Expressive Gait Synthesis Using PCA and Gaussian Modeling

Joëlle Tilmanne and Thierry Dutoit

TCTS Lab, University of Mons, Mons, Belgium,
joelle.tilmanne@umons.ac.be

Abstract. In this paper we analyze walking sequences of an actor performing walk under eleven different states of mind. These walk sequences captured with an inertial motion capture system are used as training data to model walk in a reduced dimension space through principal component analysis (PCA). In that reduced PC space, the variability of walk cycles for each emotion and the length of each cycle are modeled using Gaussian distributions. Using this modeling, new sequences of walk can be synthesized for each expression, taking into account the variability of walk cycles over time in a continuous sequence.

Keywords: motion synthesis, expressivity, PCA, variability

1 Introduction

Modeling of all kind of human behaviors is a very challenging field of study, as those behaviors which are so natural for the human eye are very often extremely difficult to model, and even more difficult to mimic. It is also the case for human motion, which is a complex phenomenon involving our physiological structure as well as our capacity to adapt to external constraints and to feedbacks from our body.

In the field of virtual human animation, various approaches can be taken to synthesize realistic human motion. In particular, there has been a lot of interest in the ways of using and re-using motion capture data [5], a technology that brings the movements of real humans into the virtual world. The main problems encountered with motion data are its variability and its high dimensionality; which make it hard to retrieve, analyze, adapt and modify motion patterns either made "on demand" or coming from an existing motion database.

Two main approaches are encountered regarding the use of motion capture data for animation. The first one consists in building a database, developing techniques to retrieve motion parts in this database, editing these motion parts if needed, and blending them together [6].

The second one uses various machine learning techniques in order to build models based on training motion capture data. The models can later be used to synthesize new motion sequences without resorting to the database initially used for training [2, 7, 1].

In this paper, we focus on the second approach, as we model walk cycles performed with eleven different styles and the cycle variation over time during each walk sequence, using a finite number of parameters. Contrarily to most studies addressing this subject, we do not only model style variations but also the variability of motion cycles over time, as these variations are an intrinsic part of the plausibility of the synthesized motions.

Our method, based on the method by Glardon et al. [7, 8] uses PCA (principal component analysis) to reduce the dimensionality of each walk cycle and to model the different style components. A Gaussian modeling of the data represented in the PC space is then conducted and enables us to model the variability of the walk cycles over time and thus to introduce some randomness in the synthesized sequences.

This paper is organized as follows. Section 2 makes a brief review of related work. The recording of the database is then presented in Section 3 and is followed by the preprocessing of the data prior PCA in Section 4. Section 5 presents the PCA of the original data and in Section 6 we explain how the PC subspace was modeled. Section 7 presents how this modeling enabled us to extrapolate and synthesize new motion sequences and the results are discussed in Section 8. Section 9 will conclude this paper by presenting future work.

2 Related work

2.1 PCA for dimensionality reduction

Principal component analysis (PCA) [13] is a widespread technique, consisting in finding, by rotation of the original set of axis, the best set of orthogonal axis (corresponding to principal components) to represent the data. Variables in the PC space are thus uncorrelated. The principal components are ordered so that the first ones retain most of the variation present in all of the original data. This techniques is thus often used to reduce the dimensionality of the original data and facilitate further processing.

This approach is widely used as a first step in motion data analysis and synthesis, mainly to reduce the dimensionality of the data vector needed to describe the pose of the character at each frame (see for instance [17, 2, 4, 10]). This is based on the assumption that despite the high dimensionality of the original motion description space, most human movements have an intrinsic representation in a low dimensional space [3].

Only a few studies use PCA not for reducing the dimensionality of the angle data, but as a way of modeling motion units composed of a sequence of frames. Thanks to their periodicity, walk cycles are especially well suited for such an algorithm. This approach as been taken for instance by Glardon et al. [7, 8] and Troje [18].

