Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 185378, 10 pages
http://dx.doi.org/10.1155/2014/185378

Research Article

An Online Full-Body Motion Recognition Method Using Sparse and Deficient Signal Sequences

Chengyu Guo,[1] Jie Liu,[1] Xiaohai Fan,[1] Aihong Qin,[2] and Xiaohui Liang[1]

[1] *State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science, Beihang University, Beijing 100191, China*
[2] *Zhejiang University of Media and Communications, Zhejiang 310058, China*

Correspondence should be addressed to Xiaohui Liang; lxh@vrlab.buaa.edu.cn

Received 18 February 2014; Revised 20 May 2014; Accepted 20 May 2014; Published 10 July 2014

Academic Editor: Yi Chen

This paper presents a method to recognize continuous full-body human motion online by using sparse, low-cost sensors. The only input signals needed are linear accelerations without any rotation information, which are provided by four Wiimote sensors attached to the four human limbs. Based on the fused hidden Markov model (FHMM) and autoregressive process, a predictive fusion model (PFM) is put forward, which considers the different influences of the upper and lower limbs, establishes HMM for each part, and fuses them using a probabilistic fusion model. Then an autoregressive process is introduced in HMM to predict the gesture, which enables the model to deal with incomplete signal data. In order to reduce the number of alternatives in the online recognition process, a graph model is built that rejects parts of motion types based on the graph structure and previous recognition results. Finally, an online signal segmentation method based on semantics information and PFM is presented to finish the efficient recognition task. The results indicate that the method is robust with a high recognition rate of sparse and deficient signals and can be used in various interactive applications.

1. Introduction

In recent years, sensor-based human motion recognition has received a great deal of attention from researchers. Sensors have been adapted for large-scale movements to avoid shading and lighting problems. This has advantages over vision-based methods for special scenes and has allowed full-body motion recognition and sensor-based motion control to be applied in various fields, such as medical rehabilitation and interactive games.

Currently, motion control tasks are based on accurate and complete accelerations, as well as signals provided by other sensors. Unfortunately, these devices are expensive and not easily portable. In practice, sparse and low-cost sensors are more attractive, but they are usually accompanied by less information, more noise, and frequent signal deletion, making it difficult to acquire or reconstruct accurate position information and accordingly harder to achieve a proper online recognition result. Therefore, reconstructing human motion from signal features based on sparse and deficient signals has recently evoked much interest.

In light of the above problems, an online motion recognition method that adopts sparse, low-cost Wii Remote sensors (Wiimotes) as input devices is proposed. Because sparse, deficient linear accelerations cannot acquire accurate position information of human motion, a predictive fusion model, which combines fused hidden Markov model (HMM) with an autoregressive process, is presented. Considering the independence of each part of the human body, a hierarchical fusion structure of fused HMM is used to deal with human motion signals, which enhances the independent and cooperative expression of the classification model. The predictive capability of the model provided by the autoregressive process ensures robustness when dealing with noisy and deficient signals. Once the online recognition process is underway, a graph model that builds the transition between different motion types filters those motion types and reduces the recognition complexity of the predictive fusion model (PFM).

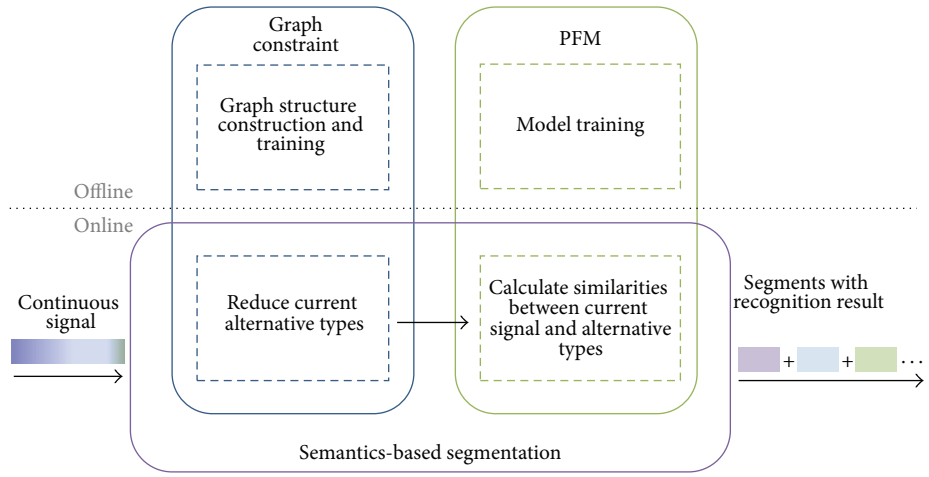


Figure 1: The structure of our method. The three main technologies include predictive fusion model (PFM), graph constraint, and semantics-based segmentation.

Moreover, a semantic-based automatic signal segmentation method is introduced to ensure the continuity of the online recognition processes.

Thus, based on sparse and deficient input signals, a human motion recognition PFM is presented that effectively supports sparse, low-cost sensors. The presented model is of a high accuracy rate and robust enough to handle insufficient and missing signals. An online motion recognition method is also proposed that does not require any position calibration. The method integrates PFM, action graph structure, and a semantic-based signal segmentation method to support user-driven virtual human motion in virtual scenes with continuous motions.

2. Related Works

As pattern recognition technologies develop, pattern recognition methods are increasingly used in the context of motion recognition. Typical methods, including self-organizing maps (SOMs), support vector machines (SVMs), and HMM approaches, can be adapted for motion recognition processes.

Methods for motion recognition vary depending on the input source. It has been shown that vision-based methods and sensor-based methods constitute two of the main research areas and are based on two types of input device, depending on the application. Poppe [1] presented a survey of vision-based human action recognition systems. Ning and Mokhtarian [2] used a shape to represent object contours extracted from each frame of a movie and constructed a tangent space based on the mean shape to approximate the linear space encompassing the datasets. Zhou et al. [3] and Min et al. [4] built a low-dimensional deformable model based on shape information from human motions in an image sequence to realize motion control. Lai et al. [5] proposed a local feature-based human motion analysis framework that extracted the features directly from local regions containing motion. Research has shown that the general idea of vision-based methods is to extract varied feature information from image sequences. In order to avoid the effect of light and shade and the inconvenience of vision-based methods when moving in a larger scene, sensor-based methods remain a hot topic in this field.

