co., Ltd., Hangzhou 311121, China
[2] School of Informatics, Xiamen University, Xiamen 361005, China
* Correspondence: jclin@xmu.edu.cn

Abstract: Extracting gait biometrics from videos has been receiving rocketing attention given its applications, such as person re-identification. Although deep learning arises as a promising solution to improve the accuracy of most gait recognition algorithms, the lack of enough training data becomes a bottleneck. One of the solutions to address data deficiency is to generate synthetic data. However, gait data synthesis is particularly challenging as the inter-subject and intra-subject variations of walking style need to be carefully balanced. In this paper, we propose a complete 3D framework to synthesize unlimited, realistic, and diverse motion data. In addition to walking speed and lighting conditions, we emphasize two key factors: 3D gait motion style and character appearance. Benefiting from its 3D nature, our system can provide various gait-related data, such as accelerometer data and depth map, not limited to silhouettes. We conducted various experiments using the off-the-shelf gait recognition algorithm and draw the following conclusions: (1) the real-to-virtual gap can be closed when adding a small portion of real-world data to a synthetically trained recognizer; (2) the amount of real training data needed to train competitive gait recognition systems can be reduced significantly; (3) the rich variations in gait data are helpful for investigating algorithm performance under different conditions. The synthetic data generator, as well as all experiments, will be made publicly available.

Keywords: 3D gait data; human motion data; neural network; gait recognition; gait synthesis

Citation: Cheng, Y.; Zhang, C.; Huang, S.; Wang, Z.; Cheng, X.; Lin, J. Synthesizing 3D Gait Data with Personalized Walking Style and Appearance. *Appl. Sci.* **2023**, *13*, 2084. https://doi.org/10.3390/app13042084

Academic Editor: Luis Javier Garcia Villalba

Received: 27 December 2022
Revised: 29 January 2023
Accepted: 3 February 2023
Published: 6 February 2023

1. Introduction

Human gait is one of the most complex phenomena combining efforts of the neural, muscle, and skeletal systems, with mutual interaction from the external environment. It contains abundant information reflecting some unique features of a person, and therefore has great potential in various applications, especially human recognition. The discrimination of human gait has been intensively investigated in the research of biomechanics, physical medicine, and psychological studies [1].

Compared with other biometric identification methods (such as fingerprints, face, or iris biometric modalities), vision-based gait recognition does not require the cooperation or even awareness of the individual under observation and therefore has been attracting increasing attention [2]. However, accurate recognition of human gait is still challenging due to (1) the inconspicuous **inter-class** differences between different people such as walking style and health condition; (2) the significant **intra-class** variations from the same person in dynamically-changing scenarios such as different walking speeds, viewpoints, clothing, and belongings; and (3) **complex surveillance environments** such as cluttered environments, illumination changes, partial occlusions, and crowds of people.

Recently, the popularity of deep neural networks stimulated its usage in gait recognition [3–5]. Adequate gait data are the prerequisite of many gait recognition algorithms, especially for mainstream supervised methods, and the size of the gait dataset should be sufficiently large in order to reflect the variety of factors and their combinations. Existing works on gait dataset construction are still far from this demanding goal. Taking

the representative OU-ISIR Gait Database [6] as an example, although declared to be the largest dataset in the world (including more than 60,000 subjects), data were collected under a fixed viewpoint and environment. A potential solution for these problems is using highly realistic synthetic data generated from virtual worlds [7–9]. We can automatically annotate synthetic data for training/validation and conveniently collect samples in extreme or challenging circumstances (such as occlusion and poor illumination). Even more important, synthetic data allow us to pre-train the surveillance system so that it can be used immediately without data collection and training. However, it is non-trivial to synthesize data for gait analysis. The major challenge is that we need to precisely control the variation of synthesized gait style to balance inter-subject and intra-subject differences.

In this paper, we present a framework to explore the gait space from various aspects (including camera, illumination, clothes, motion style, etc.). Our method has two major advantages: (1) our method can generate an unlimited number of gait data with arbitrary combinations of influential factors; (2) we can provide accurate metrics for the purpose of gait analysis, including gait energy image, inertial data, ground force, etc. Figure 1 shows the overview of the proposed framework. This work is, to the best of our knowledge, the first one to simultaneously consider the variations of both motion style and character appearance. In particular, we emphasize the synthesis of personalized yet variant gait sequences, which is rarely discussed in existing motion synthesis and editing works. The technical contributions of this paper include:

- We propose a complete framework to explore gait data generation for the purpose of synthesizing unlimited, realistic, and diverse motion data. These data could serve as the training database for learning-based methods in gait recognition, etc.
- We develop a semantic motion style exploration method that controls the motion style via a deep neural network and interpolates motion sequences of different lengths.
- We introduce an appearance parameterization model capable of generating virtual characters with different weight, height, girth, clothing, etc.

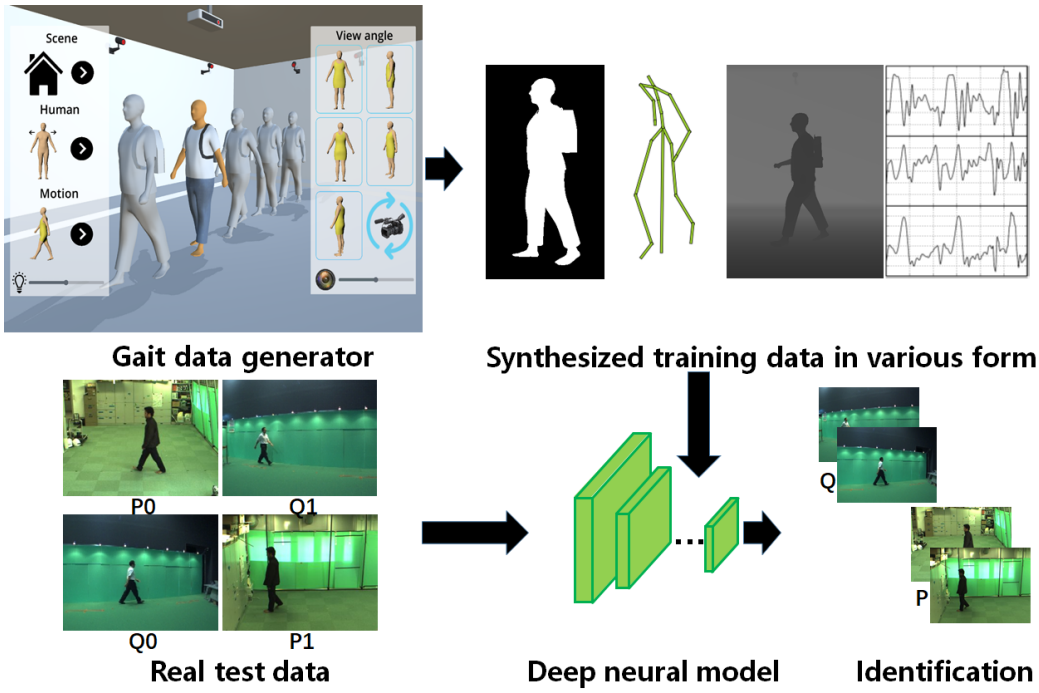


Figure 1. Synthesizing gait data for learning-based gait analysis.

