Virtual Human Motion Extension Based on Bayesian Network

Shi Qu*, Tinxin Xu, Liang Ma and Jianxun Liu
Department of Space and Air Early Warning, Air Force Early Warning Academy, Wuhan, China

*Corresponding author

Abstract. Aiming at the problem of virtual human motion reuse, the concept of motion extension is proposed, which includes lengthways extension and transverse extension. The lengthways extension extends motion along the time, and generates more frames from existing motion. The transverse extension extends motion to other characters, and multiple characters motion is generated by single character motion. Each motion in the group is similar and different from each other. Motion extension provides unified solution to important issues such as motion prediction, motion repair, and group motion. We construct the Bayesian network of the virtual human motion, which studies and reduces the dimension of the virtual human motion, and propose a method of motion extension based on Bayesian network. Experiments show that our method can generate more realistic virtual human motion based on existing motion, and provide support for the application of virtual human in the fields of bio-engineering, medicine and military.

1. Introduction
In recent years, virtual human motion synthesis has been widely used in fields such as film, television and military. Because of the complexity of the human body structure and the variability of the motion, the human body motion is also constrained by its physiological characteristics and the external environment, resulting in virtual human motion synthesis is a very heavy and difficult work. With the development of motion capture technology, the data-driven method has achieved unprecedented development. Compared with physical simulation motion synthesis method, data-driven motion synthesis method is a new mode and a fundamental change in methodology and thinking mode. The data-driven method is a more direct method. It avoids the cumbersome operation of determining the angle of the virtual human joint during the synthesis movement, directly assigns the motion capture data to the virtual human geometry model, and drives the virtual human to complete the corresponding movement based on data. It has the advantages of simple implementation, high computing efficiency, and strong sense of reality. However, with the application of virtual people becoming more and more widespread and in-depth, new difficulties have emerged in data-driven methods, which are mainly reflected in three aspects: First, the movement capture data is highly specialized, and the given movement capture data is only suitable for generating corresponding movements., Even if there are only a few changes, the entire action needs to be re-captured; Second, the contradiction between the limitation of motion capture data and the diversity of application needs still exists, and the limited motion capture data cannot meet the increasing practical application needs. The third is that some special sports are difficult to capture, high risk or even not feasible. In order to solve these problems, this paper proposes the concept of motion extension, which can generate more abundant motion from limited motion data, and proposes motion extension methods, including vertical extension and horizontal extension.
2. Related Work
As the complexity of character skeleton and motion, motion data are often high dimensional. It is often required to reduce dimension of motion data before analysis and research on them. According to mapping relation between high dimensional data and low dimensional data, dimensionality reduction techniques can be classified to linear and nonlinear.
Linear dimension reduction methods have advantages in computational efficiency, including factor analysis[1], principal component analysis[2], probability principal component analysis[3][4], and independent component analysis[5].
Linear dimension reduction methods generally try to find a linear subspace and project the original high-dimensional data to this subspace to achieve dimensional reduction. Linear dimension reduction algorithm is relatively simple, so it has higher dimensional reduction efficiency. However, there are also many limitations, mainly because the manifold structure where the high-dimensional data is located is not considered. In recent years, nonlinear dimensional reduction methods have attracted people's attention, especially after the 2000 Science magazine published two consecutive papers on nonlinear dimensional reduction methods[6][7]. The study of nonlinear dimensional reduction methods has been highly valued by researchers. Documentation[6] proposed the isometric mapping method. First, the nearest neighbor set of each point is found, and then the Dijkstra algorithm is used[8]. Look for the shortest Geodetic path between the near neighbors and get the geodesic distance between the points sets instead of the Euclidean distance that can not represent the internal manifold structure, and then enter it into Multidimensional Scaling Analysis(MDS)[9]. In the processing, the low-dimensional coordinates embedded in the high-dimensional data space are found to achieve lower dimensions. Isomap's limitation is that due to the inaccuracy of estimating the geodesic distance, poor results may be obtained in unsampled space. Documentation[10] proposed a local linear embedding method(LLE). First, the nearest point of each point is found, and then a set of weights is calculated for each point, and the current point is represented by the weighted combination of nearest points. Finally, the low dimensional embedding of the data is found by means of eigenvector optimization to keep the weight of each point unchanged. The LLE algorithm retains the relationship between neighboring points when the dimension is reduced, so that the inherent geometric structure can be preserved. Not only can the nonlinear structure of high dimensional data be found effectively, but it also has invariable characteristics such as translation and rotation. The LLE algorithm is superior to Isomap on many issues, mainly because the LLE algorithm uses sparse matrix algorithm to achieve faster and more efficient optimization. The disadvantage of the LLE algorithm is that it has poor performance on non-uniform data regions, which is due to changes in data density that will cause changes in weights. Documentation[11] combined PCA and nuclear method, the Kernel Principal Component Analysis(KPCA) is proposed. The basic idea of KPCA is to map high-dimensional data to a feature space through nonlinear mapping, and then perform principal component analysis in the feature space. KPCA has the advantage of a kernel method, that is, establishing a mapping relationship between high-dimensional data and low-dimensional data through a kernel function without explicitly determining the mapping relationship. KPCA is a non-linear low-dimensional method. Gaussian Process Hidden Variable Model(GPLVM)[12][13]. It is another effective nonlinear dimension reduction method. Unlike PPCA, it does not estimate the specific values of the mapping parameters through Gaussian process regression to realize the dimensional reduction and density estimation of high-dimensional data. Instead, it starts from the Bayesian point of view. A priori, Marginalize the mapping parameters and realize dimension reduction by maximizing the posterior probability.
3. Definition of Motion Extension
In this paper, the concept of motion extension is proposed, which includes lengthways extension and transverse extension. Transverse extension extends motion to characters and generates multi-character group motion from single character motion. Each motion in group motion is similar and different from each other; Lengthways extension is an extension along the time, with new motion that generates more frames from existing motion, and the number of frames for new motion can be customized. Motion extension provides unified solution to important issues such as motion prediction, motion repair and group motion. Transverse extension can solve group motion problem, and lengthways motion
extension can solve motion prediction and motion repair problems. We give a strict definition of lengthways and transverse extension of motion:

Transverse extension: Given the motion \( Y_1: \{ y_1, y_2, \ldots, y_N \} \) and the integer \( L \), \( L \) motions \( Y_2: \{ y_1, y_2, \ldots, y_N \} \), \( y_{i+1} \) are generated, making the \( L \) motions and \( Y_1 \) similar to each other, while there are differences in detail. This process is called motion transverse extension.

Lengthways extension: Given the motion \( Y_1: \{ y_1, y_2, \ldots, y_N \} \) and the integer \( M \), \( M \) motion positions \( y_{N+1}, y_{N+2}, \ldots, y_{N+M} \) are generated, making \( Y_2: \{ y_1, y_2, \ldots, y_N, y_{N+1}, y_{N+2}, \ldots, y_{N+M} \} \) a coherent and unified motion. This process is called motion lengthways extension.

4. Bayesian Network for Virtual Human

Gaussian process dynamical models (GPDM) is a specific latent variable model, which implements dimensionality reduction for time sequence with Gaussian process regress and models the dynamics for time sequence in latent space. Given time sequence \( Y = [y_1, y_2, \ldots, y_N]' \) in observation space where \( y_i \) is D-dimensional column vector, \( i = 1, 2, \ldots, N \), and \( X = [x_1, x_2, \ldots, x_N]' \) is corresponding representation in low dimensional latent space of \( Y \) where \( x_j \) is q-dimensional column vector, \( j = 1, 2, \ldots, N \), GPDM comprises two mappings as shown in Figure 1. One is the observation mapping from latent space to observation space parameterized by \( B \); The other is the dynamics mapping in latent space parameterized by \( A \). From the viewpoint of Bayesian, marginalize out the parameters \( A \) and \( B \) by integrating to obtain the joint distribution of latent variable and observation variable as shown in Figure 1b.

![Figure 1. Bayesian network for virtual human.](image)

The observation mapping, based on scaled Gaussian process latent variable models (SGPLVM), completes dimensionality reduction for time sequence in high dimensional observation space. The dynamics mapping in latent space is based on 2-order Markov chain. The two mappings can be described as

\[
g: \quad x_t = \sum_j a_j \psi_j(x_{t-1}, x_{t-2}) + n_{x,t} \tag{1}\]

\[
h: \quad W y_t = \sum_j b_j \phi_j(x_t) + n_{y,t} \tag{2}\]