Recent work [6–9] which has described some basic methods for gesture recognition using accelerometers shows that sensor-based methods can be adapted for recognition tasks. Sun et al. [10] and Shiratori and Hodgins [11] used low-cost sensors to monitor daily physical activities. This method is practical but the finite types of simple activities limit recognition. Niu and Abdel-Mottaleb [12] considered the continuity of signals and provided a segmentation and recognition method based on HMM. Khan et al. [13] used a hierarchical scheme for human activity recognition. Tautges et al. [14] and Wong et al. [15] generated simple full-body animations controlled by sparse and accurate 3D accelerometers attached to the extremities of a human actor; this method is able to properly deal with accurate input to recover accurate human position information. In terms of both sparse and deficient signals, learning models are more effective than generative models. Early methods of the learning model define features analysis with HMM but require improvement in the robustness for deficient signals and the recognition rate.

The present research is motivated by the above studies. A probabilistic fusion model and autoregressive process in the hierarchical model of virtual human movement are proposed, which ensures that full-body motion information can be expressed relatively independently and deals with deficient input caused by sparse, inexpensive sensors. The recognition process ensures robustness, accuracy, and efficiency.

3. Method

3.1. Overview. In this paper, a recognition model PFM to deal with offline single motion segments is proposed first. Combined with graph constraint and online signal segmentation, the model can then be applied to online motion recognition. The method consists of three main key technologies, the structure of which can be found in Figure 1.

Predictive Fusion Model. Sparse and deficient inputs require more relevant information between each input signal sequence to keep local and global information. Therefore, HMMs of different part inputs were constructed, and a probabilistic model was used to fuse these HMMs together so as to enhance the model robustness. An autoregressive process is then introduced, which ensures that the unstable signals can be adjusted based on the past signals and training signals. The model can properly deal with offline motion recognition with sparse and deficient inputs.

Graph Constraint Construction. A graph structure based on the content of motion segments is constructed, limiting the choice for the following motion type based on the current motion content. The graph structure can filter part of motion type, reducing the complexity when dealing with a large motion database and improving the recognition accuracy as well.

Semantics-Based Signal Segmentation. Because input signals are continuous and may consist of multiple motion types, a method to separate the long continuous signal into segments was proposed. This method supports online motion recognition, the basis of the PFMs and graph constraint built offline.

3.2. Predictive Fusion Model. To build a robust learning model that can acquire feature information from sparse and deficient sequential input, HMM shows a high capability of dealing with time series. Here a predictive fusion model is presented based on the structure of HMM, which not only considers the sparse and deficient signal but also considers the features of human motion.

Consider two HMMs with observations $\mathbf{O}_1$ and $\mathbf{O}_2$, which indicate two groups of signal divided from all input sources, respectively. These input sources can be Wiimotes attached to different body parts in our experiments. For each motion type, a corresponding model is needed so as to value the similarity between the current input and the model, and the highest similarity probability determines the input type. Then, the problem can be defined as finding a solution to constructing the connections between the two HMMs so as to provide an optimal estimate for this similarity probability $p(\mathbf{O}_1, \mathbf{O}_2)$. To capture the statistical dependence between two observations $\mathbf{O}_1$ and $\mathbf{O}_2$, the maximum entropy principle is used:

$$p(\mathbf{O}_1, \mathbf{O}_2) = p(\mathbf{O}_1)\, p(\mathbf{O}_2)\, \frac{p(u, v)}{p(u)\, p(v)}, \tag{1}$$

where u and v are the respective transforms of $\mathbf{O}_1$ and $\mathbf{O}_2$ and absorb some dependence between $\mathbf{O}_1$ and $\mathbf{O}_2$. Here, $\langle u, v\rangle$ should be chosen from the two components of HMM, that is, the hidden state $\mathbf{S}$ and the observation $\mathbf{O}$.

Supported by the maximum mutual information criterion in [20], it is better to connect two HMMs by the hidden state sequence for one HMM and the observation sequence for the other one, rather than two hidden states for each one. The structure is shown in Figure 2. Thus, the transforms $\langle u, v\rangle$ can

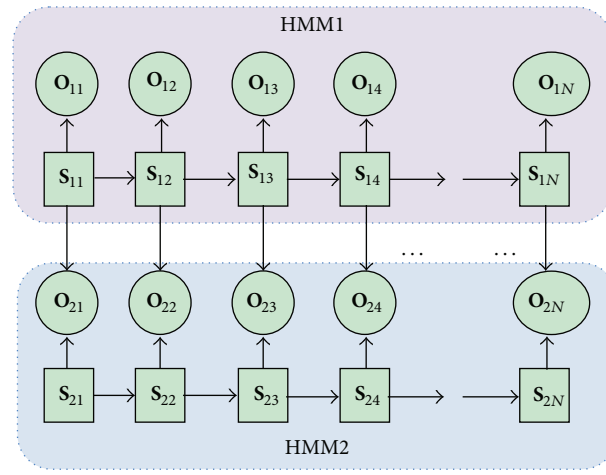


Figure 2: The structure of the model. A fusion relationship has been built between the hidden states of HMM1 and observations of HMM2.

be replaced by $\langle \mathbf{S}_1, \mathbf{O}_2\rangle$ or $\langle \mathbf{S}_2, \mathbf{O}_1\rangle$. The probability defined by (1) yields

$$\begin{aligned} p^1(\mathbf{O}_1, \mathbf{O}_2) &= p(\mathbf{O}_1)\, p(\mathbf{O}_2 \mid \mathbf{S}_1) \\ &= p(\mathbf{O}_1)\, p(\mathbf{O}_2)\, \frac{p(\mathbf{S}_1, \mathbf{O}_2)}{p(\mathbf{S}_1)\, p(\mathbf{O}_2)} \end{aligned} \tag{2}$$

or

$$p^2(\mathbf{O}_1, \mathbf{O}_2) = p(\mathbf{O}_2)\, p(\mathbf{O}_1 \mid \mathbf{S}_2), \tag{3}$$

where the structures defined by (2) and (3) are different. Equation (2) expresses the relationship between $\mathbf{S}_1$ and $\mathbf{O}_2$, indicating that the former HMM is more reliable than the latter one. The reliability of each HMM can be quantified as the weights for each part:

$$p(\mathbf{O}_1, \mathbf{O}_2) = \omega_1 p^1(\mathbf{O}_1, \mathbf{O}_2) + \omega_2 p^2(\mathbf{O}_1, \mathbf{O}_2), \tag{4}$$

where ω_1 and ω_2 represent the reliability of each body part motion. The values of ω_1 and ω_2 are determined by the selected types of actions. For general and daily activities, such as actions in our experiment, ω can be valued as 0.5, while for special occasion and activities, such as ping-pong, where the action focus is on the upper body part, ω can be valued as 0.8 and 0.2.