2. Related Work

2.1. Gait Recognition

Gait recognition methods can be roughly categorized as model-based and appearance-based [2]. As the name implies, model-based approaches [10] directly extract human body structure from the images. The main difficulty with these methods is the precise estimation of model parameters from image sequences, which requires a high image resolution with a significant computational burden. Appearance-based methods [3,4,11–13] mainly focus on extracting gait features from captured image sequences regardless of the underlying structure. Therefore, this approach can perform recognition at lower resolutions, which makes them suitable for outdoor applications when the parameters of the body structure are difficult to estimate precisely. Nonetheless, due to the human-crafted gait features, it is extremely hard for existing methods to break through feature representation bottlenecks when faced with the gait and appearance changes of a walking person with massive differences in walking speed, viewpoint, clothing, and object carrying. A joint learning strategy is adopted by some researchers to address these limitations. For example, Makihara et al. [14] jointly learned intensity and spatial metrics under a unified framework and alternately optimized it by linear or ranking support vector machines, while Zhang et al. [15] combined two complementary representations, unique gait and cross gait, to boost the performance of gait recognition. Different from these joint learning methods, Zhang et al. [16] took an opposite strategy to disentangle pose and appearance features from RGB imagery explicitly and generated gait features automatically through an LSTM-based integration of pose features over time. Our purpose is to construct an effective and powerful gait data generator with the synthetic paradigm, serving the various requirements in training and validation of these gait analysis models.

2.2. Gait Data Collection

A common problem encountered in gait recognition is the lack of enough gait data. To address the problem, researchers [6,17,18] constructed various gait database with different emphases. The SOTON database [19], mainly for gait recognition, contains hundreds of subjects with limited covariates such as views, shoes, clothing, and walking speed. The USF dataset [17], as one of the most frequently-used gait recognition dataset, is composed of 122 subjects only. The CASIA database [18] considers additional factors, such as night scenario, multi-views, different clothing, and carrying status, but contains a small number of 124 subjects. The OU-ISIR database [6], targeting age estimation, contains 60,000 subjects covering different ages. A common issue with these databases is that they are still too small to contain enough variation in those involved factors. Such variations are essential aspects of the statistically reliable evaluation in many gait analysis applications. A recent work [20], described as so far the world's largest gait database, mainly targeted performance evaluation of gait-based age estimation. There are also some general performance capture datasets [21] being used in gait motion synthesis. Some works [22] synthesized new data by distortion or blending of an existing one. However, such methods can only mimic limited and low-quality variations in viewpoint. Our method can also be considered a synthetic method but in a more intrinsic way. In addition to synthesis from images, we build a complete 3D framework to simulate factors involved in gait analysis such as environments, anthropometric measures, outfit clothing, and motion styles.

Using synthetic data for model training is a recently popular strategy to address the lack of sufficient data in learning-based methods such as object tracking [7], optic flow [23], scene understanding [24,25], pedestrian detection [26], human information estimation [27], action recognition [28], etc. However, the *reality gap* prevents the strategy from becoming practically useful: a model trained in a simulated environment may not be applicable to real-world scenarios due to numerous discrepancies between the two environments. Two solutions have been proposed to bridge the gap: extreme realism [7,8] and domain randomization [29,30]. The former attempts to improve the simulation similarity with the real-world environment, while the latter aims for exposing the model to a vast range of

simulated environments at the training stage instead of just a single synthetic one. Our work shares inspirations with [31] on the necessity to balance intra-subject and inter-subject variations. While [31] focused on face synthesis, which is different from gait, and the dynamic and subtle nature of gait increases the difficulty.

2.3. Identity Aware Data Synthesis

Learning disentangled representations has been a fundamental problem since the rise of machine learning. Many researchers are especially interested in extraction of identity attributes due to its vast involvement in various applications. Some researchers investigated the problem of identity presentation using information such as identity labels [32] and identity features [33], but this suffered from incomplete identity maintenance. FaceID-GAN [34] adopted a three-player competition architecture where the generator competes with a discriminator and an identity classifier for quality and identity preservation, respectively. However, the quality of the synthesized image is still unsatisfactory. In their followup, FaceFeat-GAN [35], a two-stage strategy is proposed, with the first stage to synthesize facial features, while the second stage is to render high-quality images. The work of Bao et al. [36] allows synthesis of faces with ID outside the training set, and no annotations of attributes are required. In [37], a novel disentanglement method with minimal supervision is presented and applied to the human head to separate identity from other facial attributes. The work of [38] deals with multimodal information, performing identity-aware textual–visual matching with latent co-attention. Identity awareness also attracted a lot of attention from 3D graphics. In [39], Zhou et al. investigated the generation of expressive talking heads from a single facial image with audio as the only input, with content and speaker information disentangled from the input audio signal. With the incorporation of multimodal context and an adversarial training scheme, Yoon et al. [40] proposed a method to generate human-like gestures that match in speech content and rhythm.

3. Diverse and Personalized Walking Motion Synthesis

Generating diverse yet personalized 3D gait data is the core of the whole paradigm. Our basic idea is to construct a motion space from a small set of motion data and then sample in the space to generate unlimited motion data. To avoid being constrained in a small interpolation space, the set of motion data consists of various motion types such as walking, running, jumping, etc. However, two problems need to be addressed:

- Given synthesized motion data, we should be able to determine whether it is a gait motion or not.
- As mentioned before, we need to balance between inter-subject and intra-subject differences carefully.

Our solution is illustrated in Figure 2: we feed a large motion data set into a deep neural network and learn an effective feature representation that is encoded in the final hidden unit layer; we then extract the motion space for synthesis from the feature representation and sample in the space to generate new data. To validate whether the synthesized data are gait or not, we perform *content embedding* on the learned high-dimensional feature vector according to a distance metric related to their content. To determine whether the synthesized gait belongs to a person or not, we introduce *ID embedding* to cluster gait data from the same person together. Our method investigated various combinations of feature vector and distance metrics to find the optimal ones for the two embeddings.

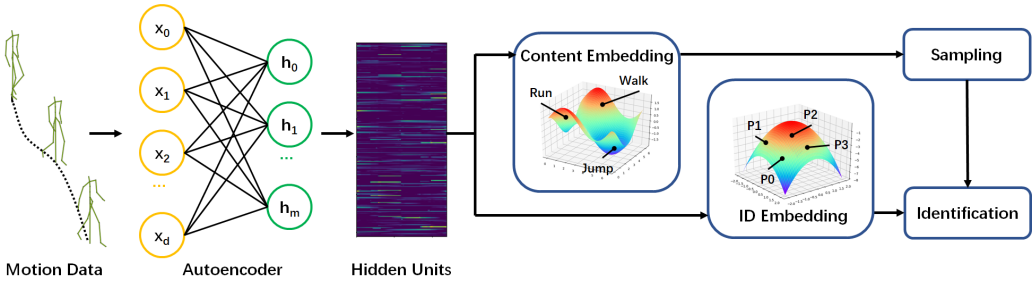


Figure 2. Flowchart of our gait synthesis framework.

3.1. Content Embedding

Feature learning. We learn feature representation from a large training set consisting of several publicly available sources [41–43] with a convolutional auto-encoder [44]. Contrary to the general motion editing method, we focus on synthesizing valid and personalized gait motion. Only one layer is used to encode the motion, as multiple layers of pooling/de-pooling can result in blurred motion after reconstruction due to the pooling layers of the network reducing the temporal resolution of the data. An additional feed-forward network will later be used to reconstruct the motion to allieviate the lack of feature abstraction due to the use of one layer only (for details please refer to [44]).

