Generating Virtual Avatars with Personalized Walking Gaits using Commodity Hardware

http://gamma.cs.unc.edu/pedvr/ (video included)

Sahil Narang*
University of North Carolina
Chapel Hill, North Carolina
sahil@cs.unc.edu

Andrew Best*
University of North Carolina
Chapel Hill, North Carolina
best@cs.unc.edu

Ari Shapiro
Institute for Creative Technologies, University of Southern California
Los Angeles, California
shapiro@ict.usc.edu

Dinesh Manocha
University of North Carolina
Chapel Hill, North Carolina
dm@cs.unc.edu

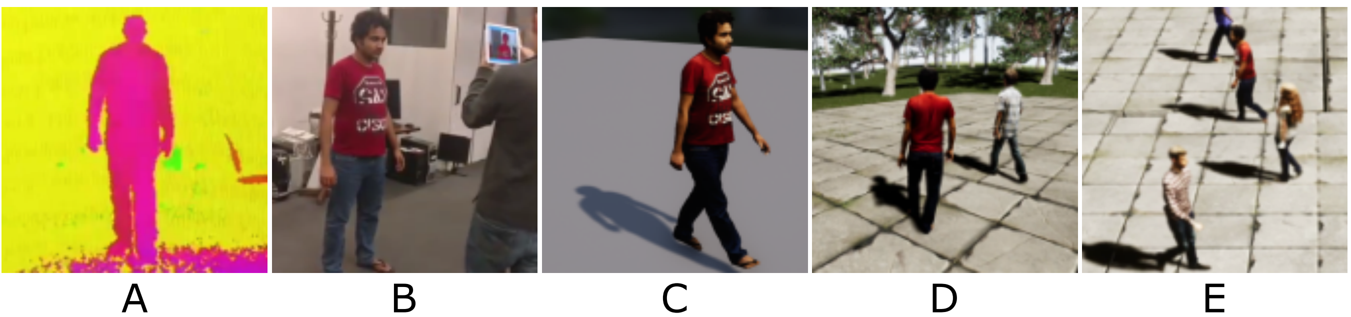

Figure 1: Our approach can quickly generate virtual avatars which move like their human subject using inexpensive, commodity hardware. Our novel gait synthesis algorithm takes as input (A) noisy gait of the user captured using Kinect v2 sensor, and automatically synthesizes a clean perceptually similar gait using a precomputed gait database; (B) we use commodity sensors to generate virtual avatars and (C) animate them with the synthesized gait with minimal artistic intervention. (D,E) A virtual avatar explores a scenario populated with virtual agents. Each agent in the crowd, as well as the virtual avatar, is captured using our scanning and gait synthesis approach, and corresponds to a real subject.

ABSTRACT

We present a novel algorithm for generating virtual avatars which move like the represented human subject, using inexpensive sensors & commodity hardware. Our algorithm is based on a perceptual study that evaluates self-recognition and similarity of gaits rendered on virtual avatars. We identify discriminatory features of human gait and propose a data-driven synthesis algorithm that can generate a set of similar gaits from a single walker. These features are combined to automatically synthesize personalized gaits for a human user from noisy motion capture data. The overall approach is robust and can generate new gaits with little or no artistic intervention using commodity sensors in a simple laboratory setting. We demonstrate our approach's application in rapidly animating virtual avatars of new users with personalized gaits, as well as procedurally generating distinct but similar "families" of gait in virtual environments.

CCS CONCEPTS

• **Computing methodologies → Interactive simulation**; **Perception**; **Virtual reality**; *Motion processing*;

KEYWORDS

perception, gait modeling, virtual avatars

ACM Reference Format:
Sahil Narang, Andrew Best, Ari Shapiro, and Dinesh Manocha. 2017. Generating Virtual Avatars with Personalized Walking Gaits using Commodity Hardware. In *Proceedings of ThematicWorkshops'17, Mountain View, CA, USA, October 23–27, 2017,* 9 pages.
https://doi.org/10.1145/3126686.3126766

*Authors contributed equally to the work.

ThematicWorkshops'17, October 23–27, 2017, Mountain View, CA, USA

ACM ISBN 978-1-4503-5416-5/17/10…$15.00
https://doi.org/10.1145/3126686.3126766

1 INTRODUCTION

Recent advances in capturing and rendering technology have enabled the rapid creation of virtual 3D avatars that resemble the human subject and can act as a representation of the human subject in immersive applications, such as military training simulations, telepresence, virtual counseling, virtual tourism, multi-player games, and treating psychological disorders.

Generating believable virtual avatars is a challenging and wide-ranging goal and involves many difficult issues: appearance and motion, as well as mechanisms of control for interactivity and autonomy, including gesture, locomotion, multi-agent interaction, and coordination. Studies have shown positive effects of visual realism on virtual avatar perception [21]. There is considerable work on capturing and rendering technologies that can enable rapid generation of photo-realistic 3D avatars of human subjects [37].

Recent work in motion realism has demonstrated important connections between the virtual avatar's motions or actions and the user's perception of immersion [6, 15, 22]. In this paper, we mainly address the problem of generating plausible gait for a virtual avatar using inexpensive sensors. A person's gait, their style of walking, is a unique feature of their motion and has been shown to be an effective biometric cue for visual identification [43]. Several studies on the perception of walking gait have found that biological motion is recognizable for both self and others in captured footage [8, 13, 20], and also when the motion is rendered on virtual avatars [32].

Despite the importance of gait in identifying with the virtual avatar, there is little work in generating personalized gaits. Recent developments in social VR applications, such as [3, 4], allow the users to create highly detailed appearances for avatars, but use either simple translations or pre-defined gaits for animating the avatars. This can largely be attributed to the fact that simulating human locomotion is regarded as a hard problem in VR, animation and multimedia literature. The current state of the art in synthesizing individual human walking gaits relies on the use of motion capture technology based on expensive and calibrated sensors. The resulting algorithms can generate natural-looking or plausible motion, though the process may require significant artistic intervention or expensive capture environments. Therefore, most databases used for human locomotion tend to consist only of tens of motion samples of one or two human actors, which are then retargeted to virtual avatars of other subjects [35]. While these approaches can be effective for certain applications, the individual style of the retargeted persons is lost and the gait motion may not provide sufficient fidelity for identification of self or others.