The observation $\mathbf{O}$ and state $\mathbf{S}$ can be unfolded as $\mathbf{O} = (\mathbf{o}_1, \ldots, \mathbf{o}_t)$, $\mathbf{S} = (\mathbf{s}_1, \ldots, \mathbf{s}_t)$, where t is the length of data sequence. The structure of this model is described in Figure 2.

Vary the basic parameters $\{\pi, \mathbf{A}, \mathbf{B}\}$ in HMM, where π stands for initial probability vector, $\mathbf{A}$ for state transition probability matrix, and $\mathbf{B}$ for observation probability vector, and the new parameters enhance the model's ability to deal with intermittent or noisy $\mathbf{O}$, where the hidden state $\mathbf{S}$ is taken into account in assuming $\mathbf{o}_t$, which can be written in the form of autoregressive process:

$$\mathbf{o}_t = e(\mathbf{s}_{t-1}, \mathbf{s}_t) + \sum_{i=1}^{p} \mathbf{c}_i(\mathbf{s}_{t-1}, \mathbf{s}_t)\, \mathbf{o}_{t-i} + \boldsymbol{\epsilon}_t, \tag{5}$$

where $\mathbf{e}_t$ is a parameter that preserves the descriptive power of the standard HMM when $\mathbf{c}_i = 0$ and $\boldsymbol{\epsilon}$ is residual error when calculating the observation $\mathbf{o}_t$. Since all current observations are affected by the current hidden state and past observations, parameter $\mathbf{B}$ of HMM can be modified as

$$\mathbf{B}^{12}(t) = \frac{1}{\sqrt{(2\pi)^D \left|\mathbf{K}\left(\mathbf{s}_{t-1}, \mathbf{s}_t\right)\right|}} \times \exp\left(-\frac{1}{2}\left(\boldsymbol{\epsilon}_t^{12}\right)^T \mathbf{K}\left(\mathbf{s}_{t-1}^1, \mathbf{s}_t^1\right)^{-1} \boldsymbol{\epsilon}_t^{12}\right), \tag{6}$$

where $\boldsymbol{\epsilon}$ can be calculated from (5).

The methods described above define the model parameters $\varphi = \{\pi_1, \mathbf{A}_1, \mathbf{e}_1, \mathbf{c}_1, \mathbf{K}_1, \pi_2, \mathbf{A}_2, \mathbf{e}_2, \mathbf{c}_2, \mathbf{K}_2, \mathbf{B}_{12}\}$, consisting of two predictive HMM parameters and the dependencies parameter $\mathbf{B}_{12}$. The training process can be summarized as follows.

(1) Calculate the parameters of two predictive HMMs with the expectation-maximization (EM) algorithm presented in [21] and Baum-Welch method in [22]. To maximize $P(\mathbf{O} \mid \boldsymbol{\varphi})$, $\mathbf{A}(\mathbf{s}_i, \mathbf{s}_j)\mathbf{B}(\mathbf{s}_i, \mathbf{s}_j)$ has to be maximized at each time t of the sequence, which can also be written as $\ln A(\mathbf{s}_i, \mathbf{s}_j) + \ln B(\mathbf{s}_i, \mathbf{s}_j)$. The terms that have to be maximized are

$$\sum_{t=1}^{T} \gamma_t\left(\mathbf{s}_i, \mathbf{s}_j\right)\left(\overbrace{\ln \mathbf{A}\left(\mathbf{s}_i, \mathbf{s}_j\right)}^{\text{transition term}} + \overbrace{\ln \mathbf{B}\left(\mathbf{s}_i, \mathbf{s}_j\right)}^{\text{observation term}}\right), \tag{7}$$

where $\gamma_t(\mathbf{s}_i, \mathbf{s}_j)$ is the probability of being in state $\mathbf{s}_i$ at time $t-1$ and in state $\mathbf{s}_j$ at time t in the Baum-Welch algorithm. To solve the terms in (7), the derivatives of the terms with respect to each variable $\mathbf{e}$ and $\mathbf{c}$ must be determined:

$$\begin{aligned} \sum_{t=1}^{T} \gamma_t \mathbf{o}_t - \sum_{t=1}^{T} \gamma_t \mathbf{e} - \mathbf{c} \sum_{t=1}^{T} \gamma_t \mathbf{O}_{\text{prior}} = 0, \\ \sum_{t=1}^{T} \gamma_t \mathbf{o}_t \mathbf{O}_{\text{prior}}^T - \sum_{t=1}^{T} \gamma_t \mathbf{e} \mathbf{O}_{\text{prior}}^T - \mathbf{c} \sum_{t=1}^{T} \gamma_t \mathbf{O}_{\text{prior}} \mathbf{O}_{\text{prior}}^T = 0, \end{aligned} \tag{8}$$

where $\mathbf{O}_{\text{prior}}$ indicates $\{\mathbf{o}_{t-1}, \mathbf{o}_{t-2}, \ldots, \mathbf{o}_{t-p}\}$. The parameters $\mathbf{e}(\mathbf{s}_i, \mathbf{s}_j)$ and $\mathbf{c}(\mathbf{s}_i, \mathbf{s}_j)$ can be calculated by solving (8). The covariance matrix $\mathbf{K}$ can then be calculated using the updated parameters $\mathbf{e}$ and $\mathbf{c}$:

$$\mathbf{K} = \frac{1}{N} \sum_{t=1}^{T} \gamma_t \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^T. \tag{9}$$

(2) Select one predictive HMM as the leading HMM and calculate the hidden state sequence for the leading HMM using the Viterbi algorithm. Then, determine the fusion parameters $\mathbf{B}_{12}$ or $\mathbf{B}_{21}$. If $\mathbf{O}$ is discrete, the following is obtained:

$$\mathbf{b}_i^{12}(j) = \frac{\sum_{t=1}^{T} \delta\left(\mathbf{o}_t^2, j\right) \delta\left(\mathbf{s}_t^1, i\right)}{\sum_{i=1}^{N} \delta\left(\mathbf{s}_t^1, i\right)}, \tag{10}$$

where N is the total hidden state number, j is the clustering number, and δ is the impulse function. When the parameter set φ of the model has been trained, the similarities $p(\mathbf{O}_1, \mathbf{S}_2)$ and $p(\mathbf{O}_2, \mathbf{S}_1)$ can be acquired by forward-backward algorithm, and the similarity $p(\mathbf{O}_1, \mathbf{O}_2)$ can be calculated by (2) or (3).

