# Temporal segmentation and seamless stitching of motion patterns for synthesizing novel animations of periodic dances

Costas Panagiotakis*, Antonis Argyros† and Damien Michel†
*Dept. of Business Administration, TEI of Crete, Agios Nikolaos, Greece
Email:cpanag@staff.teicrete.gr
†Institute of Computer Science, FORTH, Heraklion, Crete, Greece
Emails: {michel,argyros}@ics.forth.gr

***Abstract*—In this paper, we present an efficient algorithm for synthesizing novel, arbitrarily long animations of periodic dances. The input to the proposed method is motion capture data acquired from markeless visual observations of a human performing a periodic dance. The provided human motion capture data are temporally segmented into the constituent periodic motion patterns. These are further organized in a motion graph that also represents possible transitions among them. Finally, an efficient algorithm exploits this representation to come up with a previously unseen sequence of motion patterns that are stitched seamlessly into a novel, realistic dance animation. Several experiments have been conducted with real recordings of Greek folk dances. The obtained results are very promising and indicate the efficacy of the proposed approach, as well as its tolerance to dynamic and noisy human motion capture input.**

## I. Introduction

In recent years, there has been an unprecedented progress in 3D human motion capture. As a consequence, the availability of motion capture data is increasing in both volume and quality and new types of applications emerge [1], [2]. The support of these applications require new services, methods and tools for searching, retrieving, understanding and reproducing motion capture data of human activities. Several such activities (e.g. walking, running, jumping, dancing, etc.) can be considered as periodic in the sense that they consist of similar, repetitive motion patterns. The temporal characteristics of this repetition defines the rhythm of the activity. For example, the rhythm of certain types of dances is based on the repetition of motion patterns that are related to the music tempo. This holds even if the dance steps exhibit certain variations [3].

In this paper, we are interested in analyzing the motion capture data of cyclic human activities so that we are able to create new, infinite sequences of unseen animations. We are also interested in being able to synchronize these unseen animations with a music tempo that differs to the one of the original dance recording. In that direction, we capitalize on the assumption of periodic dance motion to detect the rhythm of motion, to extract the motion patterns and to perform synchronization with a given tempo.

A lot of research has been already devoted to the vision-based motion analysis in order to estimate the rhythm of motion. In [4], a method of rhythmic information extraction from 2D dance videos and music has been proposed. The rhythm of motion is estimated by the analysis of motion trajectories of points that are detected using an adaptation of the Shi-Tomasi (ST) corner detector.

Since the 2D visual information is not always sufficient to solve the problem of motion rhythm estimation with high accuracy, other methods [5] have been applied to 3D motion capture data. Kim et al. [5] proposed a method for synthesizing a new motion from unlabelled example motions while maintaining their rhythmic features. This method first captures the motion beats from the example motions to extract the basic movements. Based on those 3D motion data it constructs a movement transition graph that represents the example motions. Finally, given an input sound signal, it synthesizes a novel motion in an on-line manner while traversing the motion transition graph, which is synchronized with the input sound signal and also satisfies kinematic constraints given explicitly and/or implicitly.

The beat detection in music has been used by several methods that aim to create new unseen dance animations that are synchronized with a given music [6]–[8]. In [6], a fast, greedy algorithm analyzes a library of stock motions and generates new sequences of movements that were not described in the library. A greedy algorithm with backtracking establishes the best matching frame among the closest dance moves, takes it as a greedy choice and repeats the same process. A genetic algorithm optimizes the dance sequence by considering an initial population of valid dance figures and applying the genetic operators of crossover and mutation to create new figures.

In [7], the generation of dance performances is based on a given musical piece by matching the progressions of musical and motion patterns and by correlating musical and motion features. That method uses similarity matrices for musical and motion sequences and matched the progressions of musical and motion contents by minimizing the difference between the two similarity matrices. In [8] a music-driven semiautomatic character animation framework is presented that extracts musical features from a song and uses them to create an animation. The proposed method builds a new animation directly from musical attributes. Any type of music with noticeable beats can be used to generate a tailored animation that expresses the

Antonis Argyros is also with the Computer Science Department, University of Crete, Greece.

music. A script file that is set by the user gives a high level control over the final animation.