Main Results We present a novel data-driven algorithm that can generate virtual avatars that walk like the modeled human subject. Our approach leverages simple, inexpensive sensors that cost only a few hundred dollars and can be used in a lab or home setting. Our formulation is based on a perceptual study that evaluates the ability of participants to identify their captured gait in a sequence of pairwise comparisons. We analyze the results of the perceptual study to identify robust spatio-temporal features of human gait that relate the physical properties of gait to perceptual evaluation in terms of similarity with respect to a reference gait. We use this evaluation to generate a database of labeled pairs of gait and develop novel data-driven methods including:

- *Discriminatory Gait Features:* We present a novel gait descriptor which is used to train a neural network that can accurately label pairs of gait as "similar" or "dissimilar", with respect to user responses.
- *Personalized Gaits:* We propose a novel data-driven algorithm for synthesizing personalized gaits for new users that takes as input noisy mocap data, captured using Kinect v2 sensor, and automatically generates a perceptually similar gait in a few seconds.
- *Gait Families:* Our synthesis algorithm can be used to generate a set of gaits that share some common stylistic attributes.
- *Personalized Virtual Avatars:* The personalized gait can be used to quickly animate the virtual avatar of the subject with minimal artistic intervention. We can effectively generate personalized avatars using inexpensive commodity hardware, such as Kinect v2 and Structure sensor [2].

The overall approach offers many benefits over the state of the art. Our algorithm only needs a small database of diverse gaits to automatically generate perceptually similar gaits for a large number of users. We use features that are robust to noise in the reference gait and can synthesize smooth gaits with no artistic intervention. Furthermore, our technique for generating personalized avatars is ideal for home-based scanning.

We demonstrate several applications with multiple virtual avatars. We also evaluate the performance of our synthesis algorithm by testing it on a number of noisy gait samples of new users, captured using a Kinect v2 sensor.

2 RELATED WORK

In this section, we give a brief overview of prior work 3D avatar and gait synthesis, the perception of biological motion and automatic gait recognition.

2.1 3D Full-body Avatar Synthesis

Recent advances in low-cost scanning have lead to inexpensive solutions for efficient 3D shape acquisition and modelling of human subjects. Solutions such as [38, 47] employ multiple pre-calibrated Kinect cameras to generate full body avatars. Other solutions, such as [11, 36, 44], utilize a single 3D sensor to generate avatars. These methods rely on inexpensive sensors such as the Kinect v2, and can be used for home-based scanning and applications. On the other hand, solutions such as [37] employ 100's of cameras embedded in a static rig to generate high quality avatars.

2.2 3D Gait Synthesis

There is extensive literature in computer graphics and animation on synthesizing human gaits [45]. These include procedural methods, which apply kinematic principles based on biomechanics [18, 25], data-driven methods which generate new trajectories by blending multiple motions [9, 23, 28], and physics-based approaches [10, 26, 27, 29, 30, 40–42, 46]. Our approach is complementary to most of this work. For example, our similarity classifier can be used as an optimization constraint by biped locomotion controllers to generate gaits that are perceptually similar to a target gait.

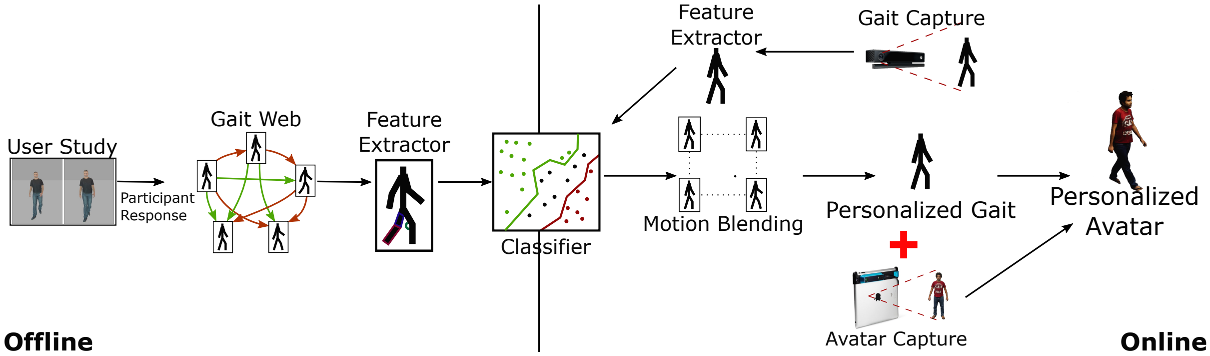

Figure 2: Overview: During an offline phase, we train a neural network to label pairs of gaits as "similar" or "dissimilar" based on the results of a user study. In the online phase, we use the Kinect v2 sensor to capture the noisy gait of a new user and automatically synthesize a clean gait that is perceptually similar to the user's gait. We use an inexpensive commodity sensor [2] to capture the avatar of the user, and animate it with the synthesized gait of the user. We can generate personalized avatars of new users for VR applications in a matter of a few minutes, with minimal artistic intervention.

2.3 Perception of Biological Motion

There is extensive literature in psychology on the perception of biological motion. Johansson [19] introduced the concept of point-light walkers which allowed for the separation and study of motion cues alone. Studies have shown that users can determine the gender of a person [24], identify individual persons [13], and recognize emotions [34] using simple point-lights walkers. Other studies have shown that users can even recognize their own point-light displays [8], which highlights the role of our motor system on the perception of motion. Troje et al. [39] transformed motion capture data into a low dimensional subspace and effectively trained linear classifiers to detect human characteristics such as gender and even emotional attributes. In contrast, our approach identifies spatio-temporal features of gait that are robust to noise and can be used to synthesize similar gaits to a reference gait.

There is considerable work related to perception of simulated human motion. Hodgins et al. [16] observed that users are more sensitive to motion variations rendered on a polygonal model in comparison to a stick figure model. Pražák [35] observed that even a small number of individual motions, as low as two or three, could be enough to make a virtual crowd look varied. Hoyet et al. [17] investigated the distinctiveness and attractiveness of a set of human motions rendered on virtual avatars. Narang et al. [32] concluded that users can recognize motion of self and others when rendered on virtual avatars. Our work is complementary as we seek to synthesize gaits for virtual avatars that are similar to the subject's gait.

