COMPUTER ANIMATION AND VIRTUAL WORLDS
Comp. Anim. Virtual Worlds 2014; **25**:303–311
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/cav.1599

SPECIAL ISSUE PAPER

Human motion variation synthesis with multivariate Gaussian processes

Liuyang Zhou[1]*, Lifeng Shang[2], Hubert P. H. Shum[3] and Howard Leung[1]

[1] City University of Hong Kong, Hong Kong
[2] Noah's Ark Lab, Huawei, Hong Kong
[3] Northumbria University, Newcastle upon Tyne , UK

ABSTRACT

Human motion variation synthesis is important for crowd simulation and interactive applications to enhance synthesis quality. In this paper, we propose a novel generative probabilistic model to synthesize variations of human motion. Our key idea is to model the conditional distribution of each joint via a multivariate Gaussian process model, namely semiparametric latent factor model (SLFM). SLFM can effectively model the correlations between degrees of freedom (DOFs) of joints rather than dealing with each DOF separately as implemented in existing methods. A detailed evaluation is performed to show that the proposed approach can effectively synthesize variations of different types of motions. Motions generated by our method show a richer variation compared with existing ones. Finally, our user study shows that the synthesized motion has a similar level of naturalness to captured human motions. Our method is best applied in computer games and animations to introduce motion variations. Copyright © 2014 John Wiley & Sons, Ltd.

KEYWORDS
human motion variation; human motion synthesis; semiparametric latent factor model; computer animation

***Correspondence**
Liuyang Zhou, Department of Computer Science, City University of Hong Kong, Hong Kong.
E-mail: liuyang.zhou@my.cityu.edu.hk

1. INTRODUCTION

There is a high demand for variation synthesis in character animation and game domains, as the quality of animation could be brought down by recurrent motion clones [1]. We observe that real humans usually perform the same action differently each time. Although motion-capture technology is able to record a person's movements, it is not practical to apply this technique to capture different variations of the same motion, as it is time-consuming and labor-intensive. As a result, it is essential to develop an effective method to generate many variations automatically based on a small set of example motions.

Human motion variation synthesis can be viewed as a special case of human motion synthesis [2]. Human motion synthesis has been widely studied for the purpose of enriching the motion database or generating motions according to certain user constraints. Style transferring is one way to synthesize a new style of motions by transferring style from one motion to another [3]. However, most style-transferring techniques cannot synthesize variations within the same style. Motion interpolation extracts style parameters to synthesize motions using interpolation [4]. However, style parameters are nonintuitive and usually difficult to be extracted. Motion editing is another way for motion synthesis; users can generate new motions through editing existing motions [5]. However, this technique usually requires a manual tuning procedure, which is not suitable for novice users. In other words, human motion variation synthesis can be considered as a one-to-many mapping procedure, while existing methods mainly focus on one-to-one mapping.

In this paper, we propose a novel generative probabilistic method to generate a large number of new variants based on a small set of example motions. We divide the kinematic skeleton into multiple partitions based on the human skeleton hierarchy, which not only reduces the complexity of human motion but also helps to model the relations between joints. We predefine the influence between joints within the same body partition based on the hierarchy of the skeleton structure. Such influence is translated into the conditional dependency relations between joints. The conditional probability distribution for each joint is calculated by the semiparametric latent factor model (SLFM) [6], which differs from the standard Gaussian process (GP) because SLFM can capture the dependency between multiple outputs. SLFM is an extension of the generally used univariate GP for regression problems involving multiple response variables. The basic idea of this model is using a set of basic GPs and then linearly mixing them to capture dependency that may exist among the output variables. In our situation, it can effectively capture the relations between different degrees of freedom (DOFs) within one joint. New variants can therefore be synthesized by sampling from the predicted distribution. Besides, compared with other nonparametric regression methods such as kernel regression, GP-based models can robustly learn from small training sets, and the parameters of kernel function can be optimized without relying on experimental cross-validation, such as the kernel width for kernel regression. Most importantly, our synthesized motions show more variations compared with existing methods, which is because we directly use the skeleton configuration feature (SCF) as the target instead of the generally used frame difference [7].

This paper is organized as follows. We will first introduce some related works in Section 2. Motion representation used in this paper will be described in Section 3. After that, we will introduce joint relation extraction based on the divided hierarchy skeleton structure and feature extraction in Section 4. In Section 5, we will detail about the distribution calculation via SLFM. The steps for human motion synthesis will be explained in Section 6. Model evaluation and user case study will be evaluated to compare our synthesized motions with motion-capture data in Section 7.

