A Statistical Model for Coupled Human Shape and Motion Synthesis

Alina Kuznetsova[1], Nikolaus F. Troje[2] and Bodo Rosenhahn[1]
[1]*Institute for Information Processing, Leibniz University Hanover, Hannover, Germany*
[2]*Bio Motion Lab, Queen's University, Kingston, Canada*

Keywords: Animation, Shape and Motion Synthesis, Statistical Modeling.

Abstract: Due to rapid development of virtual reality industry, realistic modeling and animation is becoming more and more important. In the paper, we propose a method to synthesize both human appearance and motion given semantic parameters, as well as to create realistic animation of still meshes and to synthesize appearance based on a given motion. Our approach is data-driven and allows to correlate two databases containing shape and motion data. The synthetic output of the model is evaluated quantitatively and in terms of visual plausibility.

1 INTRODUCTION

Emerging interest in 3D technologies and virtual reality introduced a need for realistic computer modeling and animation. It is a rapidly developing area and yet there are many unresolved problems, such as fast and realistic character creation. In this paper, we propose a statistical model to solve the latter problem.

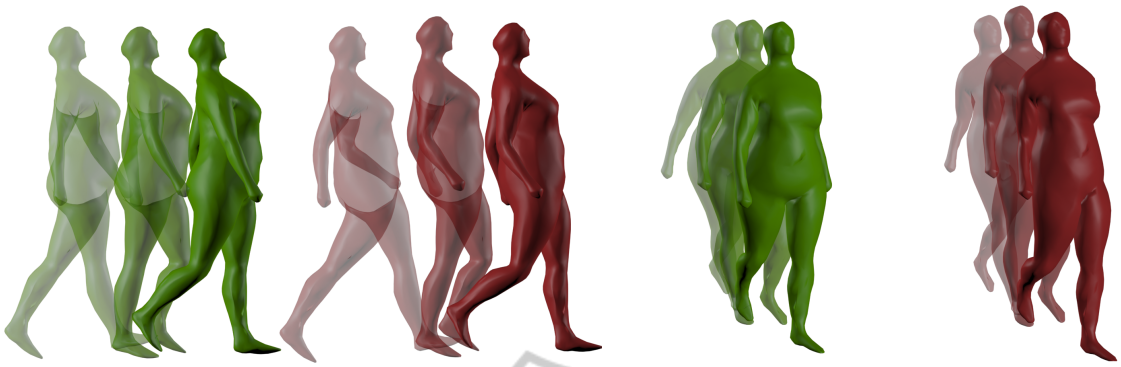
In general, character creation is a challenging task. The character generation problem is addressed either by manual editing and 3D modeling or by data-driven approaches; a well-known example of appearance creation is the SCAPE model (Anguelov et al., 2005). Motion simulation is usually considered as a separate topic, where three types of methods exist: manual motion editing, physics-based approaches and data-driven approaches. The oldest approach is manual motion editing, such as key frame animation. However, manual editing is not able to provide enough level of motion detailization and is very time-consuming, therefore data-driven approaches emerged recently, as well as physics based and control-based methods. Unfortunately, data-driven approaches usually produce physically incorrect results and the question of fitting generated motion to the concrete character (motion retargeting) is not solved completely; physics-based approaches are usually very computationally-intensive and complex, do not provide enough variability and, as experiments showed, physical fidelity of motion does not imply visual plausibility. Another way to animate a new character is to transfer motion from another character, but when the characters have different parameters (such as proportions, height, etc), such transfer can produce unrealistic results (so-called retargeting problem), as shown on the Fig 1. Here, the same shape was animated using different motions, one of which comes from a person with similar biometric parameters and the other is the motion of the person with completely different biometric parameters. Even from a sequence of images it is possible to see a mismatch between the shape and the motion.

The other disadvantage of the approaches described above is that they are difficult to apply, when the task is to animate many characters at once.

In this work, we propose a model for character appearance creation and animation by combining a statistical model of human motion with one for character appearance. In this way we address the problems described above. We are able to generate both the character's appearance and motion simultaneously, therefore avoiding the retargeting problem and extensive computations, while still producing visually plausible animations. Semantic parameters, such as weight, height and proportions of the character, serve as a link between shape and motion, and are integrated in our model, allowing excellent control over the generation process. Since our model is stochastic, it can be used for random character generation.

Our paper is organized in the following way. In Section 2, we give a short overview of existing methods for character generation and animation. In Section 3, we give technical details about data collection and processing. In Section 4 we explain the model we use for generation. In Section 5, we provide an evaluation of our approach in terms of visual perception

Kuznetsova A., Troje N. and Rosenhahn B..
A Statistical Model for Coupled Human Shape and Motion Synthesis.
DOI: 10.5220/0004215602270236
In *Proceedings of the International Conference on Computer Graphics Theory and Applications and International Conference on Information Visualization Theory and Applications* (GRAPP-2013), pages 227-236
ISBN: 978-989-8565-46-4



(a) (b)

Figure 1: Example of motion retargeting problem: (a) red figure unrealistically bends backwards, while green figure has correct balance; (b) red figure has its shoulders set backwards, as is typical for slimmer person.

of generated characters and their motions, as well as a quantitative assessment of an appearance-to-motion fit.