2.4 Automatic Gait Identification

Research in the biometrics community has sought to exploit the uniqueness of gait across individuals and to develop "gait signatures" for automatic identification of individuals [12]. Many of these methods analyze motion capture databases of human gait to determine gait descriptors that can uniquely identify an individual [5, 7]. Given their objective of automated human identification, these methods use a purely statistical approach to analyzing human gait. In contrast, our work seeks to identify similar and dissimilar gaits in terms of human perception. Nevertheless, we use their findings to guide our research in determining discriminatory gait features.

3 OVERVIEW

In this section, we introduce the notation and the terminology used in the rest of the paper. Furthermore, we give an overview of our novel gait synthesis algorithm.

3.1 Notation

We denote a scalar variable n with a lower case letter, and a vector $\mathbf{x}$ with a bold face lower case letter. We represent the configuration of a skeleton, also called a *pose*, as a state vector $\mathbf{p} \in \mathbb{R}^{3n+3}$ for a skeleton with n joints. We denote the time-varying configuration of a skeleton, $\mathcal{C} = \mathbf{p}(t)$, with an upper case calligraphic letter. Thus, $\mathcal{C}$, referred to as a motion, is comprised of a sequence of m poses such that $\mathcal{C} = \{\mathbf{p}_{t_0}, ..., \mathbf{p}_{t_{m-1}}\}$ where $\mathbf{p}_{t_0}$ denotes the pose at time $t = t_0$. Each pose can be written as $\mathbf{p} = \{\mathbf{root}, \mathbf{j}^1, \mathbf{j}^2 ..., \mathbf{j}^n\}$, where $\mathbf{root} \in \mathbb{R}^6$ denotes the position and rotational data for the root joint, and $\mathbf{j}^k = \{j^k_{yaw}, j^k_{pitch}, j^k_{roll}\}$ denotes the rotation of the kth joint. Furthermore, we denote the ith motion as $\mathcal{C}^i$.

Unless otherwise specified, each motion $\mathcal{C}$ captures a gait cycle, defined as the time interval between successive instances of the toe leaving the ground ('toe off') for the same foot [31]. A gait cycle is comprised of two distinct phases i.e. stance and swing phase, for each leg. The stance phase for the left leg is defined as the period in which the left foot is in contact with the floor, while the swing phase is defined as the period in which the left foot is off the floor moving forward to the next step.

3.2 Our Approach

Our approach aims to generate virtual avatars using commodity hardware, that reflect the appearance and walking gait of the subject(Figure 2). It is comprised of two phases:

Offline Training: We wish to be able to synthesize a gait for a new user that reflects his/her individual style. To that end, we conducted a perceptual study to evaluate the similarity of gaits rendered on virtual avatars (Section 4). Participants were asked to identify their captured gait, rendered on their virtual avatar, in a sequence of pairwise comparisons with gaits of other participants. Users were able to recognize their gait with 60.93% accuracy. In addition, we observed that the recognition ability of users varied significantly with respect to the comparison gait, suggesting that some pairs of gaits were more perceptually similar than others.

Gait-web: The sequence of pairwise gait comparisons and the corresponding responses of the user, are used to generate a database, referred to as *gait-web*. It is shown in Figure 2 as a directed graph where a vertex v_i denotes the gait of subject i, and directed edge $e : v_i \rightarrow v_j$ denotes the response of subject i when comparing his/her gait with the gait of subject j. Responses where subjects were able to recognize their gait are labelled as "dissimilar"(red), while incorrect responses were labelled as "similar"(green). We analyze the gait-web to design a novel gait descriptor that encapsulates physical features that affect the perception of gait.

Online Avatar Generation: Our approach generates a virtual avatar which walks like the modelled subject. We capture the visual features of the virtual avatar using a commodity scanner, such as [2] and automatically rig the avatar mesh using the method proposed by [14]. We capture a sample of the user's gait through a series of walking trials recorded by a Kinect2 sensor. The data captured by the Kinect2 is noisy; often, the sensor records jitter in joint positions and discontinuities in the positions of the detected subject's limbs. As such, the Kinect2 data is often not directly suitable for synthesizing a gait for a virtual avatar. Instead, we utilize this noisy data, as well as the gait-web, to synthesize a non-noisy perceptually similar gait for the virtual avatar, as explained below.

From the noisy data captured by the Kinect2 sensor, we extract robust spatio-temporal features that are used to query the classifier and label the precomputed gaits in the gait-web as "similar" or "dissimilar" with respect to the user's gait. We synthesize a new gait by blending the gaits labelled as "similar" to the user's captured gait. Our approach can automatically synthesize the gait of a user in a few seconds. The overall process to generate the personalized avatar, including motion capture, avatar capture and animation, requires only a few minutes with minimal artistic intervention.

4 PERCEPTUAL EVALUATION OF SELF MOTION

We conducted a user study to evaluate the physical features that contribute to the perceptual similarity of distinct gaits, rendered on the same avatar. The experimental details of the study are described below.

4.1 Participants

Our study consisted of 22 participants (11 male, $\bar{x}_{age}$ = 27.13, s_{age} = 6.24) recruited from the staff and students of a large west coast university. Data for 5 participants after the on-site portion was determined to be too noisy by means of visual inspection, and was discarded.

4.2 Procedure

Participants were welcomed and were instructed on the overall process and purpose of the study. They signed a consent form and provided demographic information about their gender and age. Participants were then asked to step inside a photogrammetry stage and stand still for 5 seconds for scanning. More details on the photogrammetry stage and scanning can be found in [32]. Following the 3D scan, participants were instructed on wearing the motion capture suit, and were provided help in properly wearing the suit. Once the suit was calibrated, they were instructed to perform several motions in an open unobstructed space. These included walking in a straight line for 10m at a "comfortable pace", walking in a circle of radius 3m, turning in place, side stepping, etc. We utilized the captured data to generate a questionnaire that was sent to the participants three weeks after the initial data capture. Details of the questionnaire are provided in Section 4.3.