Data from different sources are retargeted to a uniform skeleton structure (with the same hierarchy and link lengths) and converted from the joint angle representation into a joint position representation. Absolute coordinates in both Euler space and coordinates relative to skeleton root are investigated. The auto-encoder performs a one-dimensional convolution over the temporal domain, independently for each filter. The network provides a forward operation $\mathbf{\Phi}$, which receives the input vector X in the visible unit space and outputs the encoded values H in the hidden unit space:

$$\mathbf{\Phi}(X) = ReLU(\Psi(\mathbf{X} * \mathbf{W}_0 + \mathbf{b}_0)) \tag{1}$$

The backward operation $\mathbf{\Phi}^{-1}$ does a reverse task to go back from the hidden unit space to the visible unit space:

$$\mathbf{\Phi}^{-1}(H) = (\mathbf{\Psi}^{-1}(H) - b_0) * \tilde{W}_0, \tag{2}$$

where $*$ is the convolution operation in both Equations (1) and (2).

Embedding. We investigate the clustering effects of different metric definitions (l_1-norm, l_2-norm, squared Euclidean, cosine distance, and Mahalanobis distance) on either the original motion vector $\mathbf{o}_i$ input to the auto-encoder or the hidden unit vector $\mathbf{h}_i$ of the auto-encoder. The Mahalanobis distance is defined for the hidden unit vector:

$$d_c^2(\mathbf{h}_i, \mathbf{h}_j) = (\mathbf{h}_i - \mathbf{h}_j)^T \mathbf{M}_c (\mathbf{h}_i - \mathbf{h}_j), \tag{3}$$

where c denotes the motion type, such as walking, running, jumping, etc., and $\mathbf{M}_c$ denotes the Mahalanobis matrix learned with a pair learner [45]. Hidden unit vectors are paired and labeled with the ground truth of a specific motion type.

3.2. ID Embedding

The key to ID embedding is to uniquely identify different personal styles. Although the Gram matrix encodes the style of the data [44], the style here is a more general concept relating to the status of a person (such as injury, depression), not the unique personal style that consistently arises during different walking cycles, even under different statuses.

The key point is that we need to differentiate everyone clearly, while the different gait sequences of the same person should cluster together. Thus, it is quite predictable that simply using l_1-norm, l_2-norm, or even the Bhattacharyya distance defined on the Gram

matrix will not work well. Similarly, we investigate the performance of different metrics and find the Mahalanobis matrix $\mathbf{M}_g$, defined by reshaping the Gram matrix $\mathbf{g}$ into a vector,

$$d_i^2(\mathbf{g}_i, \mathbf{g}_j) = (\mathbf{g}_i - \mathbf{g}_j)^T \mathbf{M}_g (\mathbf{g}_i - \mathbf{g}_j), \tag{4}$$

serves the purpose best as shown in the experiments.

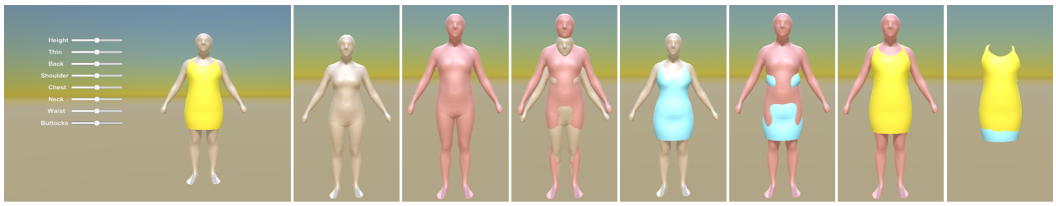
4. Consistent Subject Appearance Variation

4.1. Parametric Body Adjustment

The user is allowed to adjust the shape of the subject's body through semantic attributes like weight, height, girth, etc. (Figure 3a). We follow previous example-based strategy [46,47]. A morphable human shape model Q is constructed by applying principal component analysis (PCA) to a database of registered dense human mesh models with consistent connectivity and correspondence. (Here, we use the publicly available database provided by [46]). Linear regression is then used to learn the mapping $f : P \to Q$ from semantic attributes P to the morphable model Q (described by an s-dimensional vector):

$$\mathbf{q} = f(\mathbf{p}) = \mathbf{T} \cdot \mathbf{p} + \mathbf{r}, \tag{5}$$

where $\mathbf{T}_{s \times t}$ is the relation matrix, $\mathbf{p} = [p_1, p_2, \ldots, p_t]$ is a vector of the t semantic attributes, $\mathbf{q} = [q_1, q_2, \ldots, q_s]$ is a point in the morphable model space Q, and vector $\mathbf{r}_{t \times 1}$ is a corresponding residual. Both $\mathbf{T}$ and $\mathbf{r}$ can be solved through a least-squares solution by substituting related information from exemplary models into Equation (5).



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 3. Consistent subject appearance exploration: (**a**) interface for semantic human shape exploration: (**b**) source human model; (**c**) target human model; (**d**) overlap of source and target human model; (**e**) source human model with source cloth; (**f**) target human model with source cloth; (**g**) target human model with adapted cloth; (**h**) overlap of source and adapted cloth.

4.2. Clothing Adaptation

We initially put clothes and other accessories on a neutral human model. When a new human shape model with expected semantic attributes is generated, we also need to transfer clothes and accessories onto the new model. We design a novel surface-based deformation method to conduct the transfer. Let H^0 (with vertices $\{\mathbf{v}_0^0, \mathbf{v}_1^0, \ldots, \mathbf{v}_n^0\}$) be the neutral human mesh model, A^0 (with vertices $\{\mathbf{u}_0^0, \mathbf{u}_1^0, \ldots, \mathbf{u}_m^0\}$) be the mesh model of accessories (mainly clothes) attached to H^0, and H^1 (with vertices $\{\mathbf{v}_0^1, \mathbf{v}_1^1, \ldots, \mathbf{v}_n^1\}$) be the human mesh model with the expected semantic attributes. Our purpose is to determine the vertex coordinates $\mathbf{u}_i^0$ on A^1, the mesh model of accessories attached onto H^1. Our basic idea is to firstly estimate a transformation $\mathbf{R}_j$ from $\mathbf{v}_j^0$ to $\mathbf{v}_j^1$. For each vertex $\mathbf{u}_i^0$, we find its k neighbors on H^0. For each neighbor $\mathbf{v}_{ij}^1$, we apply its transformation to $\mathbf{u}_i^0$ to get the adapted position with respect to $\mathbf{v}_{ij}^1$:

$$\mathbf{u}_{ij}^1 = \mathbf{R}_{ij} \mathbf{u}_i^0, \tag{6}$$

and the position of $\mathbf{u}_i^1$ can finally be calculated by weighted blending of all $\mathbf{u}_{ij}^1$:

$$\mathbf{u}_i^1 = w_0 \cdot \mathbf{u}_{i0}^1 \oplus w_1 \cdot \mathbf{u}_{i1}^1 \ldots \oplus w_j \cdot \mathbf{u}_{ij}^1 \ldots \tag{7}$$

where the weighting w_j is defined as the normalized distance between $\mathbf{u}_i^0$ and $\mathbf{v}_{ij}^0$:

$$w_j = \frac{\| \mathbf{u}_i^0 - \mathbf{v}_{ij}^0 \|^2 \, (\mathbf{n}_i^{0u} \cdot \mathbf{n}_{ij}^{0v})}{\sum_{k \in N(i)} \| \mathbf{u}_i^0 - \mathbf{v}_{ik}^0 \|^2}, \tag{8}$$

where $\mathbf{n}_i^{0u}$ and $\mathbf{n}_{ij}^{0v}$ are the normals of the corresponding surfaces on $\mathbf{u}_i^0$ and $\mathbf{v}_{ij}^0$, respectively.