2 RELATED WORKS

In this section, we revise already existing works on motion and shape modeling. Since standard approaches to this problem separate creation of character and creation of its motion, we first provide a short review of methods proposed for both problems.

Motion Simulation Techniques. Many techniques address the motion simulation problem. A purely data-driven approach for motion simulation was introduced by (Wang et al., 2008) and is based on Gaussian process models for motion generation; (Li et al., 2002) proposed to use linear dynamic systems with distribution of the dynamic system parameters to control the generation; (Brand and Hertzmann, 2000) used style machines to create variations in one type of motion. Finally, (Troje, 2002) used Principal Component Analysis (PCA) analysis to build a manifold of cyclic motions.

Another approach to motion synthesis is in combining already existing motions without usage of any statistical models, for example, in (Sidenbladh et al., 2002), no motion is synthesized, but the closest real motion in the database of motions is found based on predefined metrics.

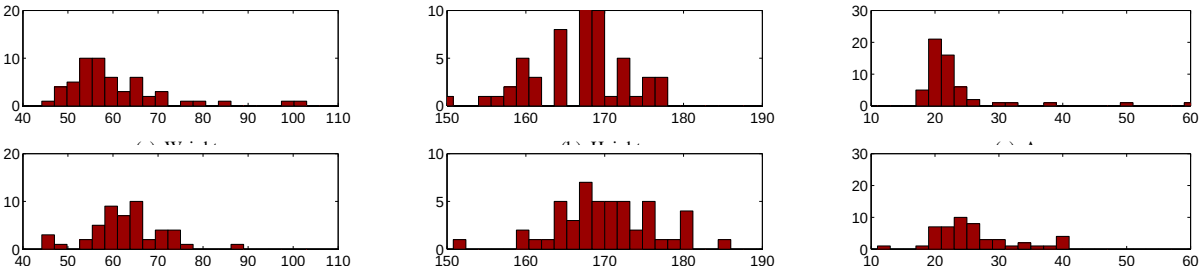
Physics-based models of human motion, such as proposed in (Liu et al., 2005), are used to make motion physically plausible, which is often required for animation of interactions. More frequently, however, different types of controllers are used to refine already generated motion, for example as proposed in (da Silva et al., 2008; Lee et al., 2009; Popovic and Witkin, 1999; Sok et al., 2007).

Mesh Simulation Techniques. The traditional way of creating character appearance is manual modeling. Recently, automated approaches were presented, such as the SCAPE model (Anguelov et al., 2005), that is now state-of-the-art technique. A slightly different statistical model is presented in (Hasler et al., 2009). Here, the use of semantic "handles" is also proposed to achieve variability of shapes.

Our Contribution. In contrast to the above mentioned works, where mesh and motion are clearly separated, we propose to combine appearance creation with motion simulation, therefore avoiding many problems of motion retargeting. To our knowledge, there are no previous works bringing motion simulation and shape creation together. We apply the model to generate joint realistic human shape and motion.

We also see our contribution in developing a method to find dependencies in two separate databases and building a model, that allows to unite the data from both databases, using the dependencies found.

In general, we can imagine many applications of the model, for example, crowd simulation, automatic animation of already existing characters or appearance-from-motion reconstruction. In mathematical sense, our model is inspired by variational models, proposed in (Cootes et al., 2001). However, we include a random component and therefore allow for more loose coupling of motion and shape, as well as for variability in sampled characters.



(d) Weight (e) Height (f) Age

Figure 2: Distribution of the semantic parameters (weight, height, age) of females in the databases. The upper row shows parameter distribution for the shapes dataset; the lower row shows parameter distribution for the motion dataset.

3 DATA PREPARATION

3.1 Data Preprocessing

Since our model is purely data-driven, we first give insights of data preprocessing.

To train the model we need two sources of data: body scans to learn human appearance, or shape, and motion data, collected during recording session. These two databases contain shapes and motions, recorded from the different people, although the distribution of their semantic parameters are approximately the same.

The shape database (Hasler et al., 2009) contains body scans of 114 different subjects. Body scans were previously registered, such that aligned meshes, containing $N = 1002$ vertices, were produced.

Into each mesh we embed a skeleton (see Fig. 4), consisting of $K = 15$ joints, and morph meshes in such a way, that the skeletons of these meshes are exactly in the same pose. This is done to exclude variability in shapes due to slight differences in pose. We compute the skeleton's joints' positions as a linear combination of nearby vertices:

$$s_k = \sum_{i=1}^{N} \omega_{ik} v_i, \tag{1}$$

where weighs ω_{ik} were derived manually based on the procedure, that was used to place markers on the body during the creation of the motion database.

As a source of motion data, we used the motion database, described in (Troje, 2002); the database contains recordings of gait motion of 100 individuals. Firstly, motion was acquired using MoCap system with 45 markers. From 45 physical markers $K = 15$ virtual markers were derived; these markers correspond to the placement of skeleton joints inside the mesh. For each motion 4 so-called eigenpostures are extracted. Each motion is then represented as a linear combination of eigenpostures, where the coefficients are periodical functions, that depend on time. The walk cycle parameter is individual for each person and therefore is stored separately. That means, that all together each motion is represented with $K \cdot 3 \cdot (4+1)+1$ coordinates.