Marker-based motion capture technologies have been used on 3D dance animations systems. In [9], a dance training system is proposed that is based on virtual reality (VR) and marker-based motion capture technologies. A prototype was implemented, in which a trainee can imitate the motion demonstrated by a virtual teacher whose actions are displayed on a wall screen. Meanwhile, the motion of the trainee is captured and analyzed by the system to provide feedback. In addition, user constraints on movement, position and timing can be used for motion synthesis. In [10], a framework is presented for motion synthesis from multiple 3D video sequences according to user constraints on movement, position and timing. Shape similarity over an adaptive temporal window is used to identify transitions between 3D video sequences. Novel 3D video sequences are synthesized by finding the optimal path in the surface motion graph between user specified key-frames for control of movement, location and timing.

In our preliminary work [3], we proposed a framework that performs temporal segmentation of periodic dances into its periods based on visual and sound information. Having segmented a given human motion and music into periods, the next task is to create a beat synchronous dance animation. To achieve this, we re-sampled the motion signal so that its tempo becomes equal to that of the target music. The main contributions of the proposed framework are the following:

- Most of the existing approaches that provide a temporal segmentation of human motion can be only applied to simple human motions, since they are heuristic and use simplifications or signal approximations without any global optimality criterion. On the contrary, in this paper, we have used an optimization approach that computes the optimal solution for the problem of temporal segmentation of human motion using 3D dance motion data [3]. An advantage of the proposed method is that it can be applied to complex multidimensional signals such as those representing dance movements creating new unseen animations.
- The existing approaches usually require special hardware using marker based motion capture devices that provide high accuracy motion data, since they are not capable of tolerating noise in their input. The proposed method tolerates noise in the representation of human motion due to the proposed motion pattern merging method. The input to the algorithm is motion capture data produced by a home-build markerless human articulation tracking algorithm that relies on the visual input provided by two RGBD cameras or several RGB cameras.
- The motion synthesis algorithms are, typically, human driven or search for the best matching between motion patterns in order to create new unseen animations. Other methods create beat synchronous animations by matching the progressions of musical and motion patterns based on the estimated correlation of musical and motion features. In this work, the synchronization with music (tempo) is optional. Moreover, the combination of the proposed pattern merging method and motion planning algorithm does not produce simple concatenations of previously observed periods of dance but rather creates smooth transitions by seamless stitching of motion patterns resulting far more realistic animations. The experimental results show that the proposed method achieves very promising results.

## II. Method Overview

Figure 1 illustrates a flow diagram of the proposed approach. The first module realizes the temporal segmentation of the complex motion capture data into motion patterns. Next, using the extracted motion patterns, we construct a motion graph and we employ an efficient algorithm for graph exploration that yields an unseen animation. Finally, in the case that a music is given, a beat synchronous animation is created by applying the method proposed in [3]. We resample each motion pattern of the resulting animation according to the estimated music tempo, so that the motion tempo becomes equal to the tempo of the given music.

The rest of the paper is organized as follows. Section III presents how the human motion is temporally segmented into periods of motion patterns. Section IV presents how these patterns are represented and explored towards creating new unseen dance animations. The synthesis of new unseen animations is presented in Section V. The experimental results obtained based on experiments on a series of datasets are provided in Section VI. Finally, Section VII concludes the paper by summarizing the most important contributions of this research.

## III. Temporal Segmentation of Periodic Human Motion

In this section, we present the analysis of motion captured data in order to segment the motion patterns of the periodic dance motion and to represent them in a motion graph. A preliminary description of the segmentation algorithm was presented at [3] and is elaborated here, also for the purpose of self-completeness.

