International Journal of Computational Vision and Biomechanics
Vol. 5 No. 1 (January-June, 2019)

MUK PUBLICATIONS
Open Access Publisher

Automatic Learning of 3D Pose Variability in Walking Performances for Gait Analysis

Ignasi Rius†, Jordi Gonzàlez‡, Mikhail Mozerov†, F. Xavier Roca†

† *Centre de Visió per Computador, Edifici O. Campus UAB. 08193, Bellaterra, Spain.*
{irius, mozerov, xavir}@cvc.uab.es

‡ *Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Barcelona, Spain*
poal@iri.upc.edu

Submitted: 26th September 2018 *Revised: 29th November 2018* *Accepted: 23rd January 2019*

This paper proposes an action specific model which automatically learns the variability of 3D human postures observed in a set of training sequences. First, a Dynamic Programing synchronization algorithm is presented in order to establish a mapping between postures from different walking cycles, so the whole training set can be synchronised to a common time pattern. Then, the model is trained using the public CMU motion capture dataset for the walking action, and a mean walking performance is automatically learnt. Additionally, statistics about the observed variability of the postures and motion direction are also computed at each time step. As a result, in this work we have extended a similar action model successfully used for tracking, by providing facilities for gait analysis and gait recognition applications.

Keywords: Human motion modeling, gait analysis and recognition, dynamic programming.

1. INTRODUCTION

Human motion analysis has received great attention from the research community during the past years. The promising applications it brings comprise automatic video surveillance, gait recognition, human body tracking, automatic video annotation, realistic motion synthesis, sports performance and medical applications among others. At present, there exist a lot of publications related to this wide and relatively-old research area [27, 25, 1] due to the number of involved tasks, which is directly proportional to the huge number of potential applications.

The nature of the open problems and techniques used in human motion analysis approaches strongly depend on the goal of the final application. Hence, most approaches oriented to surveillance demand performing activity recognition tasks in real-time dealing with illumination changes and low-resolution images. Thus, they require robust techniques with a low computational cost, and mostly, they tend to use simple models and fast algorithms to achieve effective segmentation and recognition tasks in real-time. Additionally, unlike applications which require finding body parts, most approaches treat the image as a whole and extract 2D features which are fed into classification schemes to provide the most plausible explanation of what is happening in the scene [7, 20]. Complementarily, other video-surveillance approaches are aimed to discover unusual or unseen situations, trigger an alarm when such situations are detected, and let a human operator supervise the scene. An example of this kind of systems is [8] where the system is designed to supervise a swimming pool environment so an alarm can be triggered in case there is a water-related situation. They extract several features such as speed, posture, submersion time, etc. from each of the tracked objects within the surveillance perimeter, and fed them into a polynomial network in order to detect emergency events.

In contrast, approaches focused to 3D tracking and reconstruction, require to deal with a more detailed representation about the current posture that the human body exhibits [19, 22, 6, 23]. The aim of full body tracking is to recover the body motion parameters from image sequences dealing with 2D projection ambiguities, occlusion of body parts, and loose fitting clothes among others. Thus, they require human body models able to capture the relative positions between joints and limbs. Towards this end, an "stick figure" model [14] is usually used to represent the human body configuration, where body parts are represented as segments which are connected by joints with a predefined number of Degrees of Freedom (DOF). Additionally, the stick-figure model can be fleshed out by using volumetric primitives such as cylinders, truncated cones or ellipsoids in order to model the surface of the human body [26, 5]. The number of segments and joints affects the complexity of the model, which in turn, is strongly determined by the final goal of the application.

On the other hand, gait analysis applications demand methods suitable for comparing motion sequences between individuals, between the same subject, and w.r.t. some universal representation of the same motion. They may be based on the detailed analysis of body parts trajectories [10], or in the extraction of characteristic simple image-based features for each individual from the image sequences [15, 11]. Similarly, some gait identification approaches use the information from joint trajectories, according to Johansson's studies from the early 70's [12] pointing out that the motion of the joints provides the key to recognize the behaviour and the identity of the whole figure. Other approaches to gait recognition are based on appearance cues of the individuals [2, 4, 17, 13]. For instance, in [13] they present two methods for identification of humans using gait. They extract a binary silhouette of the individual and compute the width of its outer over time. Then, these features are fed into an Hidden Markov Model (HMM) for classification.