Then, how to use the model in the process of recognition will be shown. In training process, the input signal sequences $\mathbf{O}$ are four Wiimotes attached to all four human limbs, which are divided into two groups (upper and lower limbs). M models are trained for recognition use, where M indicates the total number of motion types. In the recognition process, the models trained for each motion type are used to compute the model's similarity to the input signal sequence. The solution to the similarity probability $p_M(\mathbf{O}_1, \mathbf{O}_2)$ can be calculated using the same forward algorithm as HMM. If the similarity to any motion exceeds a certain threshold, the sequence is classified as the motion type for which the similarity probability is the largest. The recognition result and similarity probability variation trend are shown in Figure 3. The results indicate that "waving hello" is the motion most similar to the input signal of the six types. Inspect the similarity probability of these models at each time, and it can be found that PFM had a higher classification capacity than the standard model because PFM can be determined timely at 20–40th frames. More experiments with larger databases will be described in Section 4.

3.3. Graph Constraint Construction. The model detailed above can properly identify the motion type from dozens of alternative ones. However, when the number of alternative motion types grows, it not only affects the accuracy rate of recognition but also increases the computation time due to the probability calculations required for each model. Therefore, a structured method was used to reduce the scale of alternative motion types in dealing with a large database.

When a user performs continuous and varied actions, it is noticed that certain action types cannot appear when the current action type has been determined, due to the coordination of human motion. This constraint can be used to guide selection of the following motion type based on the current determinate type.

The present graph model is motivated by the methods of Li et al. [23] and the motion graph of Kovar et al. [24] but different from the methods for different purposes and results. The model can be weighted or unweighted: the weighted one is a directed graph that contains the transition possibility detailing the compatibility and transitivity between two motion types. The node of the graph is of a single motion type, such as "walk" or "run." Before constructing the graph, a training process is necessary to obtain a more precise transition probability. Hundreds of long, continuous human motions are required, and the transition probability is calculated statistically by recording the frequency of motion transitions from one motion type to another. The unweighted graph has a similar structure to the weighted graph, but the transition probability only contains two values $\{0,1\}$. The structure of the graph is shown in Figure 4.

Once the recognition process is underway, the motion type is annotated immediately after recognizing the current

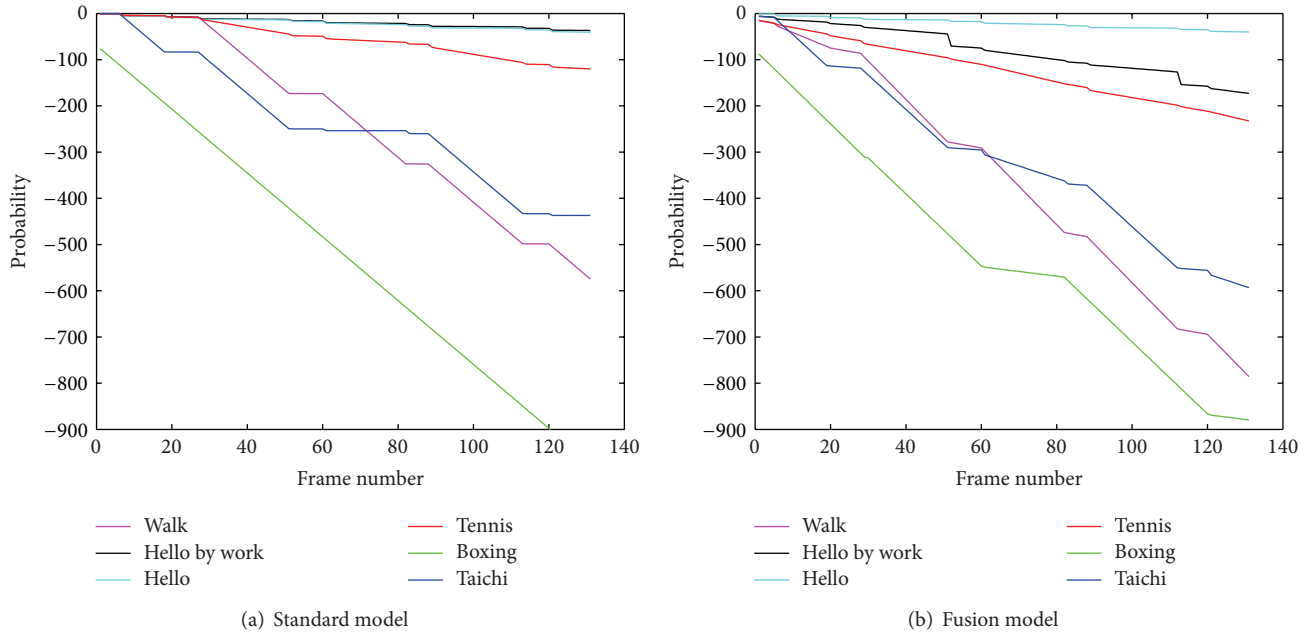
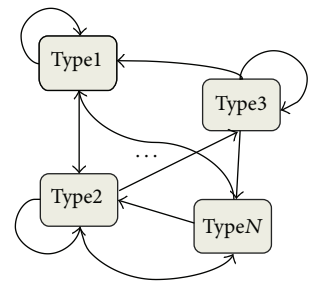


FIGURE 3: The accumulative results of similarity probability for HMM and PFM in a small scale database. The values on the y-axis indicate e^y, taking the logarithm of the probability, and the trend declines over time or frame number.