The input of this method is the time series of the joint angles of an articulated human model. This can be the result of an elaborate motion capture system (e.g., [11]) or of a much cheaper setup that is based on a contemporary RGBD sensor [12] and the accompanying software [13]. In our case, we use a middle ground that requires markerless visual observations that are acquired by two synchonized and calibrated Kinect sensors [14]. It has been shown [14] that this approach constitutes a middle ground between the accurate but expensive and less-convenient marker-based motion capture solutions and the less accurate but rather cheap markerless single Kinect approach.

We employ the following optimization approach to compute the optimal solution to the problem of temporal segmentation of human motion using 3D dance motion data. Let $S \in \Re^{m,n}$ be the given multidimensional signal of captured human motion that contains the time series of the $m$ degrees of freedom (i.e., joint angles) of the human motion. Let also $n$ be the number of temporal samples of each of these series. Assuming that the human motion is periodic, the goal of temporal segmentation is to segment $S$ into its periods. Let

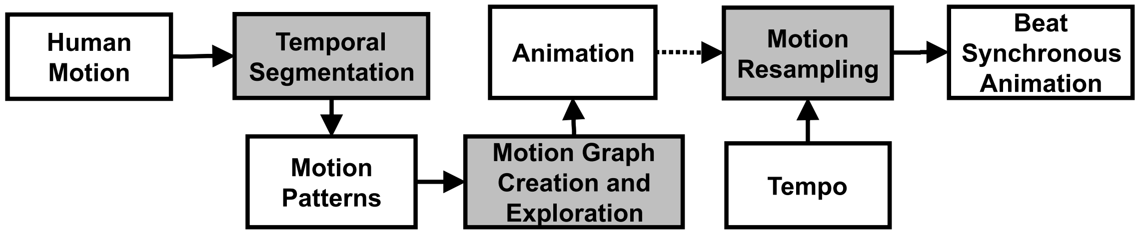



Fig. 1: The flow diagram of the proposed method. Blocks in gray color constitute the main tasks of this work.

$T_p = \{t_0, t_1, t_2, ..., t_p\}$, $1 = t_0 < t_1 < t_2 < ... < t_p \leq n$ be a temporal segmentation of $S$ into $p$ segments. The number of periods $p$ can be automatically computed by getting the global maximum of the amplitude signal of Fast Fourier Transform of a given motion signal. More details and methods for $p$ computation can be found in [3]. For each such segmentation, we define the following energy function

$$E(T_p) = \sum_{i=1}^{m} \sum_{k=1}^{p-1} d(S_i(t_{k-1} : t_k - 1), S_i(t_k : t_{k+1} - 1)). \quad (1)$$

In Eq.(1), $d(.,.)$ denotes the distance between the candidate motion patterns (signal segments) $S_i(t_{k-1} : t_k - 1)$ and $S_i(t_k : t_{k+1} - 1)$ [3]. The proposed method yields the optimal temporal segmentation by minimizing the energy $E(T_p)$. This is done by constructing a graph that represents the candidate motion patterns and the distances among them. Then, the global minimum of $E(T_p)$ is given by the sum of weights of a shortest path that is defined in this graph.

## IV. Constructing the Motion Graph

Given the extracted motion patterns, we construct a directed motion graph $G = (V, E)$ comprising a set $V$ of nodes together with a set $E$ of edges. A node of the graph corresponds to one of the estimated motion patterns. An edge from node $u \in V$ to node $v \in V$, represents the fact that the transition from motion pattern $u$ to motion pattern $v$ is possible. In order to construct the edges, we first normalize the two skeletons with respect to their 3D position. Figure 2 illustrates an example of two 3D position-normalized skeletons (blue and red skeletons). Then, for each pair of nodes $u, v \in V$, we compute the Euclidean distance $W(u, v)$ between the 3D Euclidean coordinates that correspond to the last and the first frames of motion patterns $u$ and $v$, respectively.

$$W(u, v) = |\widetilde{u(|u|)} - \widetilde{v(1)}|_2, \quad (2)$$

where $\widetilde{u(|u|)}$ and $\widetilde{v(1)}$ denote the Euclidean coordinates (after position normalization) of the last and the first frame of $u$ and $v$, respectively.