4.3 Experimental Design

The questionnaire was divided into two sections. The first section consisted of four pairs of motion clips presented in a 2-Alternative Forced Choice design. The questions posed in this section explore the effectiveness of an avatar representation for self-identification as part of ongoing research and are not reported as part of this analysis. The second section of the questionnaire focused on exploring similarities between the gaits of the subjects. Each subject was presented with a set of video pairs presented side by side. One of these was the subject's own gait presented on their virtual avatar and the other was the gait of a different participant of the same gender presented on the subject's avatar. For each pair of motion clips, the participants were asked to rate the clips using a 7 point Likert scale with values labeled (Left much better, Left Better, Left Slightly Better, No Difference, Right Slightly Better, Right Better, Right Much Better). In this response format, a value of 1 indicates a strong preference for the clip listed on the left of the comparison. Each subject compared their gait to every other subject of their gender. Subjects were posed two questions per comparison:

- **Q1** Which video shows a better depiction of yourself?
- **Q2** Which video depicts your gait (walking style)?

Variables: *Independent:* In this study, the independent variable is the specific gaits being displayed in a comparison. *Dependent:* The dependent variable in the study is the participant's response to the questions for each pairwise comparison.

4.4 Results

We recode the responses of the participants for the two aforementioned questions such that a value of 1 indicates that the participant rated *their motion* as a "Much Better" depiction of themselves, 4 indicates that they rated "Neither" clip to depict themselves and 7 indicates that they rated the *reference motion* to be a "Much Better" depiction of themselves. Figure 3 depicts the frequency of the user responses for the questions described earlier. On the question of depicting themselves (Q1), participants correctly identified the avatar with their motion with 51.56% accuracy i.e. a response of 1, 2 or 3. Similarly, on the question of depicting their gait (Q2), participants correctly identified the avatar with their motion with 60.93% accuracy. We focus on the participant responses on the explicit question of gait (Q2) for the remainder of the study, and summarize our findings as follows:

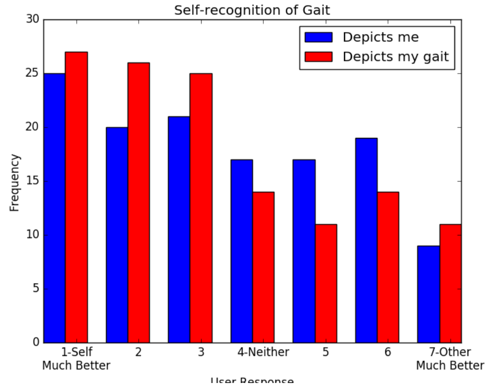

Figure 3: Frequency of user response. The figure depicts user responses for the question on depiction of self and for the question on depiction of one's gait, in a series of pairwise comparisons with reference walkers. A response of 1 indicates strong preference for self-motion, 7 denotes strong preference for the reference gait, and 4 denotes a preference for neither of the two gaits. Users correctly identified their own gaits in 60.93% of the 128 total responses. Furthermore, 21.09% of the participants rated their gait with the highest score possible i.e. 1.

(1) *Recognition Accuracy:* Participants correctly identified their own gaits in 60.93% of the 128 total responses. Furthermore, 21.09% of the participants rated their gait with the highest score possible, implying confidence in their responses.
(2) *Gait Similarity:* The participants ability to recognize his/her gait varied significantly with respect to the reference gait which suggests that some pairs of gaits are perceptually more alike than others.
(3) *Asymmetrical Responses:* The responses of participants are asymmetrical. For example, subject A recognized his/her motion in comparison to subject B, but subject B failed to recognize his/her motion in comparison to subject A.

We analyze the compared gaits to determine physical features that are discriminatory in terms of perception, and present new techniques for synthesizing perceptually similar gaits, as described in the following section.

5 DATA-DRIVEN GAIT SYNTHESIS

In this section, we use the findings of the study described above and identify physical features that affect the perceptual similarity or dissimilarity of two gaits. We propose a perceptual similarity metric comprised of these features and demonstrate its application in synthesizing perceptually similar (and dissimilar) gaits with respect to a reference gait.

5.1 Generating the Gait-Web

Based on the findings of Section 4.4, we map the user responses on the 7 point Likert scale to a label $l \in \{-1, 1\}$. A response of 3 or less denotes that the participants correctly identified their own gait, thereby implying that the compared gait was sufficiently dissimilar. Such pairs of *dissimilar gaits* are assigned the label $l = 1$. Conversely, a response of 5 or more denotes that the participants incorrectly identified the compared gait as their own, thereby implying that the compared gait was sufficiently similar. Such pairs of *similar gaits* are assigned the label $l = -1$. We use $\mathcal{F}$ to denote the set of all pairwise motion comparisons and $\mathcal{L}$ to denote the corresponding set of labels such that $\mathcal{L}_p$ denotes the label for the pth motion comparison $\mathcal{F}_p =< C^i, C^j >$. We use the term *gait-web* to describe the set of motion comparisons $\mathcal{F}$ and labels $\mathcal{L}$. We depict the gait-web pictorially using a directed graph in Figure 2.

5.2 Physical Features for Gait Perception

We identify physical features of human gait that play an important role in the perceived similarity or dissimilarity of two gaits. We analyze the set $\mathcal{F}$ of motions-pairs in the gait-web using a wide number of features, including time series data for all or some joint positions, rotations, decomposition of the entire motion; and using all combinations of the leg, arm and torso joints. We find that the rotational pattern of the hips and forearms in the saggital plane are highly correlated with the ratings-based labels of similarity and dissimilarity. We describe these features below:

5.2.1 Rotational Pattern of Hips. Let $\mathbf{p}^{Lhip}$ and $\mathbf{p}^{Lknee}$ denote the position of the left hip and left knee joints, respectively. The left hip rotation θ_k^{Lhip} at pose k can be computed as:

$$\theta_k^{Lhip} = \arccos(\frac{\mathbf{p}^{Lknee} - \mathbf{p}^{Lhip}}{\|\mathbf{p}^{Lknee} - \mathbf{p}^{Lhip}\|}.\hat{\mathbf{f}}), \tag{1}$$

where $\hat{\mathbf{f}}$ is a unit vector denoting the forward direction of the skeleton. For a given motion sampled at frequency $t = 1/\delta t$, pose k denotes the configuration of the skeleton at time $t = \delta t * k$. Thus, the set $\Theta^{LHip} = \{\theta_k^{Lhip}\}, k = 1, .., (m-1)$ for all m poses of a motion is a time series data describing the rotation pattern for the left hip [31]. The rotational patterns of the hips is periodic and can be efficiently represented by a sinusoid $\mathcal{S}_{Lhip}(A, \phi, o)$, parametrized by amplitude A, phase ϕ, and vertical offset o (Figure 5). The sinusoidal representation smooths out the noise in the time-series data, in addition to significantly reducing the dimensionality of the data. Similarly, we can compute a sinusoidal representation, $\mathcal{S}_{Rhip}$, for the rotation pattern for the right hip.