2.2 Statistical motion data modeling

Walk synthesis techniques taking into account the variability of the walk cycle over time also exist. They use statistical learning techniques to automatically extract the underlying rules of human motion, without any prior knowledge, directly from training on 3D motion capture data. Starting from the statistical models trained that way, new motion sequences can automatically be generated, using only some high-level commands from the user. Two movements generated by the same command (for example executing two walk cycles) will never be exactly identical. The result presents indeed a random aspect as can be found in the human execution of each motion, and becomes potentially more realistic than the repetition of the same motion capture sequence over and over. The motions produced that way are thus visually different, but are all stochastically similar to the training motions.

In order to take into account the high dynamic complexity of human motion, most of the researches in this path base their training on variations of hidden Markov models, Markov chains or other kind of probabilistic transitions between motions [17, 20, 14].

3 Database recording

The performance of models trained on data will highly depend on the quality of the data and its accuracy of description of the phenomenon that has to be modeled. As motion capture is the only way to obtain realistic 3D human motion data [15], it is the only way to obtain representative training data for statistical modeling of human motion.

Most motion capture recordings are performed using optical motion capture devices. This technology very often implies space constraints and treadmills have to be used to record walk databases, which impairs the naturalness of walking. Our database was created using the IGS-190 [11], a commercial motion capture suit that contains 18 inertial sensors consisting of a three axis accelerometer, a three axis gyroscope and a three axis magnetometer. This kind of motion capture suit has no space limitation and walk can thus be recorded in a more natural way. This is especially interesting for expressive gait where the subject does not always follow a perfectly straight trajectory and is thus given more freedom when he is not constrained to a given speed and trajectory like he would be with a treadmill.

The inertial motion capture suit captures directly angles between the body segments hence no mapping is necessary between tracked 3D positions of markers and joint angles.

Each motion file contains two parts: the skeleton definition and the motion data. The first part consists in defining the hierarchy of the skeleton, an approximation of the human body structure used in all motion capture systems which consists in a kinematic tree of joints modeled as points separated by segments of known constant lengths.

In the motion data part, the first three values of each frame give the 3D position of the root of the skeleton. They were discarded, as they depend on the displacement and orientation of the walk and can be recalculated given the foot contact with the ground and the leg segments lengths. The pose of the skeleton at each frame is then described by 18 tridimensional joint angles, which gives 54 values per frame to describe the motion. In our database, the data was recorded at a frame rate of 30 fps.

For the recording session, an actor wore the motion capture suit and walked back and forth on a scene. Before each capture sequence, he was given instructions about the"style" of walk that he had to act.

The walk sequences where then manually segmented into walk cycles (one cycle including two steps, one with each leg). We defined arbitrarily the boundary of our walk cycles as the moment the right heel touches the ground. As the actor walked back and forth, a turn was captured after each straight walk trajectory. Only the perfectly straight walk cycles were kept in this database, removing the turn steps and the transitions between turn and straight walk. Depending on the style of walk performed and its corresponding step length, a different number of walk cycles was recorded for each style. The eleven different styles and their corresponding number of cycles are presented in Table 1. These eleven styles were arbitrarily chosen as they all have a recognizable influence on walk, as illustrated in Fig. 1.

Table 1. Database walk styles and corresponding number of cycles recorded

	Walk Nbr	Style	Number of Cycles
	1	Proud	21
	2	Decided	15
	3	Sad	31
	4	Cat-walk	25
	5	Drunk	38
	6	Cool	23
	7	Afraid	16
	8	Tiptoeing	18
	9	Heavy	23
	10	In a hurry	19
	11	Manly	18
Total			247

4 Data preprocessing

First of all, in order to avoid problems associated with Euler angles, our original joint angle format, the motion data is converted in its quaternion form. In addition to avoid discontinuities in the angle channels, this conversion enables us to interpolate between two motion poses using the SLERP algorithm [16].