The $n^2$ pairs of edge weights $W(.,.)$, since compute all-pairs distances are computed, are sorted in ascending order and stored in vector $C_s$. Initially, the set of edges $E$ of $G = (V, E)$ is set to $E = \emptyset$. Then, the edge connecting the nodes that correspond to the first entry of $W(.,.)$ is inserted to $E$. This procedure is repeated until $G$ becomes a connected graph. The reason for seeking the connectedness of $G$ is that we require that the motion planning algorithm is able to traverse the whole graph $G$. In order to keep the graph balanced (i.e., all nodes have a similar degree), we further enforce an upper bound on graph nodes. This means that if an edge needs to be introduced to the graph, this will not happen if one of the nodes it connects has a degree more than a certain threshold $D$. In the experiments carried out in this paper we set $D = 10$). A similar version of this algorithm has been employed in [15], where the goal was to construct a connected graph of superpixels used to solve the interactive image segmentation problem.

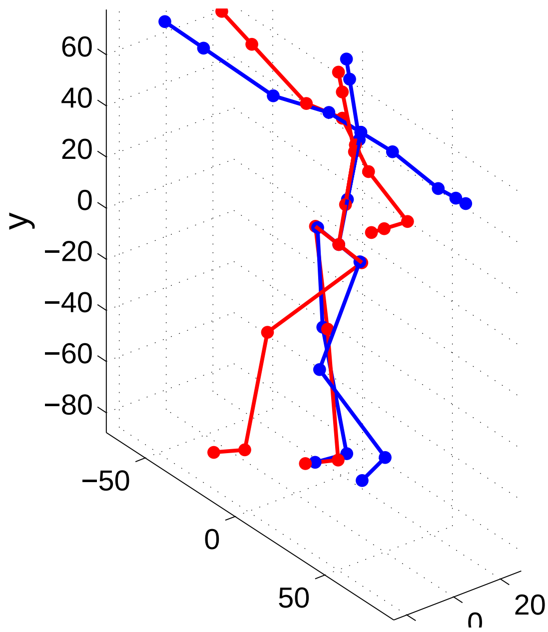



Fig. 2: An example of two skeletons that have been located in zero position.

## V. Motion Synthesis and Pattern Merging

This section presents the proposed motion synthesis algorithm that yields a series of motion patterns, traversing the motion graph $G$. Moreover, in order to smooth the transitions between successive motion patterns and to decrease the input data noise level, we proposed and perform a motion pattern stitching method.

### A. Motion Planning Algorithm

The goal of the proposed motion planning algorithm is to yield novel, previously unseen and realistic animations of the captured dance motions. This problem has been analytically studied in [2]. In this work, we have used the algorithm proposed in [2] called snake walk which, briefly stated, operates as follows. First, an arbitrary node (motion pattern) of $G$ is selected and then the next states are chosen by traversing edges of the graph. According to the snake walk algorithm, for each node $q$ we store the number $M(q)$ of times that this node has been traversed and the number $F(q)$ of times that this node has been selected and the next node was a visited node. Among all neighboring nodes of the current node, the one that maximizes $W(q)$, defined hereafter, is chosen:

$$D_u(q) = 1 - sgn(M(q)) \quad (3)$$
$$D_v(q) = deg(q) \cdot (sgn(M(q)) + sgn(F(q))) \quad (4)$$
$$D_f(q) = M(q) \cdot 2^{F(q)} \quad (5)$$
$$W(q) = \frac{deg(q)}{D_u(q) + D_v(q) + D_f(q)} \quad (6)$$

where $deg(q)$ and $sgn(.)$ denote the degree of node $q$ and the sign function, respectively. The intuition behind the maximization of the quantity of Eq. (6) is that the preferred nodes should be the least visited ones. If unvisited nodes in one step do not exist, the ones that are neighbours of unvisited nodes are selected. If there are many unvisited nodes, the nodes with high degree are preferred, because it is more likely that these nodes will drive the algorithm to unvisited nodes of the graph. The snake walk algorithm provides an approximate solution to the problem of finding a Hamilton Path in a graph. While the Hamilton Path computation is an NP-complete problem, the snake walk algorithm has a complexity that is linear to the cardinality of the set of graph vertices. It holds that for many known graphs, like connected cliques with bridges, high density graphs, which are possible motion pattern graphs, it can be proved that the above algorithm needs $O(n)$ steps to cover the graph. More details about snake walk algorithm can be found in [2].