Graph	Type1	Type2	Type3	⋯	Type$N$
Type1	1	1	0	⋯	0
Type2	0	1	1	⋯	1
Type3	1	0	1	⋯	1
⋯	⋯	⋯	⋯	⋯	⋯
Type$N$	1	1	0	⋯	0

(a)



(b)

FIGURE 4: The constraint built by unweighted graphs. (a) The transition probability between two motion types. (b) The visual graph structure constructed from the table on the left. The transition probability [type x, type y] = 0 indicates that the type x is not permitted to follow the type y.

motion signal segment type i. For the unweighted graph, the nodes which are directed from node type i are selected, and the remaining motion types are excluded without calculating the similarity probability between the upcoming input signal and the current model. Only the models that correspond to selected nodes will calculate the similarity probability. For the weighted graph, the transition probability P_{ij} between two motion types i and j measures their similarities, as follows:

$$P_j(x) = \overbrace{f(P_{ij})}^{\text{transition term}} \overbrace{p_x(O_1)p(O_2)}^{\text{pfm term}}, \tag{11}$$

where x is one of the alternative motion types for the current signal segment and f is a scaling function that reduces the effect of P_{ij}, such as a logarithmic function.

3.4. Online Semantics-Based Signal Segmentation and Motion Recognition. For the online recognition process input signals which are always continuous and long need to be separated into short segments based on different motion types. In recent studies, such as the recursive least squares (RLS) method presented by [25] and the piecewise linear representation (PLR) method presented by [26], signal segmentation problems are always located at the break point in the signal energy curve, which may lead to oversegmentation or skipping smooth transition points. Therefore, signal segmentation based only on the signal shape is not comprehensive and requires consideration of the semantic information in the signal sequence.

In order to combine the semantic information with the segmentation process, the motion content needs to be parsed by a recognition model in the online signal segmentation. PFM is introduced into the process to acquire semantic information. With specific semantic information, it can be ensured that the segmented sequence is an intact and independent

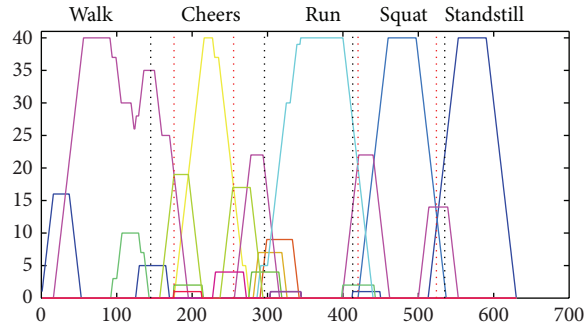


Figure 5: The accumulative process of similarity probability for HMM and PFM in a small scale database. The values on the y-axis indicate votes for each motion type at frame x. The red dotted line indicates segmentation points based on the method, and the black dotted line indicates breakpoints that the human actor expects.

motion type, which can greatly reduce the occurrence of oversegmentation. The method can be described as follows.

Let $\mathbf{O} = \{\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_T\}$ be a long sequence of n-dimensional input acceleration signal vectors and let $v[T][N]$ be a two-dimensional integer vote array of time length T, where N is the number of all motion types. The array $v[t][k]$ indicates the number of votes of type k at frame t, which indicates the current motion type at t. On the online recognition stage, a sliding window of length M ($M << T$) scans the input sequence $\mathbf{O}$ from front to back with a step length of one frame. Each time the window is moved, the PFMs are programmed to recognize the signal segment in the current window M. For example, when the window moves to frame i, similarity probabilities are calculated by PFMs of alternative types with input of signals from $\mathbf{o}_i$ to $\mathbf{o}_{i+M}$. The vote array $\{v[i][k], v[i+1][k], \ldots v[i+M][k]\}$ will then be increased by 1, where k is the winner type in the present recognition. After the window sliding to frame p, N curves can be drawn based on $\{v[1 \cdots p][1], v[1 \cdots p][2], v[1 \cdots p][N]\}$ before frame p, which is shown in Figure 5. The intersection points shown in Figure 5 can be classified as alternative segment points, and recognition results can be acquired after finishing the segmentation.

To deal with transition signals and signals that do not belong to any alternative motion type, an appropriate threshold for each PFM should be set to filter out the redundant segment points during the PFM training process. The threshold is defined as the minimum normalized probability in the training dataset, and it rejects motion signals dissimilar to the training set.

The method presented above considers the semantic information of signal sequence and acquires the recognition result based on the PFMs trained offline. The recognition process is online, and results of which will be discussed in Section 4.

4. Results and Discussions

In this section, the functions of PFM, the effect of online recognition, and various applications of this technology will be described. As is presented in the last section, the input devices used in our experiment are sparse and low-cost (see Table 1). Devices with more information provided always result in higher price. Several general portable input devices are shown in the table, and sparser and cheaper devices are chosen to conduct our experiment. The signals analyzed here were the linear accelerations without any denoising or angular information, making it difficult to calculate accurate position information, as Table 2 shows.

The Wiimotes transmitted signals to a computer via a bluetooth interface that supports an 8–10 meters distance during an experiment. The sampling time in our experiment was 25 fps, which can be adjusted to accommodate a range of precisions. The training motion signal database has been preliminary constructed, which is clustered as 28 nodes in graph structure based on the content of the motion signal segments. Each node consists of 3-4 groups of motion segments with different variants, such as walking in different styles or kicking to different positions. Each type of motion signal is captured 5 times by 4 different actors. These hundreds of motion signals are well-organized for model training. In the experiment, thousands of independent action signals and hundreds of long continuous action signals are performed by testers in real time to get the result on recognition rate, robustness, and so forth.

4.1. Performance of Model. Before the experiment, we have tested several state-based methods, such as coupled HMM and structural HMM, as Pan et al. [20] presented. The result shows that fused HMM presents a better accuracy and robustness to the others when dealing with sparse and deficient motion signals. Therefore, in this section, the functions of our PFM will be shown and the accuracy and efficiency of the recognition process will be only compared with the performance of traditional Gaussian HMM and fused HMM when dealing with sparse and deficient input.

In our experiment, the recognition effect of different actors was validated by leave-one-out and k-fold cross validation methods, and the recognition rates of the PFM are shown in Figure 6, based on 40 alternative action types from the database we built. The HMM method yielded an average recognition rate of 42%, lower than the fused HMM and PFM recognition rates. The horizontal axis in Figure 6 represents the type of input signal sample and the vertical axis represents the types of corresponding models we built. While HMM is not robust when dealing with certain special motions, the PFM presents a more robust and accurate recognition result. In the HMM, without considering the motion of different body parts, the combined acceleration information led to confusion and presented a worse classification capability than for motions of similar variance. The fused HMM considered the structure of human motion and presented a higher classification capability than the standard model. The prediction capabilities of PFM were much better for these special inputs.