5.2.2 Rotational Pattern of Forearms. We observe that the rotation of the forearm relative to the forward direction also has a significant impact on the perception of gait, determined by the user responses. Similar to Eq. 1, the left forearm rotation θ_k^{Lelbow} at pose k can be computed as:

$$\theta_k^{Lforearm} = \arccos(\frac{\mathbf{p}^{Lwrist} - \mathbf{p}^{Lelbow}}{\|\mathbf{p}^{Lwrist} - \mathbf{p}^{Lelbow}\|}.\hat{\mathbf{f}}), \tag{2}$$

where $\mathbf{p}^{Lelbow}$ and $\mathbf{p}^{Lwrist}$ denote the position of the left elbow and left wrist joints, respectively. We describe the rotation time series data for the left and right forearms over the entire motion using sinusoid's $\mathcal{S}_{Lforearm}$ and $\mathcal{S}_{Rforearm}$, respectively.

5.2.3 Other Features. We contiguously divide the time series data for a full walk cycle comprising of four steps into four equal sized bins. In doing so, a bin approximately stores time series data

Figure 4: Comparing synthesized gait with gait extracted from noisy Kinect 2 sensor data. (Top Row) We capture the gait of a subject using commodity depth sensors. The gait captured by the sensor exaggerates arm distances from the body and generates overly stiff, unnatural motion. (Bottom Row) Our synthesized similar gait captures the arm and leg forward motions of the captured gait while generating more appropriate arm swing width and more natural torso motions.

for a particular phase of the gait cycle, such as the swing phase. We compute the maximum separation distance between the two hands, d^{hDist}, and the linear velocity of the root joint, described by mean μ^{rt} and standard deviation σ^{rt}, for each bin. As a result, the original time series motion data is represented by the vector $\mathbf{d} \in \mathbb{R}^{12}$, given as $\mathbf{d} = \{d_i^{hDist}, \mu_i^{rt}, \sigma_i^{rt}\} \forall i \in \{1, 2, 3, 4\}$.

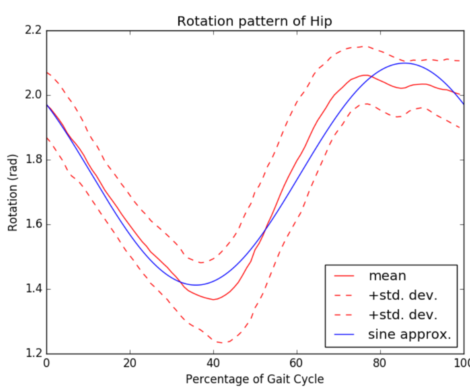

Figure 5: Perceptually important features of gait. We use the responses of the user to generate a labelled gait database, and perform an exhaustive analysis over the feature space of gaits to determine the perceptually important features. We observed that the hip rotation pattern with respect to the forward vector is a key factor in distinguishing between gaits. The figure depicts the mean and relatively large variance of the hip rotation pattern over a population of 25 healthy adults. We use a sinusoid approximation to efficiently represent the rotation pattern.

5.2.4 *Perceptual Similarity Metric.* We describe each gait in our gait-web using a *feature descriptor* $\mathbf{x} \in \mathbb{R}^{24}$ as:

$$\mathbf{x} = \{\mathcal{S}_{Lhip}, \mathcal{S}_{Rhip}, \mathcal{S}_{Lforearm}, \mathcal{S}_{Rforearm}, \mathbf{d}\}. \quad (3)$$

This feature descriptor is used as the perceptual similarity metric in order to synthesize new gaits from a precomputed database.

5.2.5 *Gait Classification.* We use a supervised learning approach to automatically classify pairs of gait as similar or dissimilar. Our classifier based on the following training set.

Training Set: We use the set $\mathcal{F}$ of motion pairs to generate a feature set $\mathcal{X}$. For each gait cycle, C^i, in the gait-web, we compute the gait descriptor $\mathbf{x}^i$. We use Principal Component Analysis (PCA) to reduce the dimensionality of the features and use $\mathbf{p}^i$ to describe the gait descriptor in terms of its principal components. Given the kth motion pair, $\mathcal{F}_k =< C^i, C^j >$, we compute $\mathcal{X}_k$ as the concatenation of $\mathbf{p}^i$ and $\mathbf{p}^j$, the corresponding feature descriptors for the gaits of subjects i and j, respectively. The set $\mathcal{L}$ of user-assigned ratings of similarity/dissimilarity is used as the associated "truth" labels. Thus, the pairs $< \mathcal{X}, \mathcal{L} >$ comprise the training set.

We reduce the dimensionality of the feature space using PCA and train a Multi-Layered Perceptron (MLP) to accurately label pairs of gaits as similar or dissimilar. The detailed analysis and implementation of this classifier is given in Section 6. Next, we describe our approach for generating labels for each gait in the database with respect to a reference gait.

5.3 Online Data-driven Gait Synthesis

Given a reference gait, our approach can automatically synthesize a "similar" gait on the fly using the similarity metric. Let C^{query} and $\mathbf{p}^{query}$ denote the reference gait and its corresponding feature descriptor in terms of principal components. We generate a test set $\mathcal{F}^{test}$, where $\mathcal{F}_i^{test}$ is computed as the concatenation of $\mathbf{p}^{query}$ and $\mathbf{p}^i$, the feature descriptor for the ith subject in the gait-web. We classify the featureset $\mathcal{F}^{test}$ to determine the set of similar and dissimilar gaits in the gait-web with respect to C^{query}. We then blend the set of database gaits determined by the classifier as similar to C^{query} with weights set to the inverse of the euclidean distance of the corresponding feature vectors.