Moreover, for each person semantic data, such as *gender, weight, heights, age*, was stored. We denote semantic attributes, connected to each recorded shape s or motion m, as $v(s)$ or $v(m)$. Firstly, we separated male from female subjects in both databases, since variation between male and female meshes is much greater then the variation within the female (or male) part of the database itself and such separation makes our model able to better capture variation within both groups. Each group has the same number of members, i.e 55 females and 59 males in the mesh database and 50 females and 50 males in the motion database. The distributions of the semantic parameters of females are presented in Fig. 2. For our experiments, we took *weight* and *height* as the most varying parameters across our data sets, but in general any set of parameters could have been taken, as long as it has significant influence on shape appearance and on motion.

3.2 Binding a Skeleton to a Mesh

Since the motion data we obtained is represented using positions of joints depending on time, we chose to use a skeleton-based approach to animate the model. Several methods are proposed to solve the problem of skeleton-based character animation. As mentioned in the previous subsection, we embedded a skeleton into each mesh. To morph a mesh in accordance with a given skeleton position, we chose to use linear skinning (Jacka et al., 2007), since the technique is both fast and robust, while providing sufficiently good results. To summarize the implementation, we first assign weights w_{ij} to each vertex of the mesh, that characterize how much bone b_j influences the vertex v_i. Let a coordinate system be attached to each bone,

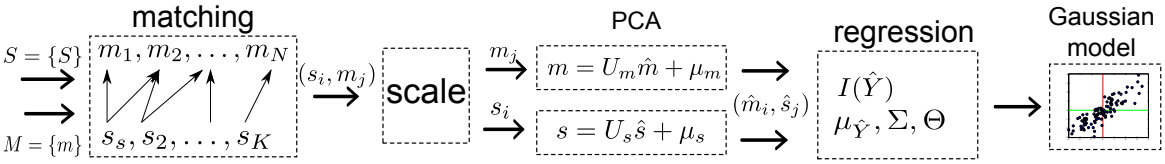


Figure 3: Block diagram of the algorithm; starting from the initial sets of motions (M) and shapes (S), Gaussian model is built.

then transformation matrix T_j transforms vertex coordinates from bone coordinate system to world coordinate system. For the new position of the bone, $\hat{T}_j$ denotes the new transformation matrix. Then the new position of the vertex is given by the formula:

$$\hat{v}_i = \sum_{j=1}^{K'} w_{ij} \hat{T}_j T_j^{-1} v_i, \quad \sum_{i=1}^{N} w_{ij} = 1 \qquad (2)$$

where $K' = K - 1$ is the number of bones (see Fig. 4). The weights are generated using the solver for the heat equation, where heat is propagated from each of the bones (Baran and Popović, 2007) until heat equilibrium is reached, and weights are set equal to the normalized temperatures.

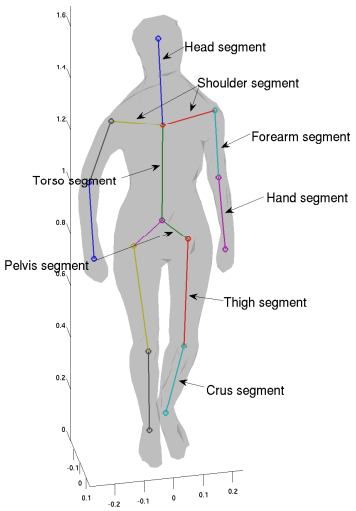


Figure 4: Alignment of a skeleton and a mesh.

4 APPEARANCE MORPHABLE MODEL

In this section, we explain our shape-motion model and its constraints and limitations.

The algorithm for model preparation and training is given in the Fig. 3. The main idea of our model is to correlate shape and motion based on semantic parameters. To achieve better alignment and to avoid the retargeting problem, we first match and align the meshes with the motions. A rough match is done based on semantic parameters, i.e. we create a set of shape-motion pairs (s_p, m_q), such that the values of the semantic parameters for each pair differ less than a given threshold ε : $P = \{(s_p, m_q), \|v(s_p) - v(m_q)\| < \varepsilon\}$ (this procedure corresponds to the first block in the Fig. 3). In the next step, we morph each shape s_p in the way, that its skeleton corresponds exactly to the skeleton of the motion m_q. To implement scaling, we modify Eq. (2) by adding a transformation matrix that represents relative scaling:

$$\hat{v}_i = \sum_{j=1}^{K'} w_{ij} \hat{T}_j S_j T_j^{-1} v_i, \quad \sum_{i=1}^{N} w_{ij} = 1, \qquad (3)$$

$$S_j = \begin{pmatrix} s_j & 0 & 0 \\ 0 & s_j & 0 \\ 0 & 0 & s_j \end{pmatrix} \qquad (4)$$

where s_j is a scaling factor. After scaling, we fit the skeleton again using Eq. (1). We iterate between these two steps several times before the bone lengths of the newly fitted skelton converge to the desired values. Since the scaling coefficients, i.e. ratio between corresponding bone lengths, are in the interval of $[0.85, 1.15]$, it does not affect the realistic look of the shapes. After scaling, the bone lengths of the mesh skeleton correspond exactly to those of motion skeleton.