The actor moved across a scene, walking back and forth. The global orientation of the actor, encapsulated in the joint angle values of the root of the

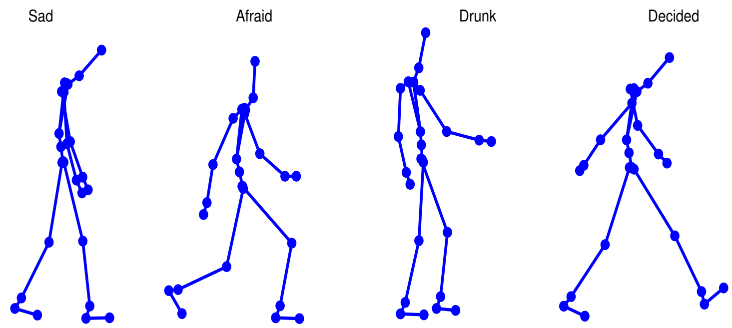

Fig. 1. Four example postures taken from the motion capture database (sad, afraid, drunk and decided walks)

skeleton, were thus rotated so that the walk sequences in the database always face the same direction.

As the number of frames for each walk cycles varies greatly across styles and over time for a single walk sequence, the time length of the walk cycles was normalized in order to give the same weight to each cycle in the subsequent PCA. This data resampling was performed using the SLERP algorithm for interpolation and all cycles were resampled to 40 frames. Furthermore, the same number of cycles had to be kept for each walk style when building the PC space. As 15 is the maximum common number of cycles (walk number 2 (*decided* style) has only 15 examples (see Tab. 1)), the first 15 cycles of each walk were kept for the PCA step, for a total of 165 cycles out of 247.

Unfortunately, PCA is a strictly linear algorithm and cannot be applied on quaternions as they do not form a linear space. The non linear quaternion rotations have thus to be converted into a linear parameterization. Our quaternion representation of joint rotations was thus reparameterized into exponential maps [9, 12] that are locally linear and where singularities can be avoided. In addition to that, this transformation maps the four values of quaternion angles to three values for exponential map representation and reduces thus the dimensionality of our data before PCA. Each walk cycle is thus represented after data preprocessing by a vector with a fixed number of variables:

$$40\, frames * 18\, joints * 3\, dimensions\ for\ exponential\ maps = 2160\, values.$$

5 Principal component analysis

Given a set of numerical variables, the aim of PCA is to describe that original set of data by a reduced number of uncorrelated new variables. Those new variables are linear combinations of the original variables. Reducing the number of variables causes a loss of information, but PCA ensures that loss of information to be as small as possible. This is done by ordering the new variables by the amount of variance of the original data set that they represent.

In our case, the variables of the original data matrix on which the PCA will be performed are the 2160 values representing each walk cycle. The observations of these variables are the:

15 *cycles per style* $*$ 11 *walk styles* $=$ 165 *observations of the walk cycle*.

When performing PCA, a mean centering is first necessary for the first principal component to describe the direction of maximum variance and not the mean of the data. We thus compute the mean vector out of our 165 walk cycles and remove it from our data matrix before PCA. The PCA can then be carried and the subspace of the principal components is calculated. Given that in our case the number of variables (P=2160) is higher than the number of observations (N=165), the number of principal components is reduced to the number of observations minus one (N-1=164). The new variables in the PC space can then be expressed as follows:

$$Z = XA \tag{1}$$

with X the original data matrix minus the mean of size NxP , Z the matrix of principal component scores (or the original data in the PC space) of size NxN-1 and A the loadings matrix (or the weights for each original variable when calculating the principal components) of size PxN-1.

As was stated before, PCA orders the principal components according to how much of information (or variance from the original data) they represent. The contribution of each component to the whole original information is represented as a cumulative percentage in Fig. 2.