The proposed model can handle imperfect signals as well as deletion of input signals. The fewest number of sensors that can retain complete full-body motion information remains to be determined. Further experiments will be conducted to

Table 1: The details of general current portable input devices applied to motion control and recognition.

Sensor	Amount	Per price	Output information
Wii Remote	4	$39.99	3D linear accelerations, 2D rotation angle
Xsens' MTx [14, 16]	4 or more	$1500	Orientation, linear accelerations, angular velocity
MEMS sensors [17, 18]	8 for gait analysis	$250–8000	3D angular velocity, Orientation, etc
HD Hero [19]	16 or more	$250	Scene videos

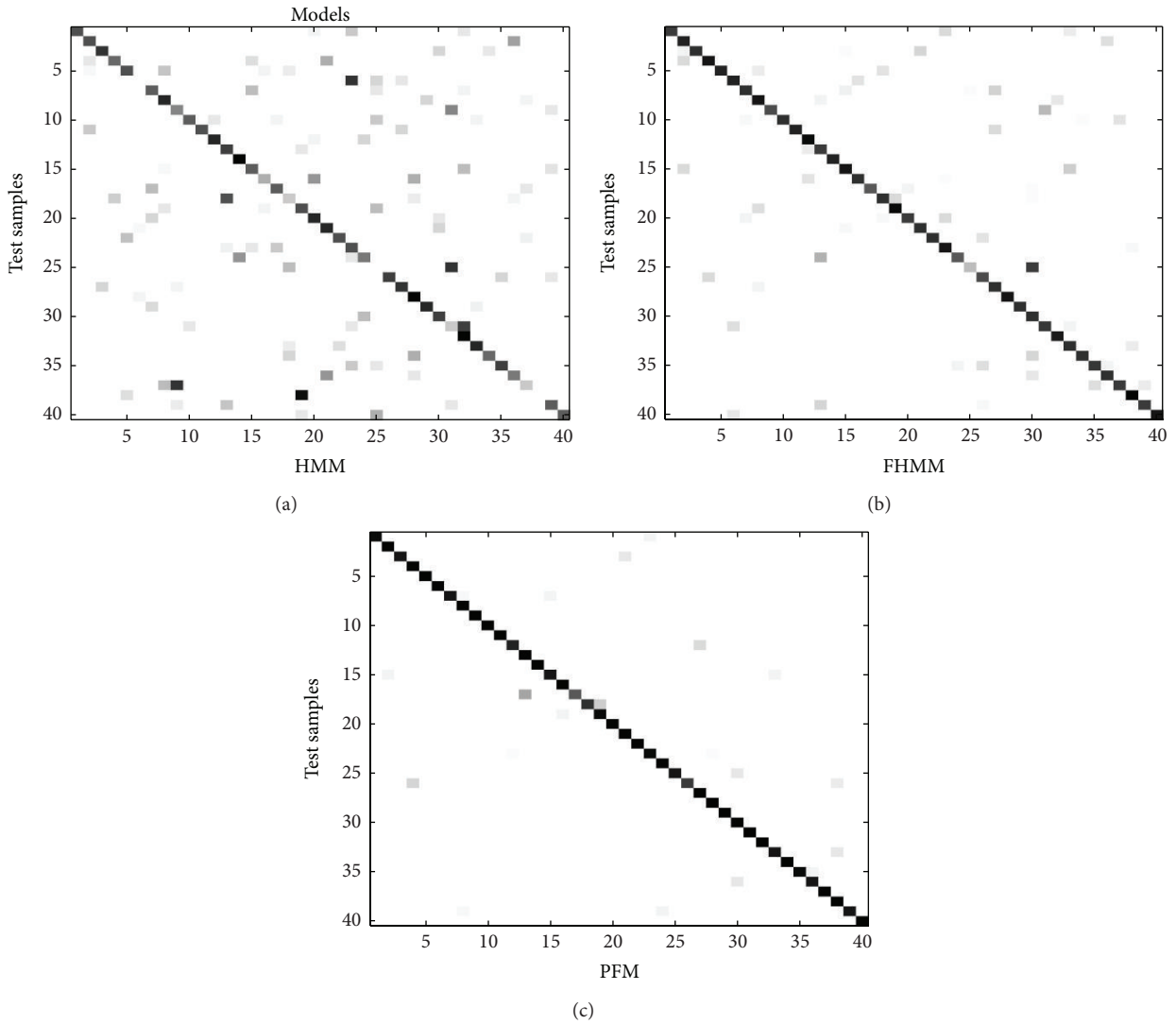


Figure 6: The recognition rate shown in the form of confusion matrices. With an increase in the recognition rate, the color of the matrix grid varies gradually from white to black.

Table 2: The offsets between actual motion data and position information, calculated by incomplete accelerations as the frame number increases.

Frame number	10	50	100	150	200
RMS value (cm)	9.55	22.13	49.68	77.86	122.46

Table 3: The PFM recognition rate for different actors with an increasing fraction of signal deletion.

Actor	Trained actor	New actor 1	New actor 2
Completed signals	0.97	0.92	0.91
One intermittent Wii	0.94	0.87	0.89
Two intermittent Wiis	0.85	0.84	0.85
One missing Wii	0.84	0.78	0.75

show the robustness and capabilities of dealing with deficient signals and to determine the requisite number of sensors in order to properly function in the motion recognition process. Table 3 shows the model's robustness with respect to deficient signals. In this experiment, different actors attached with reducing input devices are chosen. The actors here

Table 4: The accuracy rate of our segmentation method for different actors.

Actor	Actor 1	Actor 2	Actor 3	Actor 4	Actor 5
Desired points	48/50	50/50	47/50	50/50	49/50
Redundant points	8	10	5	12	7
Accuracy rate	0.86	0.83	0.9	0.8	0.87

include both trainees and newcomers. Since the action data of trainees are more standard and similar to trained motion data, the recognition rates for trainees are slightly higher than those for newcomers as the table shows. In the event that a short signal sequence from one sensor is lost, the recognition results remain unchanged from those derived from the complete signal sequence. For trained actors, the average recognition rate of HMM is 41% when two Wiis are intermittent and 73% for FHMM. This comparison shows that the classifying abilities of the PFM are greater than those of the two methods.