Since the dimensionality of the data is still high after this preprocessing step, we first reduce the dimensions by performing Principal Component Analysis (PCA) on both shape and motion coefficients separately and converting our data into PCA space of smaller dimensions. PCA transformation consists of finding orthogonal directions in space (called basis), corresponding to the maximal variance of the sample. Then, given a set of vectors $x_j \in X \subset \mathbb{R}^l$, N_x size of the set X, and the found basis $U_x = [u_x^1, \ldots, u_x^k]$, the original vectors can represented as:

$$x_j = U_x \hat{x}_j + \mu(X), \quad j = 1 \ldots N_x \qquad (5)$$

where $\mu(X) = \frac{1}{N_x} \sum_{i=1}^{N_x} X_i$ is the mean of the set X and $\hat{x}_j$ is a vector of coordinates of x_j in PCA space, dimensionality of $\hat{x}_j$ is smaller than the dimensionality

of x_j. Then,

$$\hat{x}_j = U_x^T (x_j - \mu(X)) \tag{6}$$

is the reverse transformation from $\hat{x}_j$ to x_j.

To apply PCA on our data, we firstly stretch the matrices, representing shapes and motions, into vectors; therefore we produce two sets: the set of vectors representing shapes $S \subset \mathbb{R}^{3N}$ and the set of vectors representing motion $M \subset \mathbb{R}^{15K+1}$. Than we perform PCA as described above (see the third block in the Fig 3). We denote the PCA coordinates of smaller dimensionality as $\hat{m} \in \mathbb{R}^{\hat{N}}$ and $\hat{s} \in \mathbb{R}^{\hat{K}}$ accordingly. Dimensionality of the space is chosen in such a way, that leaves 95% of variance of the sample in both cases. Now we bind mesh and motion coordinates to learn a joint Gaussian distribution over a set $[\hat{s}, \hat{m}]$, depending on the semantic parameters v. In this sense, our model is close to Active Appearance Models (AAM, (Cootes et al., 2001)), although in contrast to AAM, the relation between two sets of PCA coordinates in our model is probabilistic, i.e. they are not firmly coupled together.

Since each pair has common semantic parameters, i.e. to each pair of PCA coordinates $\hat{y} = (\hat{s}, \hat{m}) \subset \hat{Y} \in \mathbb{R}^{\hat{N}+\hat{K}}$ a vector of semantic parameters v is assigned, we can now derive control 'handles' over our model. For that, we use linear regression in the space of joint PCA coordinates. Since not all coordinates are correlated with semantic parameters, we first perform standard correlation significance analysis and find significant coefficients: $I(\hat{Y})_r = \{i \in [1,\dots,\hat{N}+\hat{K}] : \mathrm{P}(\rho(\hat{y}_i, v) = 0) < \gamma\}$, where $\rho(\hat{y}_i, v)$ is thecorrelation between the i-th coordinate of vector y and semantic values, $\gamma = 0.05$ is p-value for testing the hypothesis that no correlation exists (Kendall and Stuart, 1973). We then use coordinates from $I(\hat{Y})_r$ to build the joint regression model:

$$\hat{y}_{I(\hat{Y})_r} = \Theta v + \varepsilon, \tag{7}$$

where Θ is a matrix of regression coefficients and $\varepsilon \sim \mathcal{N}(0, \Sigma_I)$ is a normally distributed random variable with covariance matrix Σ_I.

For the rest of the coordinates we assume a joint Gaussian distribution $\mathcal{N}(0, \Sigma_f)$, where Σ_f is the joint covariance and can be easily estimated from the data. The full joint model (the construction of the joint model corresponds to the last two blocks in Fig. 3) is described by the following equations:

$$\hat{y} = \mu_{\hat{Y}} + \sqrt{\Sigma}\xi, \quad \xi \sim \mathcal{N}(0, I), \tag{8}$$

$$(\mu_{\hat{Y}})_i = \begin{cases} (\Theta v)_k, & i \in I(\hat{Y}), \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

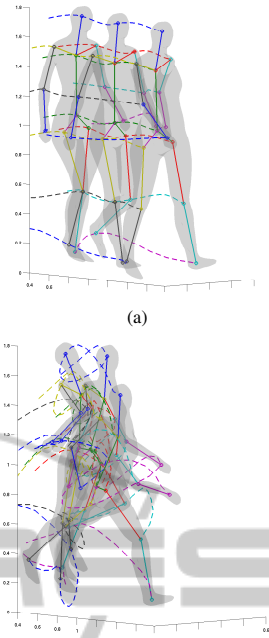


Figure 5: Two examples of conditional sampling of motion: (a) motion sampled with correlation analysis; (b) motion sampled without correlation analysis.

Here $(\Theta v)_k$ denotes the corresponding coordinate, obtained from Eq. (7). Without loosing generality, we can reorder indices in such a way, that indices in the sets $I(\hat{S})_r$ are put in the beginning, followed by indices of the elements, which are independent from semantic parameters. Then, the full covariance matrix can be written as:

$$\Sigma = \begin{pmatrix} \Sigma_I & 0 \\ 0 & \Sigma_f \end{pmatrix} \tag{10}$$

Using the proposed model for generating a character based on semantic parameters is straightforward: firstly, semantic parameters are chosen to determine the parameters of Gaussian distribution using Eq (7),(8),(10); then a random point in PCA space is sampled according to a normal distribution with determined parameters. Finally, the coordinates of the point are converted into shape-motion space using Eq (5).