2. RELATED WORKS

Motion variation can be defined as the difference within the same type of motions. For instance, when a group of people perform a punch motion, how they strike their fists and the strength of their punches can be distinctively different from each other. Some previous methods were developed for motion variation generation by adding noise to the existing ones [8]. However, variation is not merely noise or error, but rather a functional component of data itself [9]. A number of methods for generating new motions from example motions have been developed during the past years for different applications. These methods can be roughly categorized as interpolation-based methods, linear statistical methods, and nonlinear probabilistic methods.

A number of methods that employ interpolation-related techniques to generate new motions from existing example motions have been developed, such as verbs and adverbs system [10], synthesizing physically realistic variations provided by interpolation [11] and generating reactive motions using momentum-based inverse kinematics [12]. However, such interpolation-based methods have to extract style or user-specified parameters, which may not be easily defined systemically. Unlike the methods mentioned earlier, our generative model can directly learn from the training data without extracting any interpolation parameter.

Linear statistical methods have been widely adopted for motion analysis and synthesis. For example, motions can be represented as principal components, and variants can be extrapolated by a set of best-approximated principal weights from variants of similar motions [13]. Min *et al*. [5] modeled the distribution of principal motions for interactive motion generation. Kim and Neff [14] used independent component analysis (ICA) to divide the stylistic example locomotion into submotion components. New stylistically different locomotions can be synthesized through the composition of the submotion components. Although a linear statistical method is intuitive and easy to implement, human motions are high-dimensional data that may behave in a nonlinear manner, and it is difficult to extract style parameters with a linear perspective.

Probabilistic-based methods have been used to create new motions for different applications, such as motion editing [15] and style-based inverse kinematics system [16]. Style-based inverse kinematics is developed upon GP latent variable model (GP-LVM), which is originally used to visualize high-dimensional data [17]. Lawrence *et al*. [18] extended GP-LVM through hierarchies. Wang *et al*. [19] augmented GP-LVM with a dynamic prior to obtain smooth trajectories in latent space. These GP-LVM-based models are powerful methods for human motion modeling, yet they are not suitable for our application as there is no guarantee that the latent space points are densely connected to synthesize human motion variations. In addition, GP-LVM is constructed based on the independent assumptions between different observations dimensions. So the correlations between DOFs of joints cannot be modeled. Statistical methods are one direction for motion modeling, but they seldom consider physical information of motions. Wei *et al*. [20] used GP to model the nonlinear mapping between kinematic states and generalized forces. In this way, physically valid motions can be generated, which satisfies user constraints.

Style learning and transferring can also be viewed as one way to generate variations. Many researchers have developed effective methods for style-based motion generation, including style machines based on hidden Markov models [21]. GP is used for style learning and transferring [22] [15] as it is a powerful tool for regression. Wang *et al*. [23] proposed a multifactor GP model for style–content separation. Hsu *et al*. [3] transferred the input motion to a new style while preserving the original content. Style transferring is very similar to our purpose of generating new variations. Yet our approach focuses more on generating variation within the same style of motion, which can be considered as a one-to-many problem.

Among all the related research, Lau's work [7] is the most relevant to ours. They introduced the dynamic Bayesian network to model spatiotemporal variation of human motion. Conditional dependency is learned by a nonparametric kernel regression technique from training data. Yet their method was highly affected by the tuning of kernel width and the selection of neighboring instances. Ma *et al*. [4] also used a Bayesian network to model the relations between user-defined style parameters and extracted variation parameters. However, the structure of their Bayesian networks has to be manually adapted for different motions, which reduces the generalization ability of this method. In this paper, the skeleton representation is divided into multiple partitions to obtain the dependency between joints. This partition-based structure is more general as it does not need to consider the specific type about the motion.

3. MOTION REPRESENTATION

Human motion-capture data are high-dimensional time series. The hierarchy structure of each frame can be represented as root positions and joint rotations:

$$f_t = \{p_0, q_0, q_1, \ldots, q_i\} \tag{1}$$

where p_0 and q_0 are the global world 3D positions and orientations of the root joint, q_i is the rotation of the ith joint with respect to its parent joint. We encode the joint angles with parameterized exponential maps [24]. The joints are highly correlated during movements based on the articulated skeleton structure. Dividing the skeleton into partitions, either to reduce the complexity for motion recognition [25] or for partial motion synthesis to enrich motion database [2] [14], has been effective. We therefore divide the human skeleton into five partitions based on the skeleton hierarchy as shown in Figure 1, which are right arm, left arm, right leg, left leg, and torso and head. There are two reasons for us to partition the skeleton. First, it helps us extract the relations between joints as the joints are more correlated within the same body part (e.g. wrist, elbow, and shoulder). Second, we only need to take into account the joints that are in the same partition when we consider one joint, which makes the system simpler.