The transformation is estimated in an as-rigid-as-possible manner [48] with a rotation matrix $\mathbf{R}_i$ from $\mathbf{v}_i^0$ to $\mathbf{v}_i^1$ calculated by minimizing the least squares deviation:

$$E(M_i^0, M_i^1) = \sum_{j \in O(i)} w_{ij} \| (\mathbf{v}_{ij}^1 - \mathbf{v}_i^1) - \mathbf{R}_i(\mathbf{v}_{ij}^0 - \mathbf{v}_i^0)) \|^2, \tag{9}$$

where M_i consists of $\mathbf{v}_i^0$ and its one-ring neighbors $O(i)$. The rotation matrix $\mathbf{R}_i$ can be derived from the singular value decomposition of the covariance matrix $\mathbf{S}_i$:

$$\mathbf{S}_i = \sum_{j \in O(i)} w_{ij} \mathbf{e}_{ij}^0 \mathbf{e}_{ij}^{1T}, \tag{10}$$

where $\mathbf{e}_{ij}^0 = \mathbf{v}_{ij}^0 - \mathbf{v}_i^0$ and $\mathbf{e}_{ij}^1 = \mathbf{v}_{ij}^1 - \mathbf{v}_i^1$. Let $\mathbf{S}_i = \mathbf{U}_i \sum_i \mathbf{V}_i^T$, then $\mathbf{R}_i = \mathbf{V}_i \mathbf{U}_i^T$.

For the blending operator $\oplus$, we have tried various methods including simple linear blending, dual quaternion blending, etc. Results are reported in Section 5.4.

4.3. Skeleton Update

When the human shape is altered, the corresponding skeleton also needs to be updated to ensure joint positions match the body well. For each joint J_i of the skeleton, we can define a cross-section and intersect with the human surface to get a cross contour $C_i = \{\mathbf{v}_{i0}, \mathbf{v}_{i1}, \ldots \mathbf{v}_{ik}\}$. We then encode the position $\mathbf{p}_i$ of J_i with the mean value coordinate $\mathbf{b} = \{b_1, b_2, \ldots, b_k\}$ related to C_i. The J_i' position will be updated through the mean value coordinate interpolation when the human shape is changed:

$$\mathbf{p}_i = \sum_{j=1}^{k} b_j \mathbf{v}_{ij}. \tag{11}$$

5. Experimental Results

5.1. Implementation Details

We have implemented the whole paradigm and developed a gait generator with Unity 3D. The data generator and related materials will be made publicly available to support the reproducibility of our results and any future developments. The program provides various functionalities to change the view perspective, scene types, human appearance, and lighting conditions. The program can also synthesize various motion patterns, including walking, running, etc. The program runs on a standard PC with 16 GB of memory, CPU i7-9700 3.6 Ghz, and GPU RTX 1080Ti. We here focus on three locomotion patterns: walking, running, and jumping, which are typically investigated by the state-of-the-art gait recognition methods. For each motion type, we extract the motion sequence as segments of one complete cycle and apply the technique of time warping to convert into sequences of consistent frames (N = 240). In total, our database contains 161, 36, and 23 segments for walking, running, and jumping, respectively.

5.2. Visual Demonstration of Factor Variations

We first demonstrate the ability of our method to change a variety of factors, which may significantly affect the accuracy and robustness of state-of-the-art recognition methods. The factors include:

View variation. As mentioned before, our paradigm can easily generate gait data under various camera positions (see Figure 4a).

Scene variations. Our gait generator also supports the changes in the scenes, from simple lab environment to a complex real street scenario (see Figure 4b).
Human appearance variation. The user can semantically control the shape of the human model by adjusting attributes like weight (Figure 4c, top). The user can also change the accessories and clothes on a person (Figure 4c, bottom).
Lighting variation. We can also simulate different lighting conditions. This is particularly common for surveillance cameras on the street (see Figure 4d).
Motion variations. Unlimited motion data of walking, running, and jumping can be synthesized from the PCA motion space.

In extreme cases, multiple factors may arise simultaneously. For example, a person may wear different clothing and walk in different scenes, which may introduce occlusions by the external environment (e.g., the trolley in Figure 4e).

Figure 4. Variation of factors: (**a**) viewpoint variation; (**b**) scene variation; (**c**) appearance variation; (**d**) lighting variation; (**e**) extreme cases.

5.3. Gait Style Analysis

We can classify the motion data according to content embedding statistics and identify a synthesized walking series according to ID embedding statistics (Table 1). We investigate the clustering performances on different feature vectors: (1) skeleton joint position vector consisting of absolute coordinates ($\mathbf{C}_a$) and relative coordinates ($\mathbf{C}_r$); (2) hidden unit vectors of auto-encoder corresponding to absolute ($\mathbf{H}_a$) and relative ($\mathbf{H}_r$) cases; (3) Gram matrix reshaped vectors corresponding to absolute ($\mathbf{G}_a$) and relative ($\mathbf{G}_r$) cases. We also test with different distance metrics including l_1-norm ($\mathbf{L}_1$), l_2-norm ($\mathbf{L}_2$), and the Mahalanobis distance ($\mathbf{L}_m$). If we regard the motion of the same type as one cluster, the average distance between all sample points reflects the cluster size. To better differentiate different motion types, we aim to maximize the distance between two clusters and define the following measure:

$$D = \frac{D_{I-J}}{D_I + D_J} \tag{12}$$

where D_I, D_J are the average distance of all samples in each cluster I, J, and D_{I-J} is the distance between the centroids of two clusters. We define the measures $D_C = \frac{D_{W-R}}{D_W+D_R}, D_P = \frac{D_{P1-P2}}{D_{P1}+D_{P2}}$ for the content and ID embedding in Table 1.

We here present two motion types (walking and running) and two persons. Experimental results show consistent findings for other motion types and more persons. The statistics in Table 1 reveal the following facts:

- The use of absolute coordinates consistently outperforms the relative coordinates across most scenarios. The exceptions are only 5 out of 36 cases, in which the relative coordinates show minor advantages over the absolute ones. The superiority of absolute coordinates shows that the joint trajectories in the global world better reflect the motion content and personal style. We believe the movement of root joints is significant for differentiating walking and running as well as different subjects (people walk or run at different preferred speeds).
- The use of l_2-norm performs best in maximizing the distance in both content and ID embedding. This is shown in the maximal value of D_C, D_P. This shows that

such a distance metric follows the L2 principle. The potential reason for causing the under-performance of the Mahalanobis distance is a relatively small dataset.

- $\mathbf{G}_a$ is optimal for both the content and ID embedding, as shown in the bold numbers. This confirms the usefulness of our auto-encoder in extracting the latent features for both the motion content and personal style.