Another useful property of the model is the ability to animate a given mesh, i.e. create the corresponding motion, given as well as create a well-fitting appearance, i.e. mesh, based on a motion. Let an input shape s be given. Firstly, the mesh is transformed in PCA space. Secondly, we derive a sampling distribution in the PCA space of $\hat{s}$ by applying Bayes theorem. We assume normal distributions and therefore are able to sample from the conditional distribution to derive the

mesh coordinates, corresponding to the motion and vice verse:

$$\mathsf{P}(\hat{m}|\hat{s}) = \frac{\mathsf{P}(\hat{m},\hat{s})}{\mathsf{P}(\hat{s}).} \tag{11}$$

Since we assume a parametrized normal distributions we can write:

$$\mu_{\hat{m}|\hat{s}} = \mu_{\hat{m}} + \Sigma_{\hat{m}\hat{s}}\Sigma^{-1}_{\hat{m}\hat{m}}(\hat{s}-\mu_{\hat{s}}) \tag{12}$$

$$\Sigma_{\hat{m}|\hat{s}} = \Sigma_{\hat{s}\hat{s}} - \Sigma_{\hat{s}\hat{m}}\Sigma^{-1}_{\hat{m}\hat{m}}\Sigma_{\hat{m}\hat{s}} \tag{13}$$

$$\hat{m} = \mu_{\hat{m}|\hat{s}} + \Sigma_{\hat{m}|\hat{s}}\xi, \quad \xi \sim \mathcal{N}(0,I) \tag{14}$$

Afterwards, we convert the PCA coordinates of the generated motion to the original motion space. One can apply the similar procedure to derive the mesh coordinates from a given motion.

As mentioned above, we separate coordinates, correlated with semantic parameters. We do it to avoid false dependencies between coordinates and to be able to use free coordinates to our model to introduce more variability into the generation process. To show the importance of separating correlated coordinates, we perform the following experiment: we divide the databases into two disjoint parts - test and train parts, train our model using train parts of each of the databases and then perform motion sampling conditioned on a mesh from the test part of the mesh database using the model with correlation analysis and without correlation analysis. As can be seen in the Fig. 5, the model that uses all coordinates for the regression (i.e. the model trained without correlation analysis) fails to generate suitable motion, while the model with correlation analysis produces realistic motion. We explain that by the fact, that coordinates of the mesh from the test set possibly are far from the coordinates of the meshes in train set, and therefore false dependencies introduced by regression on all coordinates have strong influence and produce implausible results.

5 EVALUATION

As stated above, our model allows to generate highly realistic shape-motion correlation with relatively small afford. In this section, we evaluate simulated sequences in terms of motion fidelity, and also provide quantitative evaluation to show how our approach allows to avoid the motion retargeting problem.

There is a number of methods for visual fidelity evaluation proposed in the literature. They can be divided into interrogation approaches and automatic approaches.

Interrogation-based approaches can be applied on very different problems. In (Reitsma and Pollard, 2003), experiments with human observers were conducted to determine the sensitivity of human perception to physical errors in motions and (Hodgins et al., 1998) investigated, which types of anomalies added to the motion disturb human perception the most. In (Pražák and O'Sullivan, 2011) and (McDonnell et al., 2008) human crowd variety was investigated and influence of either motion or shape clones on the whole crowd perception was investigated.

As an example of the automated motion evaluation approaches, (Ren et al., 2005) proposed and evaluated automated data-driven method for assessing unnaturalness of the human motion; in (Ahmed, 2004) motion is evaluated in terms of feet sliding and general smoothness, while in (Jang et al., 2008) physical plausibility of motion was evaluated. Unfortunately, as also mentioned by (Ren et al., 2005), these approaches depend a lot on the type of distortion, applied to motion, and therefore are directed on the evaluation of a specific type of motion generation.

Interrogation-based approaches provide more general results and are more reliable and more suitable for our goal.

Therefore, we performed two experiments designed as questionnaires and one additional experiment for quantitative evaluation. 13 male and 10 female subjects took part in our experiment. The experimental setup is described below.

5.1 Interrogation-based Evaluation

Experiment 1. In the first experiment, we asked the participants to compare two animations, one of which was generated using the model and the second one was generated by choosing motion and mesh randomly from our databases. We prepared 3 pairs of videos of 10 s. length. The parameters of the generated (or selected) characters are given in the Table 1. To diminish bias due to greater attractiveness of one figure in comparison to another, we chose appearances, that have approximately the same semantic parameters, e.g. compared two fat humans, or two thin humans etc.

The perspective and surroundings for the animated figures, as well as textures of the figures were the same. The participants were asked to indicate, which animation from the pairs looked more realistic; they also had the opportunity to state, that both animations looked equally bad or equally good. The results of the experiment are summarized in the Table 2.