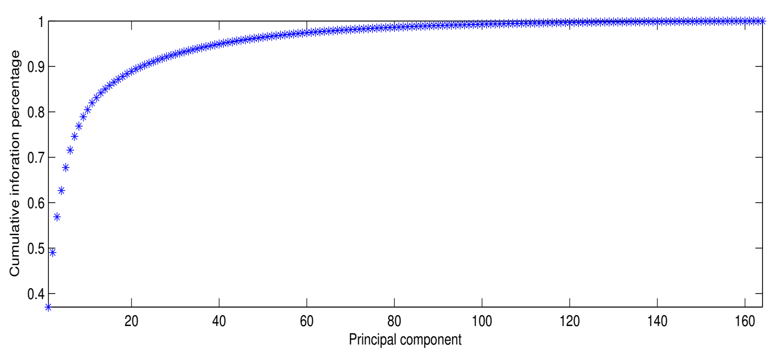

Fig. 2. Cumulated percentage of information contained in the 164 principal components

As PCA is performed in the first place to reduce the dimensionality of the original data, one has to decide how many principal components are to be kept. A very usual way of doing this is to keep the first k principal components that represent 80% of the cumulative percentage of information. Unfortunately, this is an empirical criterion as, depending on the variation present in the original data, 80% of these variations will represent very different levels of detail in the original motion. In our case, 80% of cumulated percentage was reached with 10 PCs but the data reconstruction using only 10 PCs was visually significantly impoverished compared to the original data as the style variations were smoothed. So we chose to increase the cumulated percentage of information in our PC subspace by using more principal components. Taking into account 90% of the cumulated percentage of information, which corresponds to 23 principal components, gave data reconstruction that were very difficult to differentiate from original data by the human eye.

As our original 165 walk cycles differ mainly in their style, the first PCs that represent the more variation in the original data will represent mainly the style variations. This assumption is verified and represented in Fig. 3 where the scores of the first four principal components for 15 sequences of each of the 11 different styles are illustrated. We can choose any pair of styles: one or several of the PCs will always enable us to differentiate them. For instance even if the first style (*decided*) is very similar to the second one (*sad*) if we look at the 2nd or 4th PCs, the 1st and 3rd PCs enable to differentiate those two styles very well.

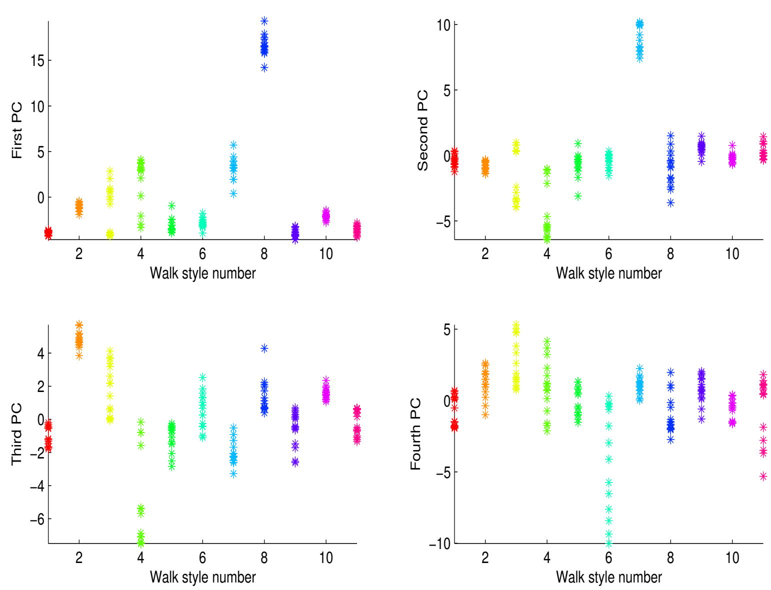

Fig. 3. Scores of the first four principal components, for the 15 occurrences of each of the eleven different styles (see Tab. 1)

6 Principal component space modeling

Once principal component analysis is performed, both Z, the matrix of variables expressed in the new PC space (scores matrix), and A, the transform matrix from the original space to the PCA space (loadings), are available. In Section 5, only 15 examples of walk cycle were kept for each motion style so that each walk style was represented by the same number of cycles. Once the PC space is determined giving the same importance to each style, the transformation matrix A can be used to transform the remaining walk cycles, that were not used for principal component analysis, into the PC space. This enables us to take fully advantage of our database for the analysis of the inter-cycles variability of each motion style.