Table 1. Motion embedding statistics. The columns of Dist and Std list the average distance and standard deviation of samples from the same motion type and person. W–R indicates the distance between the cluster centroid of walking and running, while P1–P2 indicates the distance between the cluster centroid of two persons: P1 and P2.

Feature Vector	Metric	Content Embedding Distances						ID Embedding Distances					
		Walk		Run		W–R		P1		P2		P1–P2	
		Dist D_W	Std	Dist D_R	Std	Dist D_{W-R}	D_C	Dist D_{P1}	Std	Dist D_{P2}	Std	Dist D_{P1-P2}	D_P
$\mathbf{C}_a$	$\mathbf{L}_1$	53.22	12.25	24.41	15.16	95.25	1.27	13.79	13.98	11.53	11.17	62.67	2.48
	$\mathbf{L}_2$	21.88	6.61	8.17	4.84	25.89	0.86	6.82	7.23	5.36	5.07	47.12	3.87
	$\mathbf{L}_m$	2.77	0.92	1.93	1.22	3.24	0.81	1.12	0.92	1.61	1.06	2.91	1.07
$\mathbf{C}_r$	$\mathbf{L}_1$	73.67	26.63	38.10	18.32	100.12	1.00	40.22	26.11	34.95	17.20	79.77	1.30
	$\mathbf{L}_2$	43.50	18.78	24.41	12.62	60.23	0.89	34.20	21.90	29.70	16.57	51.06	0.80
	$\mathbf{L}_m$	1.98	0.57	0.96	0.50	3.20	1.09	0.89	0.55	0.78	0.35	2.43	1.46
$\mathbf{H}_a$	$\mathbf{L}_1$	48.11	10.83	23.26	14.38	71.18	1.00	10.71	10.64	6.64	8.52	49.76	2.87
	$\mathbf{L}_2$	17.30	4.82	6.73	4.06	42.67	1.78	5.27	5.56	3.08	3.86	36.27	4.34
	$\mathbf{L}_m$	3.26	1.04	2.78	1.79	3.24	0.54	1.10	0.91	0.90	1.03	2.93	1.47
$\mathbf{H}_r$	$\mathbf{L}_1$	64.90	16.94	31.67	12.84	71.86	0.74	33.93	16.91	29.39	11.19	65.28	1.03
	$\mathbf{L}_2$	28.79	8.79	15.40	7.07	42.08	0.95	22.37	11.10	18.21	7.27	37.49	0.92
	$\mathbf{L}_m$	2.89	0.71	1.52	0.61	3.24	0.73	1.38	0.59	1.15	0.38	2.70	1.07
$\mathbf{G}_a$	$\mathbf{L}_1$	575.62	178.39	621.93	361.30	1423.82	1.19	112.60	98.07	42.51	32.51	719.86	4.64
	$\mathbf{L}_2$	**250.40**	**74.48**	**242.08**	**147.29**	**953.01**	**1.94**	**71.18**	**61.89**	**27.94**	**21.78**	**494.09**	**4.98**
	$\mathbf{L}_m$	2.16	0.74	3.17	1.96	3.23	0.61	0.82	0.91	0.33	0.33	2.90	2.52
$\mathbf{G}_r$	$\mathbf{L}_1$	2215.04	1167.85	1352.18	704.60	1954.28	0.55	1603.36	1019.97	1081.91	535.17	1008.38	0.38
	$\mathbf{L}_2$	1302.38	676.76	830.79	421.47	1312.16	0.62	1276.79	786.57	833.78	404.82	823.57	0.39
	$\mathbf{L}_m$	2.10	0.95	1.14	0.62	3.09	0.95	1.22	0.65	0.91	0.42	0.70	0.33

Note: For the two feature vectors $\mathbf{u}$ and $\mathbf{v}$, $\mathbf{L}_1 = \| \mathbf{u} - \mathbf{v} \|$, $\mathbf{L}_2 = \| \mathbf{u} - \mathbf{v} \|^2$, and $\mathbf{L}_m = \| \mathbf{u} - \mathbf{v} \|^m$ $(m > 2)$.

5.4. Subject Shape Variation

Figure 5 shows the changes in human shape when semantic parameters are altered. For cloth adaptation, we tested a different number of human shape vertices k around each cloth vertex. As we can see from Figure 6, the adaptation gets better with the increase in k; however, the discrepancy is negligible when $k \geq 6$. We also compare the different blending operators and presented the results in Figure 7, which reveals better effects are achieved with dual quaternion blending.

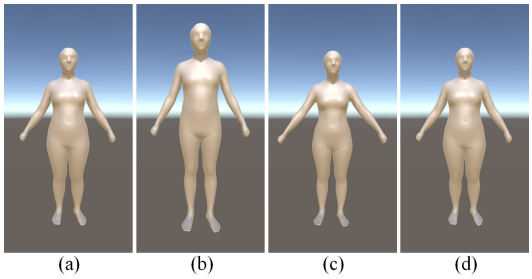


Figure 5. Parameterized editing of a human body: (**a**) Original model; (**b**) Height + 10; (**c**) Girth − 30; (**d**) Hip + 20.

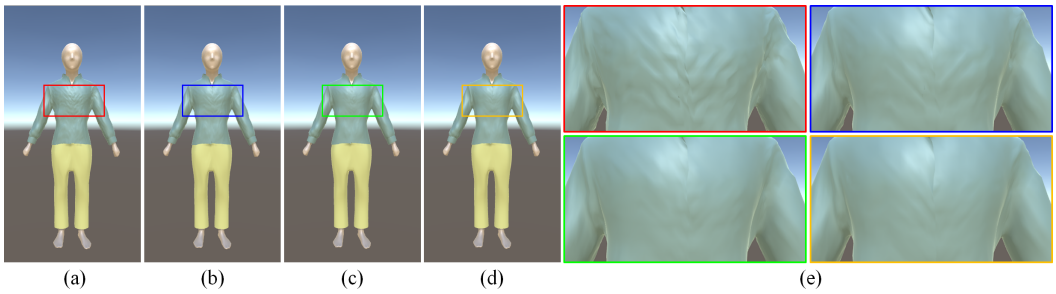


Figure 6. Effect of neighborhood size *k*: (**a**) k = 1; (**b**) k = 3; (**c**) k = 6; (**d**) k = 10; (**e**) zoom in.

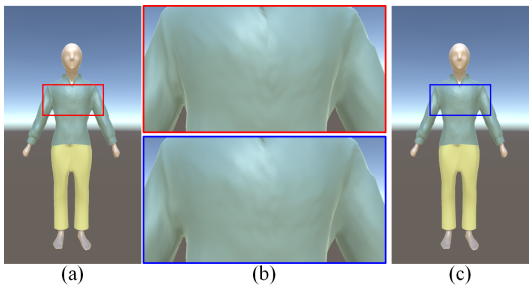


Figure 7. Effect of interpolation methods: (**a**) linear interpolation; (**b**) zoom in; (**c**) dual quaternion interpolation.

5.5. The Real-to-Virtual Gap and Closing

The real-to-virtual gap is a common problem in most methods that use synthetic data to solve real world problems. To study the problem in our paradigm, we train the GaitSet model [5] with real-world and synthetic data. GaitSet regards gait as a set of gait silhouettes. It takes a set of gait silhouettes as input, and then uses a CNN to extract frame-level features from each silhouette independently. After that, a so-called set pooling operation is used to aggregate frame-level features into a single set-level feature. Finally, a structure called horizontal pyramid mapping is used to map the set-level feature into a more discriminative space to obtain the final representation, from which gait can be recognized.