5.4 Procedurally Generated Families of Gaits

Our approach can be used to synthesize families of distinct yet stylistically similar gaits. As described in the previous section, we use an input gait C^{query} to retrieve a set of similar database gaits, and compute a weight vector $\mathbf{w}$ that reflects the similarity of the motions in the database to C^{query}. Let ϵ denote a user defined threshold, and q denote the number of similar gaits and equivalently, the dimensionality of $\mathbf{w}$. We compute a random unit vector $\hat{\mathbf{n}} \in \mathbb{R}^q$, such that $\mathbf{w}` = \epsilon\hat{\mathbf{n}}\mathbf{w}^T$. Next, we reset the weights as $\mathbf{w} = \mathbf{w} + \mathbf{w}'$, and synthesize a new gait using the new weight. Essentially, we are using a random walk approach in $\mathbb{R}^q$, starting at the original weight vector, and synthesize a new gait at each iteration. In practice, all these gaits are similar to the reference gait based on the perceptual metric.

6 PERFORMANCE & RESULTS

In this section, we describe the training methodology and accuracy of our binary classifier for identifying similar/dissimilar gaits. We also demonstrate applications of our approach for generating virtual avatars with personalized gaits and highlight its performance.

6.1 Gait Classification

	Num. points	Precision (%)	Recall (%)	F1-score (%)
Similar	7	83	88	85
Dissimilar	16	94	92	93
Total	23	91	91	91

Table 1: Classifier Accuracy. We train a Multi-layered perceptron on a database of 116 pairwise gait comparisons to predict the similarity or dissimilarity of two gaits, where the ground truth was determined using a user evaluation. This table shows the mean precision (p), recall (r) and F1-scores ($2 * p * r/(p + r)$) computed over six trials on a testing set of 23 gait comparisons.

Our gait-web comprises of 114 user responses and 25 distinct gaits. These gaits are diverse and have been evaluated on a wide number of users and used to derive our perceptual metric. We train a Multi-Layered Perceptron (MLP) comprising of one hidden layer and ReLU activation function nodes using k-fold cross validation. We observed that the discriminatory gait descriptor, described in (Section 5.2.4) achieved 91% accuracy, measured using the mean F score on the testing set (Table 1). Furthermore, the F-score's were similar for both classes of gait i.e. the prediction and recall ability of the classifier was similar for both the similar and the dissimilar gaits.

6.2 Feature Evaluation

We evaluate our approach by synthesizing similar gaits for several actors using only a small gait database of 20 gaits. As a means of comparison, we also use our classifier to determine the most dissimilar gait from the database based on the metric. Figure 7 visualizes the sine wave approximation of the rotational pattern of the left hip, one of our four gait features, for one of the test subjects. It can be observed that the sinusoid for the similar gait is a better approximation of the motion capture gait, as compared to the dissimilar gait. Overall, the synthesized gait is visually similar to the motion-captured gait, as depicted in Figure 4. We observed that both the motion capture gait and the synthesized gait have similar hip swivels and knee flexion whereas the dissimilar gait has a distinct posture and stride pattern. Further comparisons are highlighted in the video that demonstrate the benefits of our approach.

6.3 Performance

Given an input gait cycle, our algorithm can automatically synthesize a similar gait in $3 - 15$ seconds depending on the number of gaits in the database used by the blending algorithm. Overall, we can synthesize a personalized avatar for a new subject in approx $7 - 10$ minutes, including motion capture, gait extraction using commercial solutions, synthesis of similar gait, and motion retargeting to animate the virtual avatar in a virtual environment. The process requires minimal artistic intervention such as defining the region of interest in the motion capture data for skeletal tracking, and clipping the captured walk into a full gait cycle. These steps can be easily automated.

6.4 Applications

Our approach can be used to capture and generate the personalized avatars of subjects that resemble the subject in both motion and appearance. In contrast to the current state of the art which relies on elaborate and expensive motion capture systems, our approach leverages commodity hardware such as Kinect v2 and the Structure IO 3D sensor, software such as ItSeez3D scanning software [1], and our novel gait synthesis algorithm to quickly capture and animate virtual avatars.

We also leverage a small database of gait motions to create a large number of distinct motions and thus, generate crowds with motion variety (Figure 6(Right)). Each individual in a crowd can be given a distinct gait despite the lack of motion-capture data for the individuals being represented. Therefore, it improves the fidelity and perception of the crowd.

In addition to generating a similar gait for a user, our approach can synthesize families of gaits that preserve the distinctive style of a reference walker by using different blending weights. Thus we can rapidly create sets of distinct gaits with similar styles. Our technique can be leveraged to generate set of similar moving characters, e.g. soldiers or crowds, for entertainment such as movies or games using a single reference actor.

More details on the implementation and effectiveness of our approach can be found in [33].

7 CONCLUSIONS, LIMITATIONS & FUTURE WORK

We have presented a novel approach for synthesizing virtual avatars with personalized gaits using commodity hardware that requires

Figure 6: Generating crowds with diverse motions: Current animation systems often use a small set of template motions to animate large crowds, which can appear to be cloned or unnatural. (Left) All avatars are animated with an identical gait, indicated by the perfect synchronicity in arm and leg movement across characters. (Right) Our approach generates personalized gaits for each virtual avatar, leading to diverse motion styles and natural appearing crowds. The highlighted avatar exhibits particularly salient differences between the individual and cloned gait, as can been seen in the attached video.

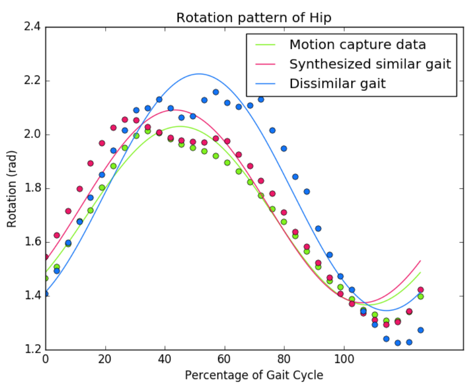

Figure 7: Comparing gait features of synthesized gait with Kinect 2 sensor data. Our data-driven gait synthesis algorithm can effectively synthesize clean gaits for new users that have similar perceptually important features, such as hip rotation patterns, to the noisy motion captured gaits.

minimal artistic intervention. Our formulation is based on a data-driven perceptual similarity metric that captures the discriminatory features of a human gait. We trained a classifier on a database of gait pairs and demonstrated its use in computing a personalized gait for a new user. Our gait synthesis algorithm is robust and can automatically handle noisy sensor data. Overall, our approach can be used to quickly model personalized avatars and is ideal for home-based scanning.