As we have seen in Section 5, PCA enabled us to capture the style variation of walk cycles in a reduced number of principal components. But as could already be seen in Fig. 3, the principal components do not only vary according to style but they also vary along several occurrences of walk cycles for a given style. This phenomenon can already be noticed in straight natural walk, but is heavily amplified in these acted style walk sequences. In this kind of walks, the variability of walk cycles over time is an intrinsic part of the plausibility of the whole sequence. A single walk cycle repeated again and again will rapidly loose all believability in the eye of the spectator. This is why we decided to model the time variability of each style expressed in the PC space rather than taking only one sample of each style for building the PC space or taking only the mean of the principal component scores for each style.

Following the same reasoning, the time variability of the duration of each walk cycle before time normalization to 40 frames per cycle was also modeled for each style independently.

Figure 4 shows the variability of the scores of the first four principal components and of the cycle duration over the 15 examples of walk cycle for style 2 (*decided walk*).

We analyzed how principal component scores varied over time for walk cycles that follow each other but there appeared to be no obvious law directing the variations, except that they always stayed in the same range and that the variations between two adjacent cycles did not exceed a given threshold. We decided to model these variations using a very basic description. Each principal component score is modeled as a Gaussian distribution, whose mean $Mean_i$ and standard deviation $StdDev_i$ are computed using all the available walk cycles for the style being modeled. The maximum variation between two adjacent walk cycles is also calculated and used later in the motion synthesis step. The same process is repeated with the time duration for each walk style.

Once this is done, our whole database is modeled using a finite number of parameters, that can then be used to produce walk sequences that vary accordingly to the original data at each cycle.

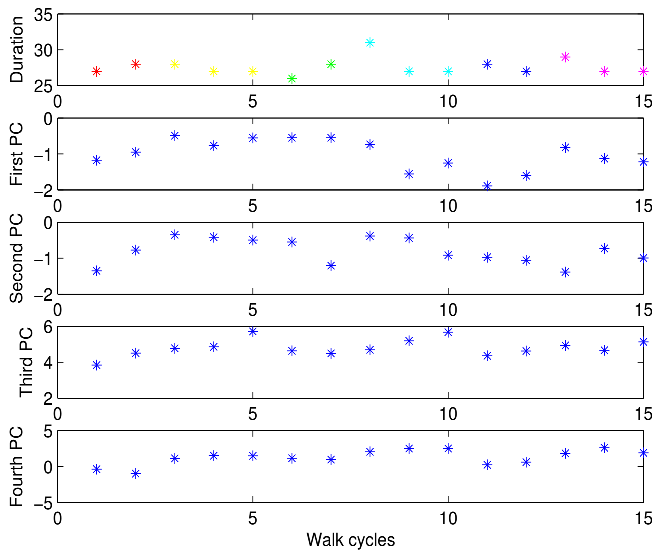

Fig. 4. Variations of cycle duration and of the scores of the first four PCs for 15 observations of walk cycle with a *decided* style. The different colors show the cycles that follow each other in one straight walk trajectory between two turns.