Training was conducted under three different settings as was done in [5]: small-sample training (with 24 subjects), medium-sample training (with 62 subjects), and large-sample training (with 74 subjects). We generate the training set by first sampling the area corresponding to walking on the motion manifold, and then classifying them into different subjects according to ID embedding. We picked up 74 subjects out of 124 subjects in the CASIA-B dataset [18] as real-world training data. They will be mixed with synthetic data according to specified proportions to form the whole training set. The rest of the CASIA-B and the OU-MVLP dataset [6] were used for testing of both the real data-trained system and the synthetic data-trained system.

From Figure 8a,c, we can observe a clear gap that was narrowed down when we gradually added real world data when training the synthetic model. Generally, the synthetic

model can gain accuracy in competing with the real model when about 30% percent of real data are added in the training. However, when running both models on the OU-MVLP test set, we found that there is a significant decrease in the accuracy but the gap was not so apparent. This observation indicates that the gap may be due not to the real-to-virtual but to the translation of the learned model to another domain. Another interesting phenomenon was that when we introduced real data in the training of the synthetic model, the accuracy could even outperform the real model. The reason behind this could be found in the hypothesis of domain randomization [29]: if the variability in the simulation is significant enough, models trained in simulation will generalize to the real world with no additional training. By mixing real and synthetic data together, we increased the variability of training data and thus led to an increase in performance.

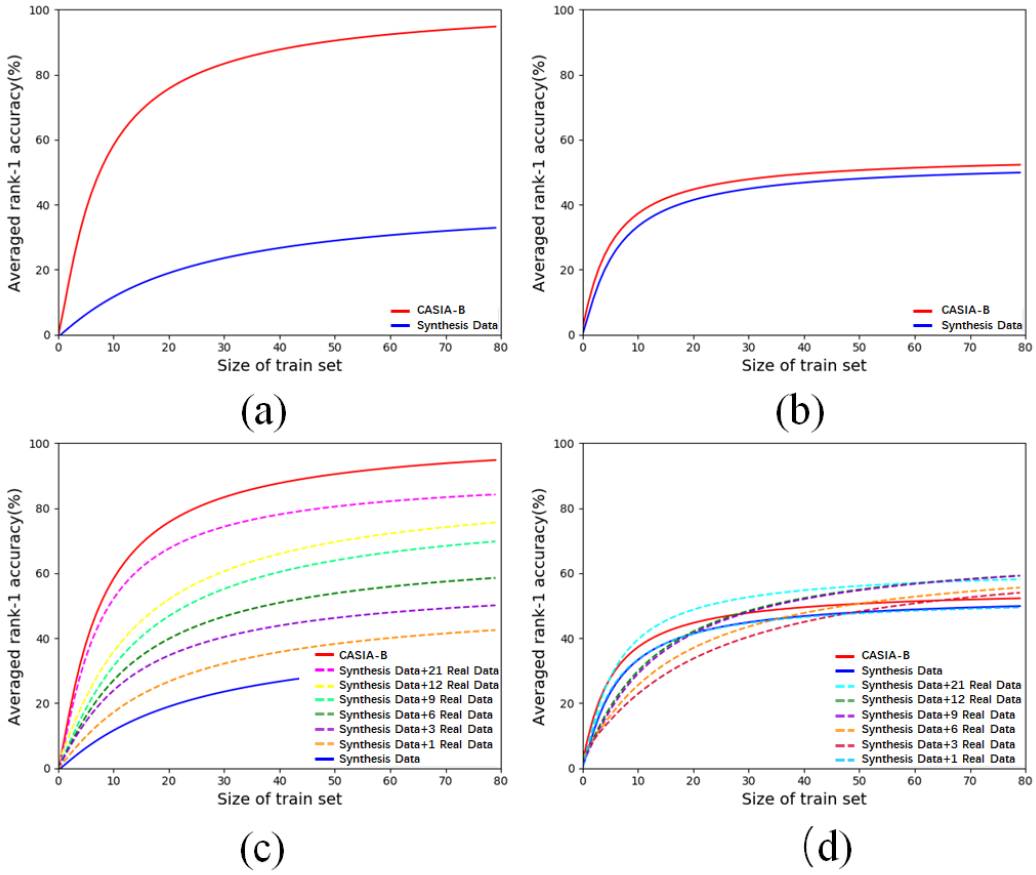


Figure 8. Performance gap and closing: the first row plots curves of accuracy against training set size for the CASIA-B test set (**a**) and OU-MVLP test set (**b**); the second row shows accuracy curves with real data added in the training of the synthetic model for the CASIA-B test set (**c**) and OU-MVLP test set (**d**).

5.6. Applications

The direct application is gait recognition, which is also the main motivation for this research. Our generator can be used to provide training data for recognition model training, to analyze and compare the performance of different recognition methods. Compared with existing datasets, a main advantage of our generator is the ability to provide many different forms of gait data. We can provide silhouette-based features like GEI images for appearance-based methods. As we can see from Figure 9a, our method provides much better silhouettes than extraction from videos using cutting-edge methods like mask-RCNN [49].

Figure 9. Performance of mask extraction: columns 1–4 (**a**) show the comparison between mask-RCNN [49] and our system; columns 5–6 (**b**) show the extraction quality of mask-RCNN under various lighting conditions; column 7 (**c**) shows mask extraction of mask-RCNN with occlusion.

We can also generate a skeleton directly for those model-based approaches and provide accelerometer data (Figure 10a) or depth images (Figure 10b) for gait analysis algorithms relying on them [50,51]. Although the accelerometer may be too smooth compared with the real one, the issue can be allieviated with appropriate noise introduced.

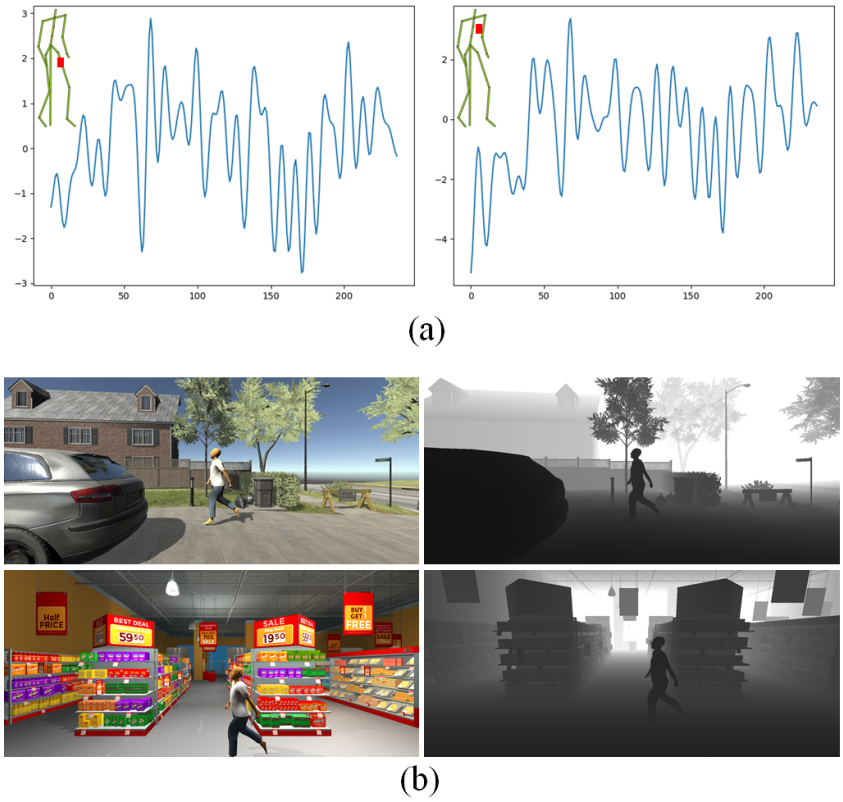


Figure 10. Simulated gait data other than silhouette provided by our system: (**a**) accelerometer data; the left shows the case with a mobile device represented by the red square in the trouser pocket, while the right shows the case in a jacket pocket; (**b**) depth map for two different scenes.