An analysis of unknown motions not included in the training datasets provides an estimate for the maximal probability of the motions most likely to be in the training datasets. Evaluation methods demonstrate the accuracy of the input signal relative to the recognition results.

4.2. Online Recognition. In an online recognition system, continuous signal processing is key for completing the task, and the results are essential for influencing and evaluating the recognition process. In our experiments, five actors were required to perform a continuous motion that included 51 motion segments used to test the segmentation accuracy rate. The accuracy rate of the segmentation experiment was evaluated by the number of desirable missing segmentation points and the number of undesirable or redundant segmentation points. Table 4 presents the segmentation results for these two measures. The desired segmentation points can be always located properly in our method, and the main factor that reduces the accuracy is the redundant segmentation points for our method. Unlike current segmentation methods of human motion signal sequence, that is, the RLS method and PLR method, whose abundant parameter and threshold groups are determined by repeated adjustments, our semantics-based method is more independent of parameters. For all this, a large number of desirable missing segmentation points for these two methods with an appropriate parameter group are always one of the main factors which may affect the accuracy rate of the segmentation. Besides, the delay of segmentation points and the accumulation of errors which always appear in these two methods can be effectively avoided in our method. Taking actor 3, for example, the desirable segmentation points are 41 for RLS and 36 for PLR, and the redundant segmentation points are 17 and 22, which is also more than semantics-based method. Figure 7 shows the final accuracy rate of these methods.

When dealing with large databases of alternative motion types, the difficulty in distinguishing features between different motion types becomes greater. The recognition capability of PFM is reduced substantially (see Table 5). Based on the graph structure we built, the alternative types of current segment recognition were fewer than the total alternative types, thereby preserving the online recognition accuracy rate rather efficiently. The high classification capability of the PFM model ensures that the results can be efficiently acquired at 30–50 frames of the signal input before the actor finishes the motion.

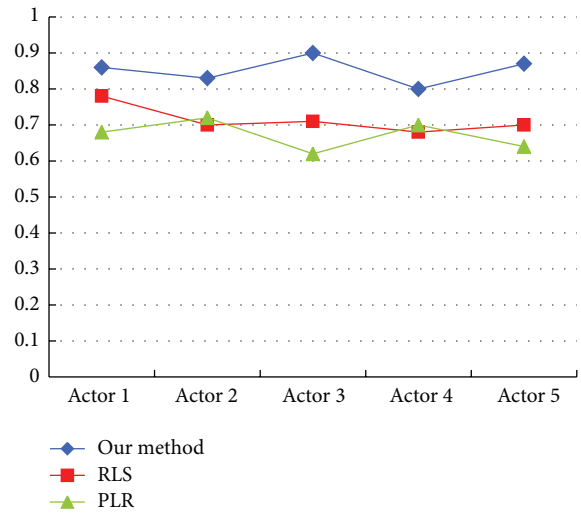


Figure 7: The segmentation accuracy rate of three methods. A long, continuous motion was performed by five actors.

4.3. Applications. The methods proposed here are applicable to a wide variety of applications, including behavioral teaching evaluations, interactive games in virtual environments, and activity validation systems in large-scale scenes.

A general application of the proposed recognition method includes driving the virtual human to generate computer animations or to simulate a virtual environment for user interactions. After the user performs the continuous motions the segmentation and recognitions are conducted efficiently, and the recognition results guide the searching process of the corresponding motion data in the database. The blending process in the motion graph technology guarantees continuity of the generated motion. Generative models, such as the Gaussian latent variable model presented by [27], can be properly embedded to synthesize more delicate and stylized motion in various applications.

In the context of educational applications, the present method can be used to evaluate activities, such as playing tennis, doing martial arts, or dancing. Students can act out motions while following a standard motion sequence that is presented in advance. The system can then evaluate the similarity of the mimicked sequence to the standard sequence.

Table 5: The recognition rate of PFM and PFM with graph constraint for trained actors with an increasing number of alternative motions.

Total alternative types	40	55	70	85	100
PFM	0.92	0.85	0.6	0.51	0.39
PFM with graph constraint	0.95	0.94	0.91	0.85	0.85

An evaluation system can be constructed by calculating the probability ratio between the input motion signal and the normative training data. The ratio provides an important evaluation criterion. The weights of the fusion model may be adjusted to standardize the motions of each appendage. Figure 8 shows an experiment based on the evaluation system described here. The proposed method was adapted to a set of complex motions associated with tennis, Tai chi, and boxing. The motions performed by the user were recognized and evaluated using our method.

Complex virtual environmental interactions constitute the main application focus of our method. Virtual environment games and special training regimens require environmental immersion and interactions with virtual objects. Our method, based on sparse, low-cost sensors, performed well in the context of these applications and can provide the user with an immersed experience.

Figure 8: The motion evaluation system. Left: an actor performs exercises with accelerometers attached to her four limbs. Right: the recognition result and the actor's performance grade.

5. Conclusion and Future Works

This paper presents a full-body motion recognition method based on sparse, low-cost accelerometers. In the online recognition process, a semantics-based signal segmentation method was adopted to acquire short motion segments, and a motion transition graph structure was constructed to reduce the amount of alternative motion types. To recognize the motion type accurately, a predictive fusion model was presented to efficiently distinguish between current motion types and alternative motion types. The models recognition capability is robust and accurate in dealing with unstable and deficient signals that provide little information for reconstructing position information. Results show that the method has a high recognition rate and can be adapted to specific input signals.

During experiments, it is found that the method had difficulty identifying the actors' orientation, as the input devices we used lack direction information for recovering whole motion information. In addition, a short pause in a continuous motion occasionally led to a redundant motion segment. In the future, in order to overcome these problems low-cost sensors will be integrated that will also provide direction information so that the input device can be more conveniently adapted to a specific interaction. The database of the motion signals and the motion data will also be expanded. Ultimately, the method will be applied to complicated scene interactions between users and the virtual environment.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the Natural Science Foundation of China (Grant no. 61170186) and the Zhejiang Leading Team of Science and Technology Innovation (2011R50019-06). The data used was obtained from HDM05 in [28] and CMU public database.