7 Walk synthesis

7.1 New PC space variable production

Once all our models are trained, they can be used to produce new sequences of motion that have the same characteristics as the original data and make thus new plausible motions. In this work, we did not study the possibility of morphing one walk style into another one in the PC space, so we only produce walk sequences for one given walk style at a time. Once the style is chosen, the first step of this synthesis is to produce new values of the scores in the PC space for each cycle of the walk sequence to be synthesized. Given the Gaussian distributions calculated in Section 6 and a *RandGauss* function that outputs Gaussian distributed random values with mean zero and standard deviation one, the new score corresponding to the i^{th} PC is calculated as follows, with as many calls to the *RandGauss* function as there are cycles to be synthesized:

$$Z_{synth}(NbrCycle, i) = Mean_i + StdDev_i * RandGauss \tag{2}$$

To help synthesizing plausible walk sequences and cycles that can be smoothly concatenated, a post-processing step is then performed to ensure that the variation between the scores of two subsequent cycles does not exceed the threshold of the original data, and reduce the gap by recalculating the concerned score if it was to be the case.

7.2 From PC space to original data format

Once we have our synthesized scores in the PC space for each one of the walk cycles, the transformation from the PC space to the original motion space can easily be performed using the following equation:

$$X_{synth} = Mean_{OrigData} + Z_{synth}(NbrCycle) * A^T \tag{3}$$

The data X_{synth} can then be brought from its exponential map form to the quaternion representation of joint angles. In the quaternion space, a resampling of the walk cycles can be performed using the SLERP algorithm, according to the synthesized durations obtained in the same manner as the PC scores:

$$Duration_{Synth}(NbrCycle) = Mean_{Dur} + StdDev_{Dur} * RandGauss \tag{4}$$

The cartesian coordinates of the root of the skeleton can then be computed. Using our knowledge of the boundaries of the synthesized walk cycles and calculating the height of each foot thanks to the known leg segment lengths, we determine which foot is in contact with the ground. From that fixed 3D position, we calculate the position of the whole body until the other foot becomes the reference, and so on for the whole sequence.

This method enables us to ensure that no foot sliding effect can occur, as the displacement of the whole body is driven by the foot contact point with the ground.

7.3 Cycles concatenation

Given that the $Mean_{OrigData}$ of the PC space to original space recomposition is the same for all sequences, and that the variations between PC scores from adjacent motion cycles were kept under values encountered in the training database, no huge differences appear between the end of one cycle and the beginning of the following one. A very simple smoothing was thus sufficient to ensure that the cycle transitions were not disturbing for the human eye.

8 Results

Thanks to the method presented in Sections 5 and 6 we modeled our original data in a PC space that makes obvious the style differences in our walk cycles. These different styles were then modeled in that PC space and the use of random Gaussian values enabled us to introduce variability over time into the synthesizing process in a very simple way, while keeping the new motion data plausible. With a finite and reduced number of parameters we are now able to produce an infinite number of new motion sequences, as one cycle is not looped over and over for each style but a new cycle is produced each time. Some examples of synthesized motion sequences can be found at *http://tcts.fpms.ac.be/˜tilmanne/*.

The method presented in this article uses very simple algorithms and modeling techniques but still outputs very interesting results even convincing to the human eye, which is very sensitive to motion naturalness. In this study, we analyzed motions presenting very different style characteristics, which is quite unusual for such a study but still very interesting as characters in the virtual world very often present exaggerated or over-acted behaviors. With as low as 23 components, eleven completely different walk styles were represented, some of them like the *drunk walk* presenting a very high intra style variability.

9 Future work

One recurrent problem with motion data analysis and synthesis is the difficulty to evaluate the produced motion sequences. The next step for this study will thus be to perform an user evaluation to assess the naturalness of the produced motion and whether the loss of information when reducing the dimensionality of the data using PCA is perceived by the user.

Several parameters influencing the final results have to be tested, like how different from each other the original walk styles appear to the subject, how the subject perceives the difference between original motions and synthesized motions, how time variability influences the naturalness of the motion compared to a single walk cycle looped, and how the the number of principal components influences the reconstructed motion.

With very simple algorithms we were able to build a perpetual walking synthesizer. As we performed our principal component analysis on the whole database, all walk styles are represented in the same PC space. Even if we did not use this property here, the aim is now to be able to produce smooth style transitions into the PC space, so that our perpetual walker could not only walk with time variability but also move its expression from one style to any other style. This could include direct trajectories in the PC space or a transitional neutral walk.