With the synthetic environment, we can easily investigate the performance of related algorithms under various extreme circumstances. Figure 9c shows how mask extraction performs under different lighting conditions, while Figure 9b shows mask extraction with occlusion. Additionally, our method can also be helpful in multi-subject gait recognition (Figure 4c bottom), which has not been investigated so far.

Finally, our method can also be used in scenarios where a new surveillance system is installed in a novel location without any instances of real gait data, similar to [52]. We can then reconstruct the 3D scene and use our method to generate training data for the recognizer.

6. Conclusions

This paper investigates how synthetic data can be used for learning-based gait analysis. A 3D gait generator is realized to successfully simulate various factors involved in gait analysis. Due to its 3D nature, the generator can provide accurate data in various forms, such as silhouettes, accelerometers, and depth maps. We can generate valid walking motion while producing diverse variations of inter-subject and intra-subject differences in style. We design an exploratory character appearance editing method to allow for altering the semantic attributes like weight, height, and girth, etc. The dressing will automatically adapt to the shape in real time. Experiments showed that the real-to-virtual gap can be effectively alleviated when real-world data are introduced to fine-tune the synthetic-trained model. We also show how the simulation of extensive gait-related factors can be helpful in investigation of gait recognition algorithms under various circumstances, especially extreme ones.

There are several directions that are worth our future efforts. So far, we can only achieve primary balance among inter-subject and intra-subject differences in walking motion synthesis. We would like to seek a richer and more semantic walking motion synthesis using linguistic descriptions, such as age, gender, energy, emotion, etc. Equipped with the generator, we would like to study more challenging gait analysis problems such as multi-subject identification and gait analysis with extreme conditions. Our current experiments are mainly designed around silhouette-based gait recognizers. We would like to explore how other types of data, such as accelerometers and depth maps provided by our generator, can be used in gait analysis.

Author Contributions: Conceptualization, J.L.; methodology, Y.C.; software, Y.C. and G.Z.; validation, Y.C. and S.H.; writing—original draft preparation, J.L. and Z.W.; writing—review and editing, X.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Research Project (No. PZ2020016) and the Natural Science Foundation of Xiamen, China (No. 3502Z20227012).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly archived datasets were analyzed in this study. These datasets can be found here: MPII Human Shape, https://humanshape.mpi-inf.mpg.de/ (accessed on 1 September 2022); CASIA-B, http://www.cbsr.ia.ac.cn/english/Gait%20Databases.asp (accessed on 1 September 2022); OU-MVLP, http://www.am.sanken.osaka-u.ac.jp/BiometricDB/GaitMVLP.html (accessed on 1 September 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, L.; Tan, T.; Ning, H.; Hu, W. Silhouette analysis-based gait recognition for human identification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 1505–1518. [CrossRef]
2. Connor, P.; Ross, A. Biometric recognition by gait: A survey of modalities and features. *Comput. Vis. Image Underst.* **2018**, *167*, 1–27. [CrossRef]
3. Feng, Y.; Li, Y.; Luo, J. Learning effective Gait features using LSTM. In Proceedings of the International Conference on Pattern Recognition, Cancun, Mexico, 4–8 December 2016; pp. 325–330.
4. Wu, Z.; Huang, Y.; Wang, L.; Wang, X.; Tan, T. A comprehensive study on cross-view gait based human identification with deep CNNs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 209–226. [CrossRef]
5. Chao, H.; He, Y.; Zhang, J.; Feng, J. GaitSet: Regarding gait as a set for cross-view gait recognition. In Proceedings of the AAAI Conference on Artifical Inteligence, Honolulu, HI, USA, 27 January–1 February 2019; pp. 8126–8133.
6. Iwama, H.; Okumura, M.; Makihara, Y.; Yagi, Y. The OU-ISIR gait database: Comprising the large population dataset and performance evaluation of gait recognition. *IEEE Trans. Inf. Forensics Secur.* **2012**, *7*, 1511–1521. [CrossRef]
7. Gaidon, A.; Wang, Q.; Cabon, Y.; Vig, E. Virtual worlds as proxy for multi-object tracking analysis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4340–4349.