References

[1] R. Poppe, "A survey on vision-based human action recognition," *Image and Vision Computing*, vol. 28, no. 6, pp. 976–990, 2010.

[2] J. Ning and F. Mokhtarian, "Human motion recognition based on statistical shape analysis," in *Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS '05)*, pp. 4–9, September 2005.

[3] H. Zhou, L. Wang, and D. Suter, "Human motion recognition using gaussian Processes classification," in *Proceedings of the 19th International Conference on Pattern Recognition (ICPR '08)*, pp. 1–4, IEEE, December 2008.

[4] J. Min, Y. Chen, and J. Chai, "Interactive generation of human animation with deformable motion models," *ACM Transactions on Graphics*, vol. 29, no. 1, article 9, 2009.

[5] Y.-C. Lai, H. M. Liao, C.-C. Lin, J. R. Chen, and Y.-P. Luo, "A local feature-based human motion recognition framework," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS '09)*, pp. 722–725, IEEE, Taipei, Taiwan, May 2009.

[6] J. Mlich, "Wiimote gesture recognition," in *Proceedings of the 15th Conference and Competition STUDENT EEICT*, vol. 4, pp. 344–349, Faculty of Electrical Engineering and Communication BUT, 2009.

[7] T. Schlömer, B. Poppinga, N. Henze, and S. Boll, "Gesture recognition with a Wii controller," in *Proceedings of the 2nd International Conference on Tangible and Embedded Interaction (TEI '08)*, pp. 11–14, Bonn, Germany, February 2008.

[8] P. Koch, W. Konen, and K. Hein, "Gesture recognition on few training data using slow feature analysis and parametric bootstrap," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '10)*, pp. 1–8, Barcelona, Spain, 2010.

[9] J. Pang and I. Singh, "Accelerometer based real-time remote detection and monitoring of hand motion," in *Proceedings of the World Congress on Engineering and Computer Science (WCECS '11)*, vol. 2 of *Lecture Notes in Engineering and Computer Science*, pp. 2078–2095, San Francisco, Calif, USA, 2011.

[10] L. Sun, D. Zhang, B. Li, B. Guo, and S. Li, "Activity recognition on an accelerometer embedded mobile phone with varying positions and orientations," in *Ubiquitous Intelligence and Computing*, pp. 548–562, Springer, Berlin, Germany, 2010.

[11] T. Shiratori and J. K. Hodgins, "Accelerometer-based user interfaces for the control of a physically simulated character," *ACM Transactions on Graphics*, vol. 27, no. 5, article 123, 2008.

[12] F. Niu and M. Abdel-Mottaleb, "HMM-based segmentation and recognition of human activities from video sequences," in *Proceedings of the International Conference on Multimedia and Expo (ICME '05)*, pp. 804–807, IEEE, July 2005.

[13] A. M. Khan, Y. Lee, S. Y. Lee, and T. Kim, "A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 5, pp. 1166–1172, 2010.

[14] J. Tautges, A. Zinke, B. Krüger et al., "Motion reconstruction using sparse accelerometer data," *ACM Transactions on Graphics*, vol. 30, no. 3, article 18, 2011.

[15] C. Wong, Z. Zhang, R. Kwasnicki, J. Liu, and G.-Z. Yang, "Motion reconstruction from sparse accelerometer data using PLSR," in *Proceedings of the 9th International Workshop on Wearable and Implantable Body Sensor Networks (BSN '12)*, pp. 178–183, May 2012.

[16] D. T. H. Lai, M. Hetchl, X. Wei, K. Ball, and P. Mclaughlin, "On the difference in swing arm kinematics between low handicap golfers and non-golfers using wireless inertial sensors," *Procedia Engineering*, vol. 13, pp. 219–225, 2011.

[17] B. Huyghe, P. Salvo, J. Doutreloigne, and J. Vanfleteren, "Feasibility study and performance analysis of a gyroless orientation tracker," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 8, pp. 2274–2282, 2012.

[18] C. Yang and Y. Hsu, "A review of accelerometry-based wearable motion detectors for physical activity monitoring," *Sensors*, vol. 10, no. 8, pp. 7772–7788, 2010.

[19] T. Shiratori, H. S. Park, L. Sigal, Y. Sheikh, and J. K. Hodginsy, "Motion capture from body-mounted cameras," *ACM Transactions on Graphics*, vol. 30, no. 4, article 31, 2011.

[20] H. Pan, S. E. Levinson, T. S. Huang, and Z.-P. Liang, "A fused hidden Markov model with application to bimodal speech processing," *IEEE Transactions on Signal Processing*, vol. 52, no. 3, pp. 573–581, 2004.

[21] P.-V. Borza, *Motion-based gesture recognition with an accelerometer [Ph.D. thesis]*, Babeş-Bolyai University, Cluj-Napoca, Romania, 2008.

[22] P. Kenny, M. Lennig, and P. Mermelstein, "Linear predictive HMM for vector-valued observations with applications to speech recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 2, pp. 220–225, 1990.

[23] W. Li, Z. Zhang, and Z. Liu, "Expandable data-driven graphical modeling of human actions based on salient postures," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1499–1510, 2008.

[24] L. Kovar, M. Gleicher, and F. Pighin, "Motion graphs," *ACM Transactions on Graphics*, vol. 21, no. 3, pp. 473–482, 2002.

[25] K. Momen and G. R. Fernie, "Automatic detection of the onset of nursing activities using accelerometers and adaptive segmentation," *Technology and Health Care*, vol. 19, no. 5, pp. 319–329, 2011.

[26] E. Keogh, S. Chu, D. Hart, and M. Pazzani, "An online algorithm for segmenting time series," in *Proceeding of the 1st IEEE International Conference on Data Mining (ICDM '01)*, pp. 289–296, San Jose, Calif, USA, December 2001.

[27] S. Levine, J. M. Wang, A. Haraux, Z. Popović, and V. Koltun, "Continuous character control with low-dimensional embeddings," *ACM Transactions on Graphics*, vol. 31, no. 4, article 28, 2012.

[28] M. Müller, T. Röder, M. Clausen, B. Eberhardt, B. Krüger, and A. Weber, "Documentation mocap database hdm," Technical Report CG-2007-2, Universität Bonn, Bonn, Germany, 2007.