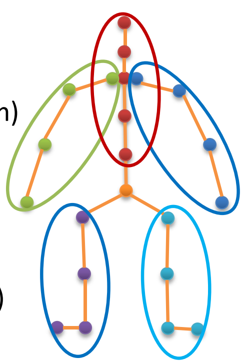


Figure 1. Five partitions of the skeleton: right arm (RA), left arm (LA), right leg (RL), left leg (LL), and torso and head (TH). The same color dots represent joints within the same partition.

4. MODEL CONSTRUCTION

In this section, we describe the way of specifying the dependency between joints with our partition-based skeleton structure and the features used in SLFM prediction model.

4.1. Partition-based Structure

Predicting the distribution of one joint based on some prior knowledge is the most important component for motion variation synthesis. In this work, we extract features from the ancestor joints of the current joint, which are defined as input features, as prior knowledge for prediction. By this, the relations between joints in the same partition are formalized as the conditional dependency between one joint and its ancestor joints.

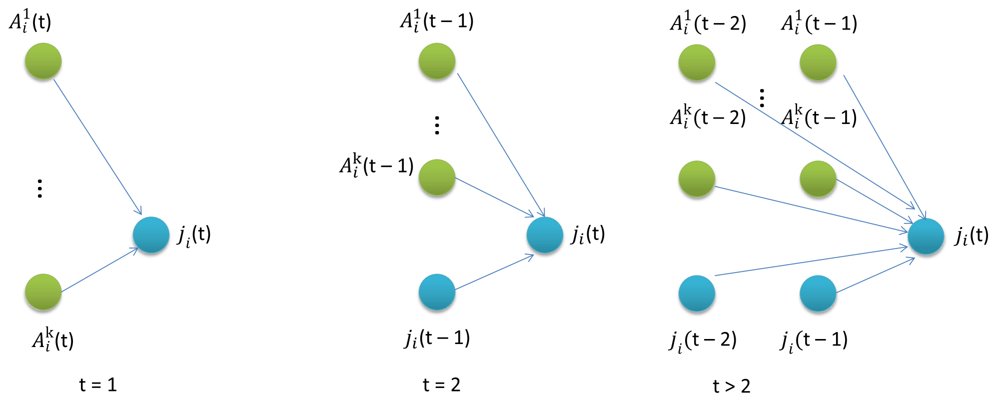
For a given joint j_i, its ancestor joints are defined as the set of all joints in the higher levels of the skeleton hierarchy within the same partition as j_i, which are denoted as A_i. Using the right wrist joint as an example, its ancestor joints are right elbow, right shoulder, and right clavicle. It should be noted that the top-level joint (e.g., right clavicle) has no ancestor joints; thus, $A_i = \emptyset$. Figure 2 shows the relationships between joint j_i and its ancestor joints. The blue dot represents the given joint j_i at time t, $t-1$, and $t-2$. The green dots are ancestor joints of j_i at time slices $t-1$ and $t-2$, which are denoted as $A_i(t-1)$ and $A_i(t-2)$, respectively. A_i^k represents the k-th joint in A_i. More specifically, A_i is one set of joints and A_i^k is one joint that belongs to A_i.

Starting from the third frame, we adopt a second-order temporal model to predict the distribution of one joint at the current frame. This is an observation we found in our initial experiments, in which the first-order model performs suboptimally. In particular, by analyzing the synthesized motions, we find that the movements tend to randomly move away from the original input data, as only one previous frame is not enough to constrain movement ranges, so we need to consider the information from more frames. To minimize the noise introduced by the newly added information, we have to select suitable features for prediction. In the following part, we will detail the feature selection for both input features and output features.

$$P(j_i[t]) = \begin{cases} P(j_i(t)|J_{pi}(t-1), J_{pi}(t-2), \dot{J}_{pi}(t-1), \dot{J}_{pi}(t-2)), & t > 3 \\ P(j_i(t)|J_{pi}(t-1), J_{pi}(t-2), \dot{J}_{pi}(t-1)), & t = 3 \\ P(j_i(t)|J_{pi}(t-1)), & t = 2 \\ P(j_i(t)|A_i(t)), & t = 1 \end{cases} \tag{2}$$

4.2. Feature Extraction

To facilitate the calculation of conditional distribution in this section, we use the term *parent joints* J_{pi} to represent the joints that will influence joint j_i. Except A_i, j_i itself will also influence j_i; thus, $J_{pi} = \{A_i, j_i\}$. The partition-based structure constructed earlier allows us to determine the conditional dependency between one joint and its parent



(a) (b) (c)

Figure 2. The graphical representation of the conditional dependency between joints. The blue dot represents joint j_i, green dots represent the ancestor joints of j_i. (a) At time $t = 1$; (b) at time $t = 2$; (c) at time $t > 2$.

joints, for example, $P(j_i(t)|J_{pi}(t-1), J_{pi}(t-2))$, where $t > 2$. P represents the conditional probability distribution. The semiparametric model SLFM will be used to model the conditional distribution for each joint. The performance of this model depends heavily on the input features, so it is important to define and select effective features.