### B. *Seamless Stitching of Motion Patterns*

The proposed a motion pattern merging method is based on a low pass filter that ensures continuity and smooth transitions between successive motion patterns and reduces the noise level of the input motion data. The use of the motion planning algorithm yields unseen and realistic animations that consist of the joint angles time series. However, discontinuities due to the transitions between motion patterns and to the noise of the given motion data do exist. The employed low pass filter is independently applied to the time series of each joint angle. Due to the smoothness and to the periodicity of joint angles' signals, we used the Fast Fourier Transform (FFT) coefficients to represent them, keeping the first $10\%$ and $20\%$ of FFT coefficients that suffice to recover the three angles for pose and the rest joint angles, respectively. The rest of the coefficients are set equal to zero. Then, by using the inverse FFT, the signal is reconstructed, correcting the signal discontinuities and reducing noise at the same time.

Another advantage of this technique is that it can be easily adapted to different joints by setting different thresholds on the number of non-zero FFT coefficients according to the feasible variation of the joint angle. Moreover, due to the application of FFT on the whole signal, we obtain differences in animation even in transitions between the two same motion patterns. This increases the motion variability and results in realistic animations, even in cases where the number of motion patterns/nodes in $G$ is relatively low like in our datasets.

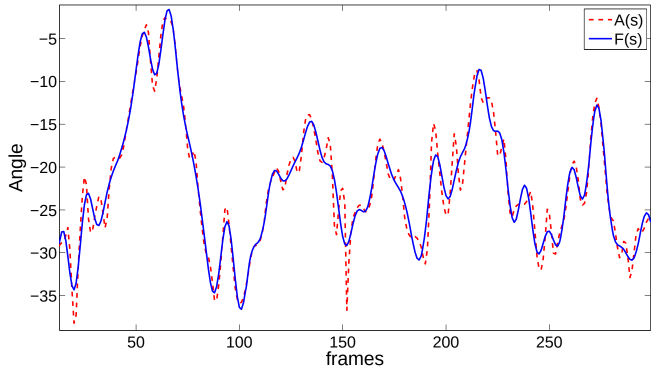


Fig. 3: An example of a joint angle signal before and after the application of the FFT-based low pass filter.

Figure 3 shows an example of a joint angle signal before, ($A(s)$) and after ($F(s)$) the application of FFT-based low pass filter. A new motion pattern starts on frame 152. It holds that the application of the low pass filter improves the signal continuity between frames 151 and 152 and reduces the noise of the original signal.