8. Martinez-Gonzalez, P.; Oprea, S.; Garcia-Garcia, A.; Jover-Alvarez, A.; Orts-Escolano, S.; Garcia-Rodriguez, J. Unrealrox: An extremely photorealistic virtual reality environment for robotics simulations and synthetic data generation. *Virtual Real.* **2020**, *24*, 271–288. [CrossRef]
9. McCormac, J.; Handa, A.; Leutenegger, S.; Davison, A.J. Scenenet rgb-d: 5m photorealistic images of synthetic indoor trajectories with ground truth. *arXiv* **2016**, arXiv:1612.05079.
10. Ariyanto, G.; Nixon, M.S. Model-based 3D gait biometrics. In Proceedings of the International Joint Conference on Biometrics, Washington, DC, USA, 11–13 October 2011; pp. 1–7.
11. He, Y.; Zhang, J.; Shan, H.; Wang, L. Multi-task GANs for view-specific feature learning in gait recognition. *IEEE Trans. Inf. Forensics Secur.* **2018**, *14*, 102–113. [CrossRef]
12. Sivapalan, S.; Chen, D.; Denman, S.; Sridharan, S.; Fookes, C. Histogram of weighted local directions for gait recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Portland, OR, USA, 23–28 June 2013; pp. 125–130.
13. Muramatsu, D.; Shiraishi, A.; Makihara, Y.; Uddin, M.Z.; Yagi, Y. Gait-based person recognition using arbitrary view transformation model. *IEEE Trans. Image Process.* **2015**, *24*, 140–154. [CrossRef]
14. Makihara, Y.; Suzuki, A.; Muramatsu, D.; Li, X.; Yagi, Y. Joint intensity and spatial metric learning for robust gait recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5705–5715.
15. Zhang, K.; Luo, W.; Ma, L.; Liu, W.; Li, H. Learning joint gait representation via quintuplet loss minimization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4700–4709.
16. Zhang, Z.; Tran, L.; Yin, X.; Atoum, Y.; Liu, X.; Wan, J.; Wang, N. Gait recognition via disentangled representation learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4710–4719.
17. Sarkar, S.; Phillips, P.J.; Liu, Z.; Vega, I.R.; Grother, P.; Bowyer, K.W. The humanID gait challenge problem: Data sets, performance, and analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 162–177. [CrossRef]
18. Yu, S.; Tan, D.; Tan, T. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In Proceedings of the International Conference on Pattern Recognition, Hong Kong, China, 20–24 August 2006; pp. 441–444.
19. Matovski, D.S.; Nixon, M.S.; Mahmoodi, S.; Carter, J.N. The effect of time on the performance of gait biometrics. In Proceedings of the IEEE International Conference on Biometrics: Theory, Applications and Systems, Tampa, FL, USA, 23–26 September 2010; pp. 1–6.
20. Xu, C.; Makihara, Y.; Ogi, G.; Li, X.; Yagi, Y.; Lu, J. The OU-ISIR gait database comprising the large population dataset with age and performance evaluation of age estimation. *IPSJ Trans. Comput. Vis. Appl.* **2017**, *9*, 1–14. [CrossRef]
21. Pavlakos, G.; Choutas, V.; Ghorbani, N.; Bolkart, T.; Osman, A.A.A.; Tzionas, D.; Black, M.J. Expressive Body Capture: 3D Hands, Face, and Body from a Single Image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 10975–10985.
22. Han, J.; Bhanu, B. Individual recognition using gait energy image. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 316–322. [CrossRef] [PubMed]
23. Dosovitskiy, A.; Fischer, P.; Ilg, E.; H?usser, P.; Hazirbas, C.; Golkov, V.; van der Smagt, P.; Cremers, D.; Brox, T. FlowNet: Learning optical flow with convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 2758–2766.
24. Handa, A.; Patraucean, V.; Badrinarayanan, V.; Stent, S.; Cipolla, R. Understanding real world indoor scenes with synthetic data. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4077–4085.
25. Sakaridis, C.; Dai, D.; Gool, L.V. Semantic foggy scene understanding with synthetic data. *Int. J. Comput. Vis.* **2018**, *126*, 973–992. [CrossRef]
26. Huang, S.; Ramanan, D. Expecting the unexpected: Training detectors for unusual pedestrians with adversarial imposters. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2243–2252.
27. Varol, G.; Romero, J.; Martin, X.; Mahmood, N.; Black, M.J.; Laptev, I.; Schmid, C. Learning from synthetic humans. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 109–117.
28. Roberto de Souza, C.; Gaidon, A.; Cabon, Y.; Manuel Lopez, A. Procedural generation of videos to train deep action recognition networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4757–4767.
29. Tobin, J.; Fong, R.; Ray, A.; Schneider, J.; Zaremba, W.; Abbeel, P. Domain randomization for transferring deep neural networks from simulation to the real world. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Vancouver, BC, Canada, 24–28 September 2017; pp. 23–30.
30. Tobin, J.; Zaremba, W.; Abbeel, P. Domain randomization and generative models for robotic grasping. *arXiv* **2017**, arXiv:1710.06425.
31. Kortylewski, A.; Schneider, A.; Gerig, T.; Egger, B.; Morel-Forster, A.; Vetter, T. Training deep face recognition systems with synthetic data. *arXiv* **2018**, arXiv:1802.05891.

32. Chen, X.; Duan, Y.; Houthooft, R.; Schulman, J.; Sutskever, I.; Abbeel, P. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets. In Proceedings of the International Conference on Neural Information Processing Systems, Barcelona, Spain, 5–10 December 2016; pp. 2180–2188.
33. Tran, L.; Yin, X.; Liu, X. Disentangled representation learning gan for pose-invariant face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1415–1424.
34. Shen, Y.; Luo, P.; Yan, J.; Wang, X.; Tang, X. FaceID-GAN: Learning a symmetry three-player GAN for identity-preserving face synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 821–830.
35. Shen, Y.; Zhou, B.; Luo, P.; Tang, X. FaceFeat-GAN: A two-stage approach for identity-preserving face synthesis. *arXiv* **2018**, arXiv:1812.01288.
36. Bao, J.; Chen, D.; Wen, F.; Li, H.; Hua, G. Towards open-set identity preserving face synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6713–6722.
37. Nitzan, Y.; Bermano, A.; Li, Y.; Cohen-Or, D. Face identity disentanglement via latent space mapping. *ACM Trans. Graph.* **2020**, *39*, 1–14. [CrossRef]
38. Li, S.; Xiao, T.; Li, H.; Yang, W.; Wang, X. Identity-aware textual-visual matching with latent co-attention. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1890–1899.
39. Zhou, Y.; Han, X.; Shechtman, E.; Echevarria, J.; Kalogerakis, E.; Li, D. MakeItTalk: Speaker-aware talking head animation. *ACM Trans. Graph.* **2020**, *39*, 1–15. [CrossRef]
40. Yoon, Y.; Cha, B.; Lee, J.H.; Jang, M.; Lee, J.; Kim, J.; Lee, G. Speech gesture generation from the trimodal context of text, audio, and speaker identity. *ACM Trans. Graph.* **2020**, *39*, 1–16. [CrossRef]
41. CMU. Carnegie-Mellon Mocap Database. Available online: http://mocap.cs.cmu.edu/ (accessed on 1 September 2022).
42. Xia, S.; Wang, C.; Chai, J.; Hodgins, J. Realtime style transfer for unlabeled heterogeneous human motion. *ACM Trans. Graph.* **2015**, *34*, 1–10. [CrossRef]
43. Müller, M.; Röder, T.; Clausen, M.; Eberhardt, B.; Krüger, B.; Weber, A. *Documentation Mocap Database HDM05*; Report No. CG-2007-2; Universitat Bonn: Bonn, Germany, 2007.
44. Holden, D.; Saito, J.; Komura, T. A deep learning framework for character motion synthesis and editing. *ACM Trans. Graph.* **2016**, *35*, 1–11. [CrossRef]
45. de Vazelhes, W.; Carey, C.; Tang, Y.; Vauquier, N.; Bellet, A. Metric-learn: Metric learning algorithms in Python. *arXiv* **2019**, arXiv:1908.04710.
46. Pishchulin, L.; Wuhrer, S.; Helten, T.; Theobalt, C.; Schiele, B. Building statistical shape spaces for 3D human modeling. *Pattern Recognit.* **2017**, *67*, 276–286. [CrossRef]
47. Chu, C.H.; Tsai, Y.T.; Wang, C.C.L.; Kwok, T.H. Exemplar-based statistical model for semantic parametric design of human body. *Comput. Ind.* **2010**, *61*, 541–549. [CrossRef]
48. Sorkine, O.; Alexa, M. As-rigid-as-possible surface modeling. In Proceedings of the Eurographics Symposium on Geometry Processing, Barcelona, Spain, 4–6 July 2007; pp. 109–116.
49. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
50. Gadaleta, M.; Rossi, M. IDNet: Smartphone-based gait recognition with convolutional neural networks. *Pattern Recognit.* **2018**, *74*, 25–37. [CrossRef]
51. Chattopadhyay, P.; Sural, S.; Mukherjee, J. Frontal gait recognition from incomplete sequences using RGB-D camera. *IEEE Trans. Inf. Forensics Secur.* **2014**, *9*, 1843–1856. [CrossRef]
52. Hattori, H.; Boddeti, V.N.; Kitani, K.; Kanade, T. Learning scene-specific pedestrian detectors without real data. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3819–3827.