As the results show, the mesh-to-motion fit delivers in general better results then just combining mesh and motion without taking into account param-

Table 1: Semantic parameters (weight in kg., height in cm.) of the samples used in the experiments.

	our model	selected mesh	selected motion
Experiment 1			
pair 1	(40, 169)	(53, 173)	(86, 176)
pair 2	(70, 155)	(44.7, 150)	(75.2, 186)
pair 3	(110, 170)	(120, 176)	(49, 168)
Experiment 2			
question 1	(90, 176)	(102, 176)	(86.3, 176)
question 2	(45, 169)	(49, 169)	(49.5168)
question 3	(77.9, 179)	(77.9, 176)	(77.9, 181)
question 4	(75, 178)	(75.6, 178)	(74.5, 179)

Table 2: Percentage of the participants, that accepted the sequence generated with our model as more realistic, less realistic and percentage of people, having difficulty to evaluate one of the motions as more natural.

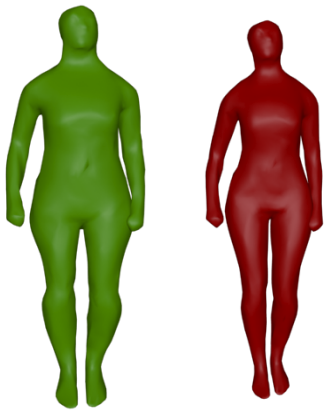
	pair 1	pair 2	pair 3	average
positive (%)	65.00	30.00	80.00	58.33
neutral (%)	15.00	15.00	10.00	13.33
negative (%)	20.00	55.00	10.00	28.33

eters of combined mesh and motion. We explain failure in the second pair of sequences due to generally lesser attractiveness of the appearance sampled from the model (see Fig. 6). In general, we observed, that mismatch between mesh and motion due to difference in weight is a lot easier to detect then mismatch due to difference in height, because in the latter case proportions of the person, unless some extreme cases are taken (e.g. a six-year child and a tall adult) are the same and therefore motion can fit quite well. Some more examples of subjects, sampled using our model, are given in the Fig 7.

Taking into account, that during artificial character animation it is usually difficult to find real motion data that fits exactly to the animated shape, our approach is beneficial in terms of delivered result.

Experiment 2. In the second experiment, we asked our participants to evaluate several sets of videos. In each set, four videos were presented. The videos in each set were produced as follows:

- A_1: an animation was generated by matching mesh and motion based on semantic parameters;
- A_2: an animation was generated by sampling mesh and motion from the model with the same semantic parameters;
- A_3: a mesh was taken from the database of meshes and a motion was sampled from our model given the mesh;
- A_4: an animation, where a motion was taken from the database of motion and a mesh was sampled from our model given the motion.



(a) (b)

Figure 6: Shapes used in the second pair of sequences in experiment 1 for comparison of visual fidelity (a) mesh sampled from the model (b) mesh from the original database.

(a) weight:40kg,(b) weight:61kg,(c) weight:77kg,(d) weight:110kg, height:169cm height:171cm height:178cm height:170cm

Figure 7: Examples of simulated characters.

The parameters of the samples are given in Table 1. The participants should grade each video with marks from 1 to 4, where 1 means very visually plausible and 4 mean completely unrealistic. In Table 3, the results are given in the form of percentage of participants, that evaluated sequences with 1, 2, 3 or 4 respectively.

As the results show, that our model produces slightly better results as match of semantic parameters, which is a doos result given that the two data bases were only matched via semantic parameters during training.The second most realistic animation is delivered by motion sampling, conditioned on the mesh, while appearance sampling (i.e. mesh sampling) based on motion generates the same or even slightly lower results, then semantic matching. We

Table 3: Percentage of the participants, given marks from 1 to 4 and mean note for each algorithm. The notes can vary in the interval [1,4], where 1 means very realistic, and 4 means not visually plausible.

alg.	1	2	3	4	average note
A_1	15.00	30.00	42.50	12.50	2.52
A_2	33.75	51.25	13.75	1.25	1.82
A_3	20.00	40.00	28.75	11.25	2.31
A_4	18.75	25.00	33.75	22.50	2.60

attribute this to the fact that motion still does not provide enough information to sample appearance accurately.

We explain the superiority of our sampling methods with several arguments: firstly, mesh and motion sampled with our model are aligned better to each other, and secondly, due to smoothing on the PCA stage, small artifacts, appearing occasionally in matched sequences, are smoothed away, and therefore visually plausible result can be produced.

5.2 Quantitative Evaluation

There exist several approaches to solve the motion retargeting problem, and none of them is perfect. However, we can avoid this problem by generating mesh and motion simultaneously. We evaluate the mesh-motion fit by comparing bone lengths of the mesh skeleton b_j^s and corresponding bone lengths of the motion skeleton b_j^m in terms of amount of scaling needed:

$$s_j = 1 - \frac{b_j^s}{b_j^m} \tag{15}$$

Here j corresponds to a body part and s_j denotes the amount of scaling required to fit mesh to motion. For the motion and mesh from the same person, i.e. when no scaling is required, the amount of scaling equals zero. As shown in the Figure 8, the model captures dependencies in bone lengths in mesh with bone lengths in motion and allows to generate pairs of mesh and motion already scaled properly, so that no scaling is required afterwards. We also evaluate mesh-motion fit in terms of conditional mesh sampling, when motion is given, and conditional motion sampling, when mesh is given. We create pairs from the real data using semantic matching and for the same range of semantic parameters we sample pairs from the model. The results (Fig. 8) confirm, that mesh and motion in the pairs, sampled from the model, are already properly aligned to each other.