Where \( g \) is the dynamics mapping from last two latent states \( x_{t-1}, x_{t-2} \) to next latent state \( x_t \), parameterized by \( A = [a_1, a_2, \ldots]' \); \( h \) is the observation mapping from latent space to observation space, parameterized by \( B = [b_1, b_2, \ldots]' \). \( y_t \) is weighted by \( W = \text{diag}(\omega_1, \omega_2, \ldots, \omega_p) \) to make a difference among each dimensionality, \( \psi_j \) and \( \phi_j \) are nonlinear basic function, \( n_{x,t} \) and \( n_{y,t} \) are isotropic Gaussian noise. \( g \) and \( h \) are implemented with Gaussian process regress, which describes probabilistic distribution of function. From viewpoint of Bayesian, we do not estimate the precise value of parameter \( A \) and \( B \) like traditional fitting but lay emphasis on the uncertainty about them. Gaussian process is determined by mean function and kernel function (covariance function) similar to that
Gaussian stochastic variable is determined by mean and covariance. Training data has been pretreated minus mean, so mean function is set to zero. In order to implement Gaussian process, maximize the likelihood of training data by tuning parameters in kernel function.

Given the prior for hyper parameters \( p(\alpha, \beta) \propto \prod_i \alpha_i \prod_j \beta_j \), the joint probability of all variable is defined as

\[
p(X, Y, \alpha, \beta) = p(Y | X, \beta, W)p(X | \alpha)p(\alpha, \beta)
\]

### 5. Virtual Human Motion Lengthways Extension

The technique of generating more frame motion can automatically generate more frame new motion based on existing motion, which looks realistic and the number of frames can be controlled by user easily. In last section, we map the motion data from observation space to latent space based on GPDM:

\[
Y = [y_1, y_2, \cdots, y_N] \rightarrow X = [x_1, x_2, \cdots, x_N]^T
\]

The mean prediction, which predicts with probability mean, is used to predict latent states in latent space. At the time of \( t \), the latent state \( x_t \) is Gaussian

\[
x_t \sim N(\mu_x([x_{t-1}, x_{t-2}]), \sigma_x^2([x_{t-1}, x_{t-2}]))I
\]

According to mean prediction, given the latent states \( x_{t-1}, x_{t-2} \) at the time \( t-1 \) and \( t-2 \), the latent state at the time \( t \) can be predicted as

\[
x_t = X_{k_t} \cdot K^{-1}_x k^t_x([x_{t-1}, x_{t-2}])
\]

Similarly, the mean prediction is also used to predict \( y_t \).

\[
y_t = Y^T \cdot K^{-1}_y k^t_y(x_t)
\]

### 6. Virtual Human Motion Transverse Extension

Transverse motion extension extends motion along characters, and it generates multiple character group motion from single character motion automatically. Multiple character group motion is that many characters performance the same motion together, in which every character motion is similar to each other and there is difference among them. To implement transverse motion extension, train GPDM with single motion to obtain corresponding latent states which form a smooth latent trajectory, then, sample near it in latent space to obtain more latent trajectories with hybrid Monte Carlo (HMC). After optimize new latent trajectories, map them from latent space to observation space to generate new motions which form group motion together. Optimize latent trajectory \( X \) according to

\[
X = \text{Max}\{p(X | \alpha)\}
\]

The new latent trajectories sampled by HMC lie near the original latent trajectory of training data, and the motion poses are similar if their latent states are nearby, so the new motions reconstructed by new latent trajectories are similar to the training motion, but there is difference among them. This is suitable for multiple character group motion.

### 7. Results and Analysis

In this section, we design two experiments to prove the algorithm in this paper correct and effective. The training motions in experiments come from the Carnegie Mellon University motion capture database. In our experiments, training motion data is 50 dimensional, and the latent space is 3 dimensional, that is \( D = 50, q = 3 \).
7.1. Motion Lengthways Extension
In this experiment of motion lengthways extension, the training motion is the running motion which consists of 290 frames as shown in Figure 2(a). In order to compare, we take the first 145 frames as training data to extend, and compare the extension result with the last 145 frames. Train GPDM with the first 145 frames to obtain corresponding latent states, and extend the latent trajectory to 290 frames in latent space.

After extend latent trajectory in latent space, map the extension result to observation space to generate new motion as shown in Figure 2, in which (a) is the training motion, (b) is the new motion extended from the first 145 frames of training motion, (c) is the detail at the start of extension. Figure 2(a) and 2(b) are sampled each 20 frames, and Figure 2(c) is sampled each 5 frames. From Figure 2(c), the extended frames and training frames hang together very well. From Figure 2(b), the extended frames are all new poses not coming from the training motion. By comparison of Figure 2(b) and 2(a), we can see that the extended frames are similar to the corresponding training frames, so the extended motion obeys the motion rule well.