In this work, we use the SCF [15] as the output feature to reconstruct poses. SCF is represented by joint angles and is parameterized by exponential maps [24]. We do not use a joint position as the joint representation, as constraining bone lengths defined in the skeleton hierarchy introduces extra system complexity. It should be noted that the output features of the root joint are the translation along the ground plane, and rotation around the vertical axis of the root relative to the root at the last frame. With this information, we can reconstruct the movement path by integrating the elapsed time and root transformation. Our output feature differs from that used by Lau *et al*. [7], in which the frame difference is used as the output of their regression model. However, the adjacent frames of one motion are often very similar owing to the high frame sampling rate, which limits the range of movement and hence the variation of the synthesized movement.

The SCF is also used as the component of the input features. In addition, a dynamic feature is introduced to convey the temporal information between frames. The dynamic feature is defined as the positional velocity between the current frame and previous frame. For the joints at the first frame, we only use SCF from its ancestor joints as the input feature. For the second frame of each joint, the input feature only includes SCF from its parent joints as there is no dynamic feature for the first frame. For the third frame, we use the SCF from the first two frames and the dynamic feature from the second frame as the input feature. Both the SCF and dynamic feature from the previous two frames are used as the input features for the subsequent frames. After defining source and output features, we use the SLFM to model the distribution of one joint at different time slices as given by Equation (2), where $\dot{J}_{pi}(t-1)$ and $\dot{J}_{pi}(t-2)$ represent the dynamic feature at time $t-1$ and $t-2$, respectively. In the following section, we will show the details of modeling the conditional distribution of one joint by SLFM.

5. COMPUTING CONDITIONAL DISTRIBUTION BY SEMIPARAMETRIC LATENT FACTOR MODEL

Estimating the conditional probability distributions of the multivariate variable $j_i(t)$ from the input features, that is, Equation (2), is the key component of our synthesis method. Traditional methods model this distribution using a parametric model, such as using a multivariate normal distribution and then optimizing its related means and covariance matrices by maximizing the posterior distribution of training instances. However, these methods involve many parameters and often suffer from local optima.

In this paper, we use the recently developed multivariate GP model SLFM [6] to estimate the distribution for each joint. In our case, our purpose is to predict the distribution of each joint. In our initial experimental testing, GP was adopted to synthesize variations. However, the movements tended to be ambiguous as GP cannot model the relations between the outputs. The DOFs of each joint are highly correlated, so it is more appropriate to consider them together rather than treating them independently. SLFM also inherits the advantages of the GP method, for example, SLFM can be robustly learned from small training data sets and the parameters of the similarity function can be optimized without relying on experimental cross-validation. Moreover, the variation trend of mean function can be easily adapted by changing the combination factors of different style kernel functions.

The structure of our SLFM for modeling the correlations between three DOFs of one joint is shown in Figure 3.

Comp. Anim. Virtual Worlds 2014; **25**:303–311 © 2014 John Wiley & Sons, Ltd.
DOI: 10.1002/cav

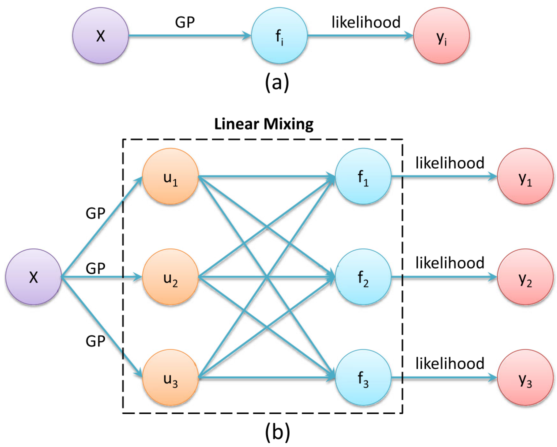


Figure 3. The graphical models of (a) standard Gaussian process and (b) our used semiparametric latent factor model for three degrees of freedom (DOFs) of a joint. X represents the input features, and y_i corresponds to the i-th DOF.