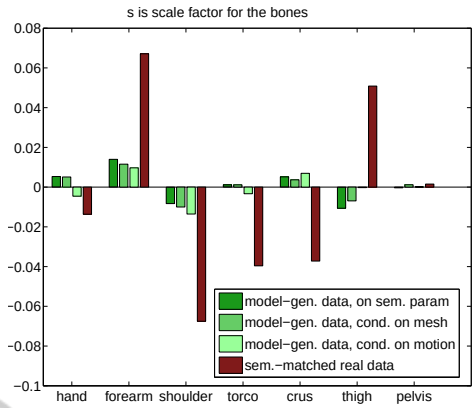


Figure 8: The mean amount of scaling required for each of the body parts; green color corresponds to the pairs sampled from the model, red color corresponds to the matched pairs. When no scaling is required, the amount of scaling equals 0.

5.3 Using the Model in Crowd Simulation

As mentioned above, we also propose to use the model for automatic character appearance and motion simulation of crowds. We leave out of scope of this paper the question of crowd behavior simulation, since it is an active area of research and numerous approaches exist (see, for example, (Guy et al., 2010), (Narain et al., 2009)). By controlling the distribution of the semantic parameters, it is possible to generate a set of character meshes, that is specific to the scene. In our example, we generate the crowd by randomly placing sampled characters in space. For creating the crowd in Fig. 9 and 10, we used a uniform distribution of the semantic parameters *weight* and *height*. However, one can consider more advanced ways of setting semantic parameters, if some specific distribution of shapes and motions is required.

More demonstrations are provided in the supplementary material.

6 FUTURE WORK

As our experiments have shown, there are several limitations of our model. First of all, our model was applied to a cyclical motion, so its extension to arbitrary type of motion can represent some difficulties. Secondly, the model can be used to generate samples based on semantic parameters, when the semantic parameters are not very far from the range of the parameters used for training. However, when it is not the case, produced results can look unrealistic. We hope, that increasing size and variability of the datatbases

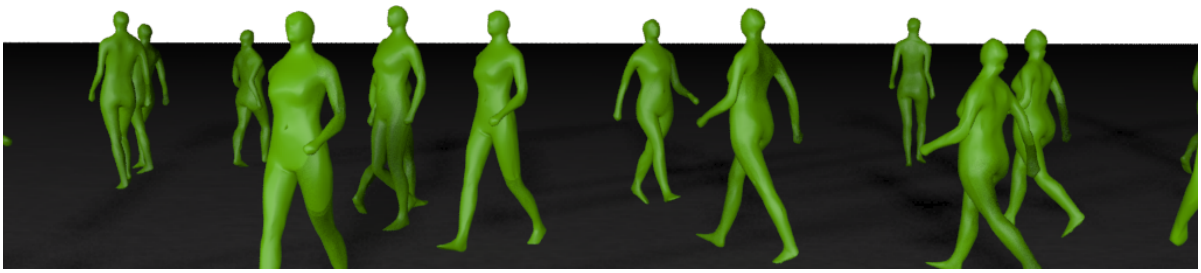


Figure 9: Example of crowd generation with semantic parameters *weight* and *height* uniformly distributed on $[40, 70]$ and $[150, 180]$ accordingly. The animation itself can be viewed in additional materials.

Figure 10: Example of crowd generation with semantic parameters *weight* and *height* uniformly distributed on $[50, 100]$ and $[150, 200]$ accordingly. The animation itself can be viewed in additional materials.

will help to generate characters with a wider range of semantic parameters.

Therefore, in future we are planning to extend our work on more complex motions, as well as achieve better mesh-motion fitting in the model training phase. More complex motions should firstly be aligned using techniques such as time warping (Hsu et al., 2007) and then the similar procedures as in (Troje, 2002) should be applied to produce vector-based representation of motions, suitable for analysis.

Another interesting topic is the use of more advanced stochastic models for mesh-motion coordinates coupling to be able to capture non-linear dependencies.

7 CONCLUSIONS

In our work, we proposed a parametrized model combining shape and motion, that can be applied to generate realistic characters together with appropriate motion. While providing enough variability, the model allows tight control over the generation process with the help of semantic parameters, and has a probabilistic formulation.

Although the usage of two independent databases can be seen as a weakness of our approach, we want to stress that the method was suggested to find dependencies in the initially unrelated databases and use them to bridge these databases. Such an approach can be advantageous when it is not possible or difficult to collect motion and shape data together from the same subjects.

Our model can be used to generate completely new meshes and motion, as well as to generate a specific motion for a given mesh or a specific shape using an existing motion.

We also evaluated our model in three experiments, that showed superiority of the proposed model in terms of visual plausibility of created samples in comparison to simple matching based on semantic parameters. We furthermore avoid the retargeting problem, since our model already contains necessary dependencies in it.

The model can be easily extended to bigger datasets, since all the algorithms used here in general are computationally inexpensive.