10 Acknowledgment

The authors would like to thank the comedian Sebastien Marchetti for his participation in the motion capture database recording.

J.Tilmanne receives a PhD grant from the Fonds de la Recherche pour l'Industrie et l'Agriculture (F.R.I.A.), Belgium.

References

1. Brand, M., Hertzmann, A.: Style machines. Proceedings of the 27th annual conference on Computer graphics and interactive techniques, 183–192 (2000)
2. Calinon, S., Guenter, F., Billard, A.: On Learning, Representing, and Generalizing a Task in a Humanoid Robot. IEEE Transactions on Systems, Man and Cybernetics, 37(2), 286–298 (2007)

3. Elgammal, A., Lee, C. S.: The Role of Manifold Learning in Human Motion Analysis Human Motion Understanding, Modeling, Capture and Animation, 1–29 (2008)
4. Forbes, K., Fiume, E.: An efficient search algorithm for motion data using weighted PCA. Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation, 67–76 (2005)
5. Forsyth D. A., Arikan, O., Ikemoto, L., O'Brien, J., Ramanan, D.: Computational Studies of Human Motion: Part 1, Tracking and Motion Synthesis. Foundations and Trends in Computer Graphics and Vision 1:2/3, Now Publishers Inc. (2006)
6. Geng, W., Yu, G.: Reuse of Motion Capture Data in Animation: A Review. Computational Science and Its Applications (ICCSA), 620–629 (2003)
7. Glardon, P., Boulic, R., Thalmann, D.: Pca-based walking engine using motion capture data. Computer Graphics International, 292–298 (2004)
8. Glardon, P., Boulic, R., Thalmann, D.: A Coherent Locomotion Engine Extrapolating Beyond Experimental Data. Proceedings of Computer Animation and Social Agent (CASA), Geneva, Switzerland, 73–84 (2004)
9. Grassia, F. S.: Practical parameterization of rotations using the exponential map. Journal of Graphics Tools 3, 29–48 (1998)
10. Grudzinski, T.: Exploiting Quaternion PCA in Virtual Character Motion Analysis. Computer Vision and Graphics, 420429 (2009)
11. IGS-190. Animazoo. www.animazoo.com, 2010.
12. Johnson, M. P.: Exploiting quaternions to support expressive interactive character motion. PhD Thesis (2002)
13. Jolliffe, I. T.: Principal Component Analysis, Springer Series in Statistic, 2nd ed. Springer-Verlag, New York (2002)
14. Li,Y., Wang, T., Shum, H.: Motion texture: a two-level statistical model for character motion synthesis. Proc. of SIGGRAPH '02, 465–472, New York, NY, USA, (2002)
15. Menache, A.: Understanding motion Capture for Computer Animation and Video Games. San Francisco, CA, USA: Morgan Kauffman Publishers Inc. (1999)
16. Shoemake, K.: Animating Rotations with Quaternion Curves. Proc. of SIGGRAPH'05, San Francisco, 19(3), 245–254 (1985)
17. Tanco, L. M., Hilton, A.: Realistic synthesis of novel human movements from a database of motion capture examples. Proc. of the Workshop on Human Motion (HUMO'00), 137–142, Washington DC, USA (2000)
18. Troje, N. F.: Retrieving information from human movement patterns. Understanding Events: How Humans See, Represent, and Act on Events. Oxford University Press, 308–334 (2008)
19. Urtasun, R., Glardon, P., Boulic, R., Thalmann, D., Fua, P.: Style-Based Motion Synthesis. Computer Graphics Forum, 23(4), 799–812 (2004)
20. Wang, Y. , Liu, Z., Zhou, L.: Automatic 3d motion synthesis with time-striding hidden markov model. Proc. International Conference on Machine Learning and Cybernetics (ICMLC'05) 3930, 558–567, Guangzhou, China (2005)