Similar to standard GP, each output y_i with $i \in \{1, 2, 3\}$ is independently generated from its own latent function $f_i(x)$. The difference between standard GP and SLFM is that $f_i(x)$ is a linear mixing of some basic GPs $u_i(x)$, which can capture the dependencies that exist among DOFs. The kernel function of latent function $f_i(x)$ is expressed as

$$f_i(x, x') = \sum_{p=1}^{3} \phi_{i,p}^2 u_p(x, x') \tag{3}$$

where $u_p(x, x')$ is the kernel function of the p-th GP for the input feature instances x and x' and $\{\phi_{i,1}, \phi_{i,2}, \phi_{i,3}\}$ are the mixing weights for $f_i(x)$. SLFM can be viewed as an augmented GP, which models output dependencies by sharing kernel hyper-parameters (i.e., the parameters of $u_p(x, x')$) of basic GPs. As a result, we can still use the same learning and prediction method as GP. The basic training procedure is as follows. First, we use a series of motion of the same motion type as coarse training data. Then, we use the feature extraction method described in Section 4.2 to extract SCF and dynamic features as the training input and output pairs of SLFM. Finally, using the conjugate gradient descent [26] methods, we obtain the optimal parameters for SLFM. A review of the inference and learning of GP can be found in [26].

6. MOTION SYNTHESIS

After obtaining the learned SLFM, we can synthesize a bunch of variations by sampling from the predicted distribution of each joint. Subsequent frames can be iteratively synthesized based on its previous synthesized frames and the learned conditional distribution model. For the joints that have no ancestor joints at the first frame, we sample from the data distribution directly. More specifically, we calculate the mean and standard deviation from the training data of these joints that have no ancestor joints at the first frame. In order to enhance the naturalness of the motions synthesized by our method, we apply the blending technique proposed in [27] to enhance the motion quality. Alternatively, automatic foot strike cleanup methods [28] can be applied to achieve a similar purpose.

7. EXPERIMENTAL RESULTS

In this section, we will evaluate the proposed method in different aspects. All experiments were conducted on a desktop computer with Intel Core 2 Duo 3.17-GHz processor. We implemented the system with C++/MATLAB. There were two data sets used in our experiments. The first one was the generally used HDM05 [29] database. This database consists of 130 different motion classes, with multiple trials performed by five subjects in each class. The motions were performed at a sampling rate of 120 Hz. We chose three types of motion from this database: walking, single-leg hopping, and jumping jack. On average, 0.17 second is required to synthesize 1 second of motion for this data set. The second data set included Tai Chi motions captured by an optical motion-capture system in our own laboratory. The frame rate is 60 Hz. We chose Tai Chi motion because of its large range of movement and its complicated movement features, which can be used to verify the robustness of the proposed method. The motions were performed 10 times by a professional Tai Chi Master. On average, 0.14 second is required to synthesize 1 second of motion for this data set.

7.1. Model Evaluation

For each type of the motion, we choose 10 example motions as training data set to learn SLFM. Figure 4 shows the synthesized Tai Chi and jumping jack motions by our method. We can observe that the synthesized motion preserves the major features of the original one but differs in both spatial and temporal domains.

For the Tai Chi motion, the range of movement varied at different phases. For the jumping jack motion, the heights of swing arms and jumping legs also varied among different variants. The synthesized variants of walking motion had different degrees of arm swing as well as stride length. For the single-leg hopping motion, the maximum height reached by the jumping motion varied across the variants. More animation results for all the motions can be found in our accompanying demo video. The readers are referred to the demonstration video for further animation results.

To verify that the synthesized motion is similar to the input one, we visualized the results by applying principal component analysis (PCA) on both motions. The visualization of the first PCA dimension is depicted in Figure 5. We can observe that the synthesized motions (the gray lines) have a similar variation trend as the original input data (the red line), so the motion type of the original input data can be well preserved, while variations are added. Other PCA

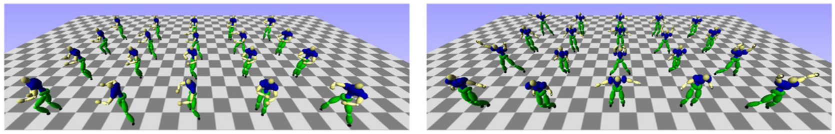


Figure 4. Human motion variants synthesized by our approach. Left side: the variants synthesized from Tai Chi motion. Right side: the variants synthesized from jumping jack motion.