Finally, motion synthesis applications usually deal with complex models having a large number of DOF [16, 22, 24]. Here, the pursued objective is to provide realism and natural motion to animations rather than merely describing the motion performed. For example, in [24] they use a database of pre-recorded motion capture sequences and learn an statistical model for segments of the original motion capture data. Then, they are able to re-use previously recorded motion subsequences in the actual animation, providing realism and soft transitions between motions.

Complementarily, we present an action-specific model of human motion suitable for many applications, that has been successfully used for full body tracking [19, 18]. In this paper, we explore and extend its capabilities for gait analysis and recognition tasks. Additionally, we present a method for synchronizing similar motion sequences in order to allow comparison between them. Our action-specific model is trained with 3D motion capture data for the walking action from the Carnegie Mellon University's (CMU) Graphics Lab Motion capture database. In our work, human postures are represented by means of a full body 3D model composed of 12 limbs. Limbs' orientations are represented within the kinematic tree using their direction cosines [28]. As a result, we avoid singularities and abrupt changes due to the representation. Moreover, near configurations of the body limbs account for near positions in our representation at the expense of extra parameters to be included in the model. Then, Principal Component Analysis (PCA) is applied to the training data to perform dimensionality reduction over the highly correlated input data. Additionally, the main modes of variation of human gait are naturally represented by means of the principal components found. This leads to a coarse-to-fine representation of human motion which relates the precision of the model with its complexity in a natural way, and makes it suitable for different kind of applications which demand more or less complexity in the model.

Subsequently, all the walking performances are synchronised using a Dynamic Programming (DP) algorithm and a mean manifold for a set of training performances is computed. As a result, we can analyse intra-performance differences in each time step. In other words, we can quantify the difference between the same part of two different performances of the same action, enabling to achieve gait analysis for sports performance or medical applications among others. Finally, we learn a mean direction of motion for subsequences of a determined length, and extract statistics from the synchronised dataset that characterise the variation observed in each step between different training performances. This leads, together with the computed mean performance, to gait identification applications since we can establish classification boundaries according to the variation observed from the mean performance. Both the action-specific model and the synchronization algorithm constitute the main contribution of this paper.

The remainder of this paper is organised as follows. Section 2 details the composition of the motion database used for training, the human body model employed, and explains the method used for synchronising the whole training set. Then, Section 3 describes the action specific model and explains the procedure for learning its parameters from the synchronised training set. Section 4 introduces how this model is used for gait analysis and gait recognition applications and some experimental results are shown. Finally, Section 5 concludes this paper and outlines the future research lines.

2. MOTION DATABASE SYNCHRONIZATION

In order to train and test our approach, we used the CMU Graphics Lab Motion capture database. The motion data was acquired at 120 fps with a Vicon Motion Capture System, using a 41 markers set. The database contains a total of 2622 performances classified in 23 different motion categories such as walking, boxing or running, and were performed by different subjects. We encourage the reader to refer to their website for further details on the acquisition procedure, markers' positions and database organization.

2.1 Human Body Model

The body model employed in our work is composed of twelve rigid body parts (hip, torso, shoulder, neck, two thighs, two legs, two arms and two forearms) and fifteen joints, see Fig. 1(a). These joints are structured in a hierarchical manner, constituting a kinematic tree, where the root is located at the hip.

However, postures in the CMU database are represented using the *XYZ* position of each marker that was placed to the subject in an absolute world coordinates system. Therefore, we must select some principal markers in order to make the input motion capture data usable according to our human body representation. Figure 1(b) relates the absolute position of each joint from our human body model with the markers' used in the CMU database. For instance, in order to compute the position of joint 5 (head) in our representation, we should compute the mean position between the RFHD and LFHD markers from the CMU database, which correspond to the markers placed on each side of the head. Notice that our model considers the left and the right parts of the hip and the torso as a unique limb, and therefore we require a unique segment per each. Hence, we compute the position of joints 1 and 4 (hip and neck joints) as the mean between the previously computed joints 2 and 3, and 6 and 9 respectively.

We use directional cosines to represent relative orientations of the limbs within the kinematic tree [28]. As a result, we represent a human body posture ψ using 36 parameters, i.e.

$$\psi = \{\theta_1^x, \theta_1^y, \theta_1^z, ..., \theta_{12}^x, \theta_{12}^y, \theta_{12}^z\}, \tag{1}$$

where $\theta_l^x, \theta_l^y, \theta_l^z$ are the relative directional cosines for the limb l, i.e. the cosine of the angle between a limb l and each axis x, y, and z respectively. Subsequently, let us define a particular performance Ψ_i of an action as a time-ordered sequence of F_i postures such as

$$\Psi_i = \{\psi_i^1, ..., \psi_i^{F_i}\}, \tag{2}$$

where the index i denotes the number of performance. Finally, an action $A_k = \{\Psi_1, ..., \Psi_{I_k}\}$ is defined by all the I_k performances that belong to that action.