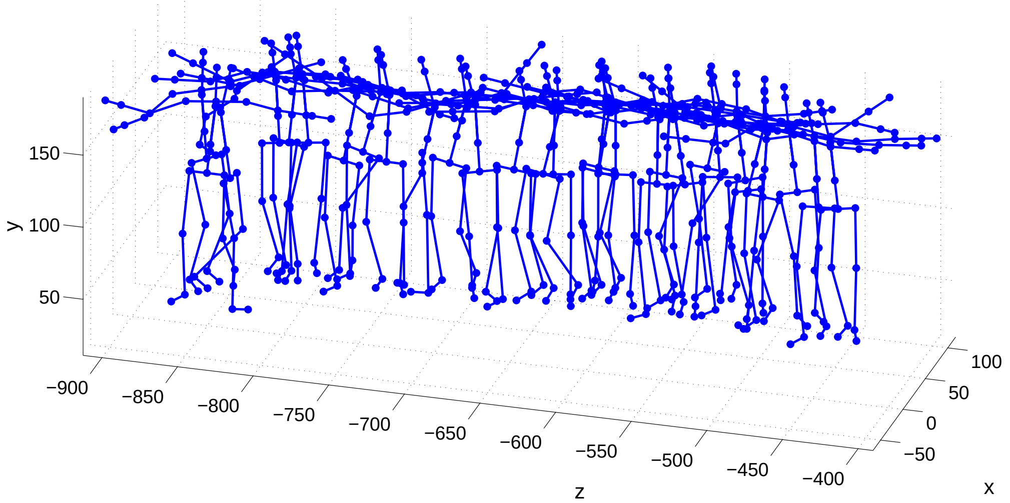
## VI. Experimental Results

In this section, we present experimental results from a number of experiments. We have tested our approach using 3D motion captured data containing the folk dances "Siganos" and "Syrtos" from the island of Crete [14]. We used several recordings of a professional dancer. Frame capturing was performed at a rate of 30 fps. Like most of traditional Cretan dances, a theme can be repeated with an infinite number of variations [3]. The number of recorded motion patterns of Siganos and Syrtos were 63 and 32, respectively. Based on the proposed framework, we synthesized several dance animations (more than 100,000 frames) of those dances.

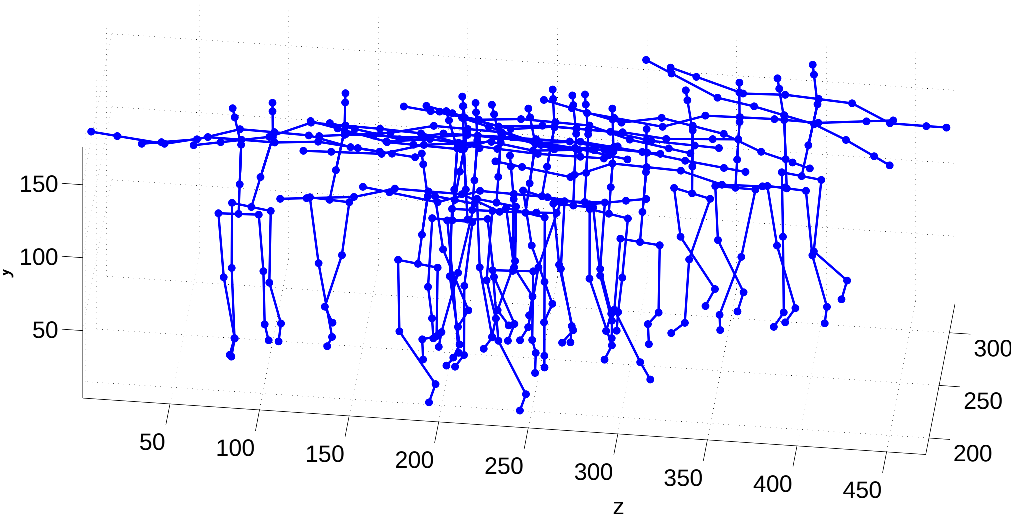
Figure 4 depicts snapshots of several human skeletons during the synthesis of Siganos and Syrtos dances. The snapshots are captured once every 50 frames, while the total number of frames that correspond to the Siganos and Syrtos animations are 1400 and 1000, respectively. Figure 5 illustrates seven frames of unseen dance animation for Siganos during a human body rotation. The snapshots are captured once every 50 frames. These figures show the high variability of human poses and the seamless stitching of motion patterns during Siganos and Syrtos animations.

A more complete set of video results containing several dance animation videos can be downloaded at [1]. The synthetic dance animations are smooth and realistic. Although the number of motion patterns are not high, the variability of resulting animations is high due to the proposed motion synthesis and pattern merging method.

[1] https://sites.google.com/site/costaspanagiotakis/research/dancer




(a)




(b)

Fig. 4: Snapshots of several human skeletons during the synthesis of **(a)** Siganos and **(b)** Syrtos dances.