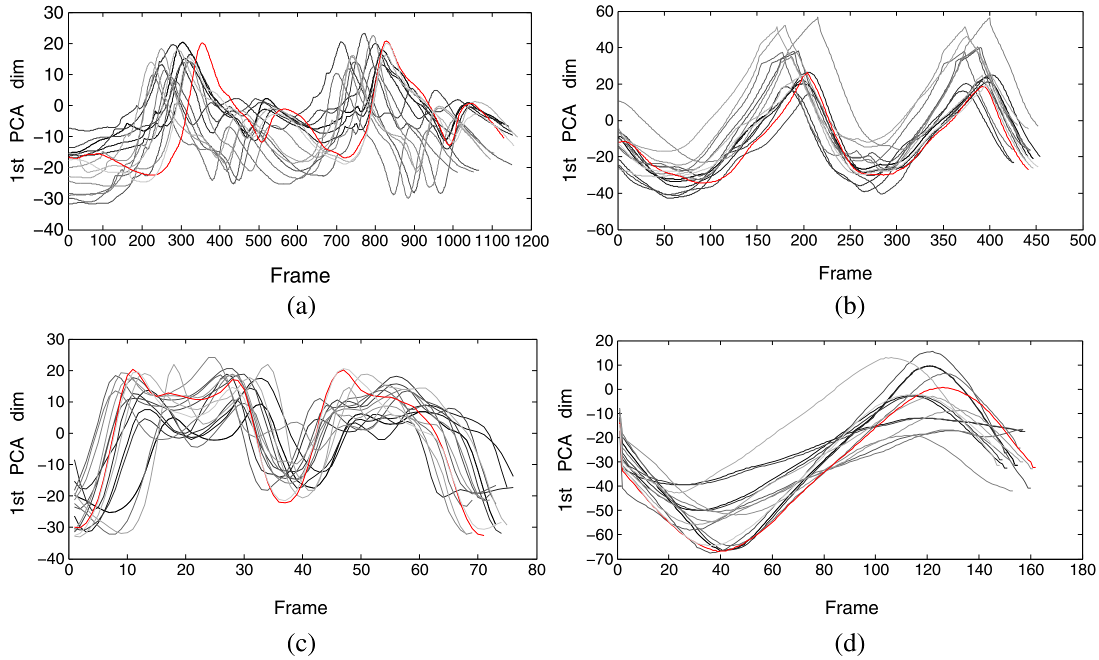


Figure 5. Plots of 15 variants and one of the training data. Each curve represents the first principal component analysis dimension of one motion, where the red curve represents one of the training data and others represent the synthesized results. (a) Tai Chi motion; (b) walking; (c) single-leg hopping; (d) jumping jack.

dimensions also have similar moving trends; because of space limitation, we will not show them here.

7.2. User Study

To evaluate the naturalness of the motions synthesized by our method, we compared the synthesized motions with motion-capture data by conducting a user study evaluation. A total of 10 participants were invited. All of them had little or no experience about motion-capture and 3D animation. We created a set of motions consisting of motion-capture data and synthesized motion data. Participants were asked to give a score for each motion based on its naturalness without knowing whether the displayed motion is captured or synthesized. The score ranges from 1 to 10 (inclusive), where 1 means the most unnatural and 10 means the most natural. We performed a two-way analysis of variance [30] on the obtained scores. There are two factors that influence the score, one is motion type (Tai Chi, walking, hopping, or jumping jack) and the other factor is method (motion capture or our method). The result is shown in Table I. There was no statistically significant

Table I. Two-way analysis result between the naturalness and two factors (method and motion type).

Source	SS	df	MS	*F*	Prob > *F*
Motion type	1.00	3	0.33333	0.67	0.5775
Method	6.45	9	0.71667	1.43	0.2069
Interaction	9.75	27	0.36111	0.72	0.8113
Error	20.00	40	0.50000		
Total	37.20	79			

SS; Sum of Squares, df; degree of freedom, MS; Mean Square.

difference in the perceived naturalness between motion types ($p = 0.5775$) and methods ($p = 0.2069$). There was also no statistically significant interaction between the two factors ($p = 0.8113$). It verifies that our method produces motions as natural as motion-capture data.

7.3. Comparison with Related Works

We also compared our method with other methods that can be used to generate motion variations. An intuitive way to generate variations is to directly add noise to the original motion data. Here, we adopt the generally used Perlin noise [8], and the synthesized variations can be found in the accompanying video. From the results, jerks can be observed during the movement. This is because variations in real human movement are not merely noise but are constructing components of the motion itself. We also synthesized motion variations with the method proposed by Lau *et al*. [7]. The relations between joints are modeled with a second-order dynamic Bayesian network. New motions can be synthesized by sampling from the predicted distribution with kernel regression. The motions synthesized by [7] showed less variations as they use the frame difference as the output feature while our approach use SCF as the output directly. Motion variations generated by our method and [7] can be found in the accompanying video for comparison.

8. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a novel generative probabilistic model to synthesize variations of human motion. Our focus differs from style-transferring methods, which transfer the style of one motion to another. To be specific, we focus on variation generation within the same style of motion. Our method is appealing for human motion variation synthesis because our approach can learn from small training sets and the parameters of motion model can be optimized without relying on experimental cross-validation. The usage of SLFM can model the relation between the DOFs of each joint, which is more natural for human motion modeling.

The skeleton representation is divided into multiple partitions to obtain the influence between joints. This partition-based representation not only reduces the complexity of human motion but also helps us define the influence between joints. The conditional dependency between joints is predefined for the joints within the same body partition based on the hierarchy structure.

We have demonstrated the effectiveness of our method by comparing the synthesized motions with motion-capture data. We applied a blending technique to enhance the naturalness of the resultant motion. We will conduct more experiments to evaluate the proposed method on data sets with different sampling rates. It will be also interesting to encode the contact information as a new feature into our model as described by Min *et al*. [31]. Moreover, SLFM allows different length scales for each input feature; this can automatically determine the importance of each input features to achieve feature selection.

Our approach may not perform well when applied in a long, complex sequence of movements owing to inaccurate time alignment. One simple solution is to carry out time warping for the output results to synthesize different lengths of motions. Another way is to extract the timing information beforehand as in [5] and add it back to the synthesized motion. Generating variations in the semantic level can be another possible future direction.

ACKNOWLEDGEMENT

The work described in this paper was supported by a grant from City University of Hong Kong (Project No. 7004045).

REFERENCES

1. McDonnell R, Larkin M, Dobbyn S, Collins S, O'Sullivan C. Clone attack! Perception of crowd variety. *ACM Transactions on Graphics* August 2008; **27**(3): 26:1–26:8.
2. Oshita M. Smart motion synthesis. *Computer Graphics Forum* 2008; **27**(7): 1909–1918.
3. Hsu E, Pulli K, Popović J. Style translation for human motion. *ACM Transactions on Graphics* July 2005; **24**(3): 1082–1089.
4. Ma W, Xia S, Hodgins JK, Yang X, Li C, Wang Z. Modeling style and variation in human motion. In *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. SCA '10, Madrid, Spain. Eurographics Association: Aire-la-Ville, Switzerland, Switzerland, 2010; 21–30.
5. Min J, Liu H, Chai J. Synthesis and editing of personalized stylistic human motion. In *Proceedings of the 2010 ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*. I3D '10, Washington, D.C. ACM: New York, NY, USA, 2010; 39–46.
6. Teh YW, Seeger M. Semiparametric latent factor models. In *Workshop on Artificial Intelligence and Statistics*, Barbados, 2005, 10; 333–340.
7. Lau M, Bar-Joseph Z, Kuffner J. Modeling spatial and temporal variation in motion data. *ACM Transactions on Graphics* December 2009; **28**(5): 171:1–171:10.
8. Perlin K. Real time responsive animation with personality. *IEEE Transactions on Visualization and Computer Graphics* 1995; **1**(1): 5–15.
9. Harris CM, Wolpert DM. Signal-dependent noise determines motor planning. *Nature* August 1998; **394**: 780–784.
10. Rose C, Bodenheimer B, Cohen MF. Verbs and adverbs: multidimensional motion interpolation using radial basis functions. *IEEE Computer Graphics and Applications* 1998; **18**: 32–40.
11. Safonova A, Hodgins JK. Construction and optimal search of interpolated motion graphs. *ACM Transactions on Graphics* July 2007; **26**(3): 402–408.
12. Komura T, Ho ESL, Lau RWH. Animating reactive motion using momentum-based inverse kinematics. *Computer Animation and Virtual Worlds* 2005; **16**(3–4): 213–223.
13. Urtasun R, Glardon P, Boulic R, Thalmann D, Fua P. Style-based motion synthesis. *Computer Graphics Forum*; **23**(4): 799–812.