![Figure 2. Lengthways motion extension.](image)

7.2. Motion Transverse Extension
In this experiment of motion transverse extension, the training motion is the football motion which consists of 362 frames as shown in the top line of Figure 3. Train GPDM with it to obtain corresponding latent states, and extend the original latent trajectory to 5 new ones by hybrid Monte Carlo sampling in latent space.

After extend in latent space, map the extended latent trajectories back to observation space to generate new motions as shown in Figure 3, in which the extended motions are similar to the training motion, but there is difference among them. This is suitable for multiple character group motion.

![Figure 3. Transverse motion extension.](image)
8. Conclusions
The motion extension technique, proposed in this paper, can generate more frame motion and multiple character group motion based on existing single character motion. Motion extension does not repeat existing motion by rote but generate new motions following the motion rule of training data, which is depicted by probabilistic distribution of motion pose. The extended frames do not exist in training motion, and the new motion is similar to training motion but there is difference among them. Motion extension can extend the applicable field of existing motion.

Acknowledgments
Thanks to the anonymous reviewers for their helpful comments, Neil Lawrence for his on-line source code, and Carnegie Mellon University for human motion capture data.
This work is supported by the National Nature Science Funding No. 61602506, No. 11805278 and the Hu Bei Province Nature Science Funding No. 2016CFB307.

References
[1] Moby Parsons, Howard D. Routman, etc. Patient-reported outcomes of reverse total shoulder arthroplasty: a comparative risk factor analysis of improved versus unimproved cases[J]. JSES Open Access, 2019, pp. 174-178.
[2] Lanlan Feng, Yipeng Liu, etc. Robust block tensor principal component analysis[J]. Signal Processing, 2019, pp. 132-137.
[3] Sun Hongwei, Gu Ming and Sun Jiaguang, “A Coding Algorithm Using PCA-Based Correlation Vector Quantization,” Journal of Computer-aided Design & Computer Graphics (17:8), 2005, pp.1662-1666.
[4] Gift N, Lorraine B, Isobel C G. Probabilistic principal component analysis for metabolomic data[J]. BMC Bioinformatics, 2010, 11(1): 571-582.
[5] Jonatan Henriquez, Werner Kristjanpoller. A combined Independent Component Analysis–Neural Network model for forecasting exchange rate variation[J]. Applied Soft Computing Journal, 2019, pp.36-42.
[6] Wang, J. M., Fleet, D. J., Hertzmann, “A Gaussian Process Dynamical Models,” In Proc. NIPS 2005. Vancouver, Canada. 2005, pp.1441-1448.
[7] Y. Yacoob and M.J. Black, “Parameterized Modeling and Recognition of Activities,” Computer Vision and Image Understanding (73:2), 1999, pp.232-247.
[8] Keith Grochow, Steven L. Martin, Aaron Hertzmann, Zoran Popović. “Style-based Inverse Kinematics,” ACM Transactions on Graphics (Proceedings of SIGGRAPH 2004) (23:3), 2004, pp.522-531.

[9] Pedro G. Vieira, Marcelo M.R. de Melo, etc. Quercus cerris extracts obtained by distinct separation methods and solvents: Total and friedelin extraction yields, and chemical similarity analysis by multidimensional scaling[J]. Separation and Purification Technology, 2019, pp. 289-296.

[10] Huang Lili. Real Time Localization Algorithm based on Local Linear Embedding Optimization in Mobile Sensor Networks[M]. 2019 7th International Conference on Machinery, Materials and Computing Technology, 2019, pp.658-662.

[11] R.Rosipal, M.Girolami, L.J. Trejo, A. Cichochi, “Kernel PCA for Feature Extraction and De-Noising in Non-linear Regression,” Neural Computing & Applications (10:3), 2001, pp. 231-243.

[12] Keith Grochow, Steven L. Martin, Aaron Hertzmann, Zoran Popović. “Style-based Inverse Kinematics,” ACM Transactions on Graphics (Proceedings of SIGGRAPH 2004) (23:3), 2004, pp.522-531.

[13] QU Shi, WEI Ying-mei, KANG Lai, WU Ling-da. “Pose Synthesis of Virtual Character Based on Statistical Learning,” The International Symposium on Computer Network and Multimedia, Wuhan, China, 2009, pp.36-39.