Our approach has some limitations. It assumes that the precomputed database has sufficient number of gaits of diverse style. The final set of gaits is a function of the number and variations available in the precomputed database. In many ways, the accuracy of the gait classification scheme and consequentially, the quality of the synthesized gaits depends on this database. Furthermore, the performance of our blending approach can vary based on the choice of weights. The choice of physical features relevant to gait perception may not work well in all cases. For example, it is unclear whether the proposed features are critical for other motions such as turning, sidestepping etc.

There are many avenues for future work. One of the challenges is to develop a large database of clean gaits that represent different samples corresponding to age, height, gender, ethnic background, etc. We may like to re-evaluate our proposed gait descriptor on a large database. We would like to investigate the perception of other types of motion such as turning, sidestepping etc. and evaluate our proposed feature descriptor on such non-periodic motion. We would like to further evaluate the performance of our gait synthesis algorithm in terms of the appearance or perception of the avatars in a VR environment. We would also like to completely automate our pipeline and enable animating avatars without any artistic or user intervention. Finally, we would like to evaluate the perceptual benefits of our gait generation algorithm for social VR and crowd simulation applications.

ACKNOWLEDGEMENTS

We would like to thank Andrew Feng and Sin-hwa Kang from the Institute for Creative Technologies, USC for their contribution to this work. This work was supported by Institute for Information & communications Technology Promotion (IITP) grant (No. R0184-15-1030, MR AvatarWorld Service and Platform Development using Structured Light Sensor) funded by the Korea government (MSIP), National Science Foundation award 1305286 and ARO contract W911NF-16-1-0085.

REFERENCES

[1] 2016. Mobile 3D Scanner App for iPad. (2016). https://itseez3d.com/

[2] 2016. Structure Sensor for Ipad. (2016). https://structure.io/

[3] 2017. High Fidelity. (2017). https://highfidelity.io/

[4] 2017. Sansar, Linden Research, Inc. (2017). https://www.sansar.com/

[5] Faisal Ahmed, Padma Polash Paul, and Marina L Gavrilova. 2015. DTW-based kernel and rank-level fusion for 3D gait recognition using Kinect. *The Visual Computer* 31, 6-8 (2015), 915–924.

[6] Jeremy N Bailenson, Kim Swinth, Crystal Hoyt, Susan Persky, Alex Dimov, and Jim Blascovich. 2005. The independent and interactive effects of embodied-agent appearance and behavior on self-report, cognitive, and behavioral markers of copresence in immersive virtual environments. *Presence* 14, 4 (2005), 379–393.

[7] Michal Balazia and Petr Sojka. 2016. Walker-Independent Features for Gait Recognition from Motion Capture Data. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*. Springer, 310–321.

[8] T Beardsworth and T Buckner. 1981. The ability to recognize oneself from a video recording of oneâĂŹs movements without seeing oneâĂŹs body. *Bulletin of the Psychonomic Society* 18, 1 (1981), 19–22.

[9] Myung Geol Choi, Manmyung Kim, Kyunglyul Hyun, and Jehee Lee. 2011. Deformable Motion: Squeezing into Cluttered Environments. *Comput. Graph. Forum* 30, 2 (2011), 445–453. http://dx.doi.org/10.1111/j.1467-8659.2011.01889.x

[10] Stelian Coros, Philippe Beaudoin, and Michiel van de Panne. 2010. Generalized biped walking control. In *ACM Transactions on Graphics (TOG)*, Vol. 29. ACM, 130.

[11] Yan Cui, Will Chang, Tobias Nöll, and Didier Stricker. 2012. KinectAvatar: fully automatic body capture using a single kinect. In *Asian Conference on Computer Vision*. Springer, 133–147.

[12] David Cunado, Mark S Nixon, and John N Carter. 2003. Automatic extraction and description of human gait models for recognition purposes. *Computer Vision and Image Understanding* 90, 1 (2003), 1–41.

[13] James E Cutting and Lynn T Kozlowski. 1977. Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the psychonomic society* 9, 5 (1977), 353–356.

[14] Andrew Feng, Dan Casas, and Ari Shapiro. 2015. Avatar Reshaping and Automatic Rigging Using a Deformable Model. In *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games (MIG '15)*. ACM, New York, NY, USA, 57–64. DOI: http://dx.doi.org/10.1145/2822013.2822017

[15] Andrew Feng, Gale Lucas, Stacy Marsella, Evan Suma, Chung-Cheng Chiu, Dan Casas, and Ari Shapiro. 2014. Acting the Part: The Role of Gesture on Avatar Identity. In *Proceedings of the Seventh International Conference on Motion in Games (MIG '14)*. ACM, New York, NY, USA, 49–54. DOI: http://dx.doi.org/10.1145/2668064.2668102

[16] Jessica K Hodgins, James F O'Brien, and Jack Tumblin. 1998. Perception of human motion with different geometric models. *IEEE Transactions on Visualization and Computer Graphics* 4, 4 (1998), 307–316.

[17] Ludovic Hoyet, Kenneth Ryall, Katja Zibrek, Hwangpil Park, Jehee Lee, Jessica Hodgins, and Carol O'Sullivan. 2013. Evaluating the Distinctiveness and Attractiveness of Human Motions on Realistic Virtual Bodies. *ACM Trans. Graph.* 32, 6, Article 204 (Nov. 2013), 11 pages. DOI: http://dx.doi.org/10.1145/2508363.2508367

[18] Rune Skovbo Johansen. 2009. *Automated semi-procedural animation for character locomotion*. Ph.D. Dissertation. Aarhus Universitet, Institut for Informations-og Medievidenskab.

[19] Gunnar Johansson. 1973. Visual perception of biological motion and a model for its analysis. *Perception & psychophysics* 14, 2 (1973), 201–211.

[20] Daniel Jokisch, Irene Daum, and Nikolaus F Troje. 2006. Self recognition versus recognition of others by biological motion: Viewpoint-dependent effects. *Perception* 35, 7 (2006), 911–920.