14. Kim Y, Neff M. Component-based locomotion composition. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. SCA '12, Lausanne, Switzerland. Eurographics Association: Aire-la-Ville, Switzerland, Switzerland, 2012; 165–173.
15. Ikemoto L, Arikan O, Forsyth D. Generalizing motion edits with gaussian processes. *ACM Transactions on Graphics* February 2009; **28**(1): 1:1–1:12.
16. Grochow K, Martin SL, Hertzmann A, Popović Z. Style-based inverse kinematics. *ACM Transactions on Graphics* August 2004; **23**(3): 522–531.
17. Lawrence ND. Gaussian process latent variable models for visualisation of high dimensional data. In *In NIPS*, Vancouver, Canada, 2004; 2004.
18. Lawrence ND, Moore AJ. Hierarchical gaussian process latent variable models. In *Proceedings of the 24th International Conference on Machine Learning*. ICML '07. ACM: New York, NY, USA, 2007; 481–488.
19. Wang JM, Fleet DJ, Hertzmann A. Gaussian process dynamical models. In *NIPS*: Vancouver, Canada, 2005; 1441–1448.
20. Wei X, Min J, Chai J. Physically valid statistical models for human motion generation. *ACM Transactions on Graphics* May 2011; **30**(3): 19:1–19:10.
21. Brand M, Hertzmann A. Style machines. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '00, Louisiana, USA. ACM Press/Addison-Wesley Publishing Co.: New York, NY, 2000; 183–192.
22. Ma X, Le BH, Deng Z. Style learning and transferring for facial animation editing. In *Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. SCA '09, New Orleans, Louisiana. ACM: New York, NY, USA, 2009; 123–132.
23. Wang JM, Fleet DJ, Hertzmann A. Multifactor Gaussian process models for style-content separation. In *Proceedings of the 24th International Conference on Machine Learning*. ICML '07, Corvalis, Oregon. ACM: New York, NY, USA, 2007; 975–982.
24. Grassia FS. Practical parameterization of rotations using the exponential map. *Journal of Graphical Tools* March 1998; **3**(3): 29–48.
25. Deng L, Leung H, Gu N, Yang Y. Generalized model-based human motion recognition with body partition index maps. *Computer Graphics Forum* 2012; **31**(1): 202–215.
26. Shang L, Chan AB. On approximate inference for generalized Gaussian process models. *ArXiv e-prints* November 2013.
27. Feng AW, Huang Y, Kallmann M, Shapiro A. An analysis of motion blending techniques. In *The Fifth International Conference on Motion in Games*, Rennes, France, 2012; 232–243.
28. Kovar L, Schreiner J, Gleicher M. Footskate cleanup for motion capture editing. In *Proceedings of the 2002 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. SCA '02, San Antonio, Texas. ACM: New York, NY, USA, 2002; 97–104.
29. Müller M, Röder T, Clausen M, Eberhardt B, Krüger B, Weber A. Documentation Mocap database HDM05. *Technical Reports CG-2007-2*, Universität Bonn, 2007.
30. Thomas E. Miller, Rupert G. Jr.: Beyond ANOVA, basics of applied statistics. Wiley, New York, Chichester, Brisbane, Toronto, Singapore 1986, 317 s., 30.65; isbn 0-471-81922-0. *Biometrical Journal* 1988; **30**(7): 874–874.
31. Min J, Chai J. Motion graphs++: a compact generative model for semantic motion analysis and synthesis. *ACM Transactions on Graphics* November 2012; **31**(6): 153:1–153:12.

AUTHORS' BIOGRAPHIES

Liuyang Zhou is currently a PhD student in the Department of Computer Science, City University of Hong Kong. His research interests include character animation, human computer interaction, and machine learning.

Lifeng Shang is a Researcher of Huawei Noah's Ark Lab. He obtained his PhD degree in Computer Science from the University of Hong Kong in 2012. Before joining Huawei, he was a Postdoctoral Research Fellow at the City University of Hong Kong 2012–2013. His research interests include statistical machine learning, computer vision, and information retrieval.

Hubert P. H. Shum is a Senior Lecturer (Associate Professor) at Northumbria University. Before joining the university, he worked as a Lecturer in the University of Worcester, a Postdoctoral Researcher in RIKEN Japan, as well as a research assistant in the City University of Hong Kong. He received his PhD degree from the School of Informatics in the University of Edinburgh. His research interests include character animation, machine learning, human computer interaction, and physical simulations.

Howard Leung is currently an Assistant Professor in the Department of Computer Science at City University of Hong Kong. He received a BEng degree in Electrical Engineering from McGill University, Canada, in 1998, an MSc degree and a PhD degree in Electrical and Computer Engineering from Carnegie Mellon University in 1999 and 2003, respectively. His current research projects include interactive applications with motion capture, synthesis of single-character motions with variations, synthesis of multiple character interacting motions, brain informatics, intelligent tools for Chinese handwriting education, and Web-based learning technologies. He is also supervising the 3D Motion Capture Laboratory (http://mocap.cs.cityu.edu.hk), which promotes the 3D motion-capture technology for research and teaching. At present, he is the Treasurer of the Hong Kong Web Society.