Directional cosines constitute a good representation method for body modeling, since it does not lead to discontinuities, in contrast to other methods such as Euler angles or spherical coordinates. Additionally, unlike quaternions, they have a direct geometric interpretation. However, given that we are using 3 parameters to determine only 2 DOF for each limb, such representation generates a considerable redundancy of the vector space components. Additionally, the human body motion is intrinsically constrained, and these natural constraints lead to highly correlated data in the original space. Therefore, we aim to find a more compact representation of the original data to avoid redundancy. To do this, we consider a set of performances corresponding to a particular action A_k, and perform PCA to all the postures that belong to that action. Then, we project all the training postures to the PCA space, i.e.

$$\tilde{\psi} = [\mathbf{e}_1, ..., \mathbf{e}_b]^T (\psi - \overline{\psi}), \tag{3}$$

where ψ refers to the original posture, $\tilde{\psi}$ denotes the lower-dimensional version of the posture represented in the PCA space, $[e_1,..., e_b]$ is the PCA space transformation matrix that correspond to the first b selected eigenvectors, and $\tilde{\psi}$ is the mean of all the postures. The resulting PCA-like space where postures are represented will be denoted as Ω^{A_k}. As a result, we obtain a lower-dimensional representation of human postures which is more suitable to describe human motion, since we found that each dimension on the PCA space describes a natural mode of variation of human motion [9]. Choosing different values for b lead to models of more or less complexity in terms of their dimensionality. Hence, while the *gross-motion*[2] is explained by the very first eigenvectors, subtle motions in the PCA space representation requires more eigenvectors to be considered. The projection of the training sequences into the PCA space will constitute the input for our sequence synchronization algorithm.

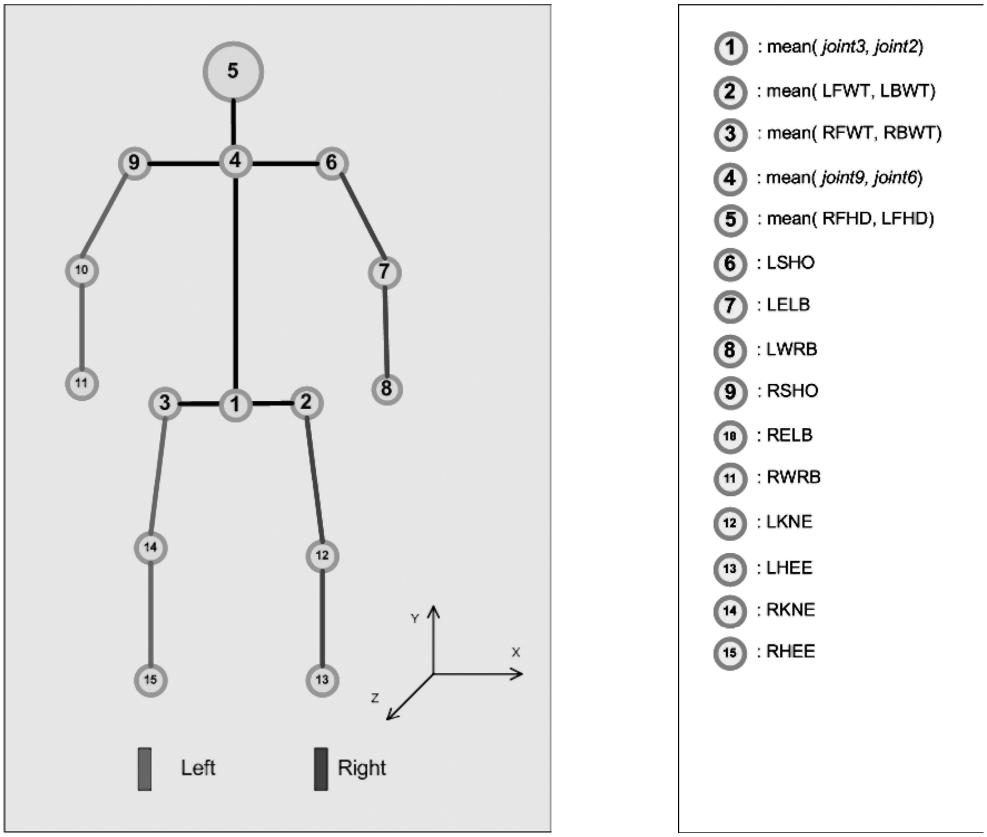

Figure 1: Details of the human body model used (a) and the relationship to the markerset employed in the CMU database (b).