[21] Tomoko Koda and Pattie Maes. 1996. Agents with faces: the effect of personification. *Proceedings 5th IEEE International Workshop on Robot and Human Communication. RO-MAN'96 TSUKUBA* (1996), 189–194. DOI: http://dx.doi.org/10.1109/ROMAN.1996.568812

[22] Elena Kokkinara and Rachel McDonnell. 2015. Animation realism affects perceived character appeal of a self-virtual face. In *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games*. ACM, 221–226.

[23] Lucas Kovar, Michael Gleicher, and Frédéric Pighin. 2002. Motion graphs. In *ACM transactions on graphics (TOG)*, Vol. 21. ACM, 473–482.

[24] Lynn T Kozlowski and James E Cutting. 1977. Recognizing the sex of a walker from a dynamic point-light display. *Perception & Psychophysics* 21, 6 (1977), 575–580.

[25] D Le, Ronan Boulic, and Daniel Thalmann. 2003. Integrating age attributes to virtual human locomotion. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 34, VRLAB-ARTICLE-2007-007 (2003).

[26] Yoonsang Lee, Sungeun Kim, and Jehee Lee. 2010. Data-driven biped control. In *ACM Transactions on Graphics (TOG)*, Vol. 29. ACM, 129.

[27] Yoonsang Lee, Kyungho Lee, Soon-Sun Kwon, Jiwon Jeong, Carol O'Sullivan, Moon Seok Park, and Jehee Lee. 2015. Push-recovery stability of biped locomotion. *ACM Transactions on Graphics (TOG)* 34, 6 (2015), 180.

[28] Jianyuan Min and Jinxiang Chai. 2012. Motion graphs++: a compact generative model for semantic motion analysis and synthesis. *ACM Transactions on Graphics (TOG)* 31, 6 (2012), 153.

[29] Igor Mordatch, Martin De Lasa, and Aaron Hertzmann. 2010. Robust physics-based locomotion using low-dimensional planning. In *ACM Transactions on Graphics (TOG)*, Vol. 29. ACM, 71.

[30] Igor Mordatch, Jack M. Wang, Emanuel Todorov, and Vladlen Koltun. 2013. Animating human lower limbs using contact-invariant optimization. *ACM Transactions on Graphics* 32, 6 (2013), 1–8. DOI: http://dx.doi.org/10.1145/2508363.2508365

[31] M Pat Murray, A Bernard Drought, and Ross C Kory. 1964. Walking patterns of normal men. *J Bone Joint Surg Am* 46, 2 (1964), 335–360.

[32] Sahil Narang, Andrew Best, Andrew Feng, Sin-hwa Kang, Dinesh Manocha, and Ari Shapiro. 2017. Motion recognition of self and others on realistic 3D avatars. *Computer Animation and Virtual Worlds* 28, 3-4 (2017).

[33] Sahil Narang, Andrew Best, Ari Shapiro, and Dinesh Manocha. 2017. Personalized Gaits for Social Virtual Reality. (2017). http://gamma.cs.unc.edu/pedvr/PersonalAvatarTech.pdf Available at http://gamma.cs.unc.edu/pedvr.

[34] Frank E Pollick, Helena M Paterson, Armin Bruderlin, and Anthony J Sanford. 2001. Perceiving affect from arm movement. *Cognition* 82, 2 (2001), B51–B61.

[35] Martin Pražák and Carol O'Sullivan. 2011. Perceiving human motion variety. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization*. ACM, 87–92.

[36] Ari Shapiro, Andrew Feng, Ruizhe Wang, Hao Li, Mark Bolas, Gerard Medioni, and Evan Suma. 2014. Rapid avatar capture and simulation using commodity depth sensors. *Computer Animation and Virtual Worlds* 25, 3-4 (2014), 201–211. DOI: http://dx.doi.org/10.1002/cav.1579

[37] Jeremy Straub and Scott Kerlin. 2014. Development of a large, low-cost, instant 3D scanner. *Technologies* 2, 2 (2014), 76–95.

[38] J. Tong, J. Zhou, L. Liu, Z. Pan, and H. Yan. 2012. Scanning 3D Full Human Bodies Using Kinects. *IEEE Transactions on Visualization and Computer Graphics* 18, 4 (April 2012), 643–650. DOI: http://dx.doi.org/10.1109/TVCG.2012.56

[39] Nikolaus F Troje. 2002. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of vision* 2, 5 (2002), 2–2.

[40] Yao-Yang Tsai, Wen-Chieh Lin, Kuangyou B Cheng, Jehee Lee, and Tong-Yee Lee. 2010. Real-time physics-based 3d biped character animation using an inverted pendulum model. *IEEE transactions on visualization and computer graphics* 16, 2 (2010), 325–337.

[41] Michiel Van De Panne. 1996. Parameterized gait synthesis. *IEEE Computer Graphics and Applications* 16, 2 (1996), 40–49. DOI: http://dx.doi.org/10.1109/38.486679 arXiv:arXiv:1011.1669v3

[42] Kevin Wampler and Zoran Popović. 2009. Optimal gait and form for animal locomotion. *ACM Transactions on Graphics* 28, 3 (2009), 1. DOI: http://dx.doi.org/10.1145/1531326.1531366

[43] Liang Wang, Tieniu Tan, Huazhong Ning, and Weiming Hu. 2003. Silhouette analysis-based gait recognition for human identification. *IEEE transactions on pattern analysis and machine intelligence* 25, 12 (2003), 1505–1518.

[44] Ruizhe Wang, Jongmoo Choi, and Gerard Medioni. 2012. Accurate full body scanning from a single fixed 3d camera. In *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*. IEEE, 432–439.

[45] van H Welbergen, van BJH Basten, A Egges, ZM Ruttkay, and MH Overmars. 2010. Real Time Character Animation: A Trade-off Between Naturalness and Control. *Computer Graphics Forum* 29, 8 (2010).

[46] KangKang Yin, Stelian Coros, Philippe Beaudoin, and Michiel van de Panne. 2008. Continuation methods for adapting simulated skills. *ACM Transactions on Graphics* 27, 212 (2008), 1. DOI: http://dx.doi.org/10.1145/1360612.1360680

[47] Ming Zeng, Jiaxiang Zheng, Xuan Cheng, and Xinguo Liu. 2013. Templateless quasi-rigid shape modeling with implicit loop-closure. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 145–152.