## VII. Summary

In this paper, we have proposed a framework for the analysis and exploration of motion patterns that has been applied to synthesize realistic animations with the option of synchronization with any given music. The segmentation of input motion capture data can be applied to any type of complex motion data, resulting the motion patterns that correspond to the nodes of motion graph. The proposed method improves the motion signal continuity during motion patterns transitions and reduces the noise of the original motion signal. Moreover, the proposed method is able to synthesize realistic animations with high motion variability even in cases where the number of motion patterns is low.

The proposed approach has been successfully tested on dances containing cyclic activities such as traditional dances from Crete. Regarding future work, we plan to apply the proposed framework to other types of periodic dances.

## Acknowledgments

This research has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF) - Research Funding Project: THALIS-UOA-ERASITECHNIS MIS 375435 and also from the European Communitys Seventh Framework Programme (FP7/2007-

(a) #0 (b) #50 (c) #100 (d) #150 (e) #200 (f) #250 (g) #300

Fig. 5: Frames of new unseen dance animation for Siganos dance.

2013) under grant agreement No. 288146, HOBBIT.

## REFERENCES

[1] L. Kovar, M. Gleicher, and F. Pighin, “Motion graphs,” in *SIGGRAPH*, 2002.

[2] C. Panagiotakis and G. Tziritas, “Snake terrestrial locomotion synthesis in 3d virtual environments,” *The Visual Computer*, vol. 22, no. 8, pp. 562–576, 2006.

[3] C. Panagiotakis, A. Holzapfel, D. Michel, and A. A. Argyros, “Beat synchronous dance animation based on visual analysis of human motion and audio analysis of music tempo,” in *International Symposium on Visual Computing*, 2013.

[4] W. Chu and S. Tsai, “Rhythm of motion extraction and rhythm-based cross-media alignment for dance videos,” *IEEE Transactions on Multimedia*, no. 99, pp. 1–1, 2012.

[5] T. Kim, S. Park, and S. Shin, “Rhythmic-motion synthesis based on motion-beat analysis,” in *ACM Transactions on Graphics (TOG)*, vol. 22, no. 3, 2003, pp. 392–401.

[6] G. Alankus, A. A. Bayazit, and O. B. Bayazit, “Automated motion synthesis for dancing characters,” *Computer Animation and Virtual Worlds*, vol. 16, no. 3-4, pp. 259–271, 2005.

[7] J. W. Kim, H. Fouad, J. L. Sibert, and J. K. Hahn, “Perceptually motivated automatic dance motion generation for music,” *Computer Animation and Virtual Worlds*, vol. 20, no. 2-3, pp. 375–384, 2009.

[8] D. Sauer and Y.-H. Yang, “Music-driven character animation,” *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, vol. 5, no. 4, p. 27, 2009.

[9] J. C. Chan, H. Leung, J. K. Tang, and T. Komura, “A virtual reality dance training system using motion capture technology,” *IEEE Transactions on Learning Technologies*, vol. 4, no. 2, pp. 187–195, 2011.

[10] P. Huang, A. Hilton, and J. Starck, “Human motion synthesis from 3d video,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1478–1485.

[11] Vicon, “Vicon: Motion capture systems,” 2013. [Online]. Available: {http://www.vicon.com}

[12] Microsoft Corp. Redmond WA, “Kinect for Xbox 360.”

[13] OpenNI, *OpenNI User Guide*, OpenNI organization, November 2010, last viewed 19-01-2011 11:32. [Online]. Available: http://www.openni.org/documentation

[14] D. Michel, C. Panagiotakis, and A. A. Argyros, “Tracking the articulated motion of the human body with two RGBD cameras,” *Machine Vision and Applications*, 2013 (under review).

[15] C. Panagiotakis, H. Papadakis, E. Grinias, N. Komodakis, P. Fragopoulou, and G. Tziritas, “Interactive image segmentation based on synthetic graph coordinates,” *Pattern Recognition*, vol. 46, no. 11, pp. 2940–2952, 2013.