Possible applications of our model include animating existing characters, creating appearance based on motion, as well as creating a visually plausible crowd.

ACKNOWLEDGEMENTS

We would like to thank Prof D. Fleet for the very helpful discussions and good advice.

REFERENCES

Ahmed, A. A. H. (2004). *Parametric synthesis of human animation*.

Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., and Davis, J. (2005). Scape: shape completion and animation of people. *ACM Trans. Graph*, 24:408–416.

Baran, I. and Popović, J. (2007). Automatic rigging and animation of 3d characters. In *ACM SIGGRAPH 2007 papers*, SIGGRAPH '07, New York, NY, USA. ACM.

Brand, M. and Hertzmann, A. (2000). Style machines. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '00, pages 183–192, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co.

Cootes, T. F., Edwards, G. J., and Taylor, C. J. (2001). Active appearance models. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(6):681–685.

da Silva, M., Abe, Y., and Popović, J. (2008). Interactive simulation of stylized human locomotion. In *ACM SIGGRAPH 2008 papers*, SIGGRAPH '08, pages 82:1–82:10, New York, NY, USA. ACM.

Guy, S. J., Chhugani, J., Curtis, S., Dubey, P., Lin, M., and Manocha, D. (2010). Pledestrians: a least-effort approach to crowd simulation. In *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '10, pages 119–128, Aire-la-Ville, Switzerland, Switzerland. Eurographics Association.

Hasler, N., Stoll, C., Sunkel, M., Rosenhahn, B., p. Seidel, H., and Informatik, M. (2009). A statistical model of human pose and body shape. computer graphics forum28.

Hodgins, J. K., O'Brien, J. F., and Tumblin, J. (1998). Perception of human motion with different geometric models. *IEEE Transactions on Visualization and Computer Graphics*, 4(4):101–113.

Hsu, E., da Silva, M., and Popović, J. (2007). Guided time warping for motion editing. In *Proceedings of the 2007 ACM SIGGRAPH/Eurographics symposium on Computer animation*, SCA '07, pages 45–52, Aire-la-Ville, Switzerland, Switzerland. Eurographics Association.

Jacka, D., Merry, B., and Reid, A. (2007). A comparison of linear skinning techniques for character animation. In *In Afrigraph*, pages 177–186. ACM.

Jang, W.-S., Lee, W.-K., Lee, I.-K., and Lee, J. (2008). Enriching a motion database by analogous combination of partial human motions. *Vis. Comput.*, 24(4):271–280.

Kendall, G. and Stuart, A. (1973). *The Advanced Theory of Statistics. Vol.2: Inference and: Relationsship*. Griffin.

Lee, Y., Lee, S. J., and Popović, Z. (2009). Compact character controllers. In *ACM SIGGRAPH Asia 2009 papers*, SIGGRAPH Asia '09, pages 169:1–169:8, New York, NY, USA. ACM.

Li, Y., Wang, Y. L. T., and yeung Shum, H. (2002). Motion texture: A two-level statistical model for character motion synthesis. In *ACM Transactions on Graphics*, pages 465–472.

Liu, C. K., Hertzmann, A., and Popovic, Z. (2005). Learning physics-based motion style with nonlinear inverse optimization. *ACM Trans. Graph*, 24:1071–1081.

McDonnell, R., Larkin, M., Dobbyn, S., Collins, S., and O'Sullivan, C. (2008). Clone attack! perception of crowd variety. In *ACM SIGGRAPH 2008 papers*, SIGGRAPH '08, pages 26:1–26:8, New York, NY, USA. ACM.

Narain, R., Golas, A., Curtis, S., and Lin, M. C. (2009). Aggregate dynamics for dense crowd simulation. In *ACM SIGGRAPH Asia 2009 papers*, SIGGRAPH Asia '09, pages 122:1–122:8, New York, NY, USA. ACM.

Popovic, Z. and Witkin, A. P. (1999). Physically based motion transformation. In *SIGGRAPH*, pages 11–20.

Pražák, M. and O'Sullivan, C. (2011). Perceiving human motion variety. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization*, APGV '11, pages 87–92, New York, NY, USA. ACM.

Reitsma, P. S. A. and Pollard, N. S. (2003). Perceptual metrics for character animation: sensitivity to errors in ballistic motion. In *ACM SIGGRAPH 2003 Papers*, pages 537–542. http://www.odysci.com/article/1010112995491572.

Ren, L., Patrick, A., Efros, A. A., Hodgins, J. K., and Rehg, J. M. (2005). A data-driven approach to quantifying natural human motion. *ACM Trans. Graph*, 24:1090–1097.

Sidenbladh, H., Black, M. J., and Sigal, L. (2002). Implicit probabilistic models of human motion for synthesis and tracking. In *In European Conference on Computer Vision*, pages 784–800.

Sok, K. W., Kim, M., and Lee, J. (2007). Simulating biped behaviors from human motion data. *ACM Trans. Graph.*, 26(3).

Troje, N. F. (2002). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. In *Journal of Vision*, volume 2, pages 371–387.

Wang, J. M., Fleet, D. J., and Hertzmann, A. (2008). Gaussian process dynamical models for human motion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(2):283–298.