2.2 Composition of the Training Set

Given that we are focused in modeling the walking action, we only use the walking sequences from the CMU database. As a result our training set is composed of 12 subjects showing different performances of the walking action. In turn, each walking performance consists in a variable number of cycles ranging from 1 to 5. Subsequently, each recorded performance is split into its composing walking cycles. We used the angle between the left and right legs as the criterion for splitting walking cycles. A full cycle is defined as all the body postures in between two consecutive maximums of the angle between both legs when the left leg remains in the back. Incomplete cycles and erroneous sequences were discarded from the training set. As a result, we finally end up with a set of 16891 body postures corresponding to 126 walking cycles performed by 12 different actors showing different speeds and different body configurations while performing the same action. Table 1 details the composition of our training set. The number of each subject and recorded performance corresponds to the same indexes used in the CMU database.

Table 1
Detail of the Training Set Composition

Subject id.	*Index of selected performances*	*# recorded performances*	*Total # of walking cycles*	*Total # body postures*
2	{1, 2}	2	3	372
5	{1}	1	3	448
7	{1, 2, 3, 6, 7, 8, 9, 10 ,11}	9	15	2027
8	{1, 2, 3, 6, 9, 10}	6	9	1058
12	{3}	1	3	482
16	{15, 16, 21, 22, 31, 32, 47}	7	15	1977
35	{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 28, 29, 30, 31, 32, 33, 34}	23	42	5782
38	{1, 2}	2	4	540
39	{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14}	13	26	3260
43	{1}	1	2	263
49	{1}	1	3	491
55	{4}	1	1	191
Total		67	126	16891

2.3 Synchronization Algorithm

As stated before, the training sequences are acquired under very different conditions, showing different durations, velocities and accelerations during the performance of a particular action. As a result, it is difficult to perform useful statistical analysis to the raw training set, since we cannot put in correspondence postures from different cycles of the same action. Therefore, a method for synchronizing the whole training set is required so that we can establish a mapping between postures from different cycles.

Inspired by techniques used in the stereo-matching and image procesing literature [3, 21], we developed a novel dense matching algorithm based on Dynamic Programming (DP), which allows us to find an optimal solution for synchronizing the pre-recorded motion sequences of the same class in the presence of different speeds and accelerations. Towards this end, we first compute the similarity between each pair of training sequences with a given metric. Then, in order to extract from the input data set the best time scale pattern for synchronization, an intra-class minimum global distance criterion is used. Finally, all walking cycles are synchronised to the computed time pattern. The detailed explanation of the process is as follows.

The projection of the training sequences into the PCA space constitutes the input for our sequence synchronization algorithm. Hereafter, we consider a multidimensional signal $\mathbf{x}_i\ (t)$ as an interpolated expansion of each training performance $\tilde{\Psi}_i$ such as

$$\mathbf{x}_i(t) = \tilde{\psi}_i^f \quad if \quad t = (f-1)\delta f; \qquad f = 1,\ldots,F; \tag{4}$$

where the time domain of each action performance $\mathbf{x}_i(t)$ is $[0, T)$.

Before starting synchronising the dataset, all the walking cycle performances are resampled, using cubic spline interpolation, so that all the performances have exactly the same number of frames F. The longest performance from the training set is chosen to be the one which determines the number of frames F of the rest of the set. As a result, all the input sequences $\mathbf{x}_i(t)$ have the same period T.

The problem of synchronizing two multidimensional signals $\mathbf{x}_n(t)$ and $\mathbf{x}_m(t)$ is similar to the matching problem of two epipolar lines in a stereo image. For stereo matching a Disparity Space Image (DSI) representation is usually employed [3, 21]. The DSI approach assumes that a 2D DSI matrix has dimensions time p and and disparity d, ranging from $0 \le p < P$, and $-D \le d \le D$. Let $E(d, p)$ denote the DSI cost value assigned to each DSI matrix element (d, p) calculated by

$$E_{n,m}(p,d) = \left|\mathbf{x}_n(p\delta t) - \mathbf{x}_m(p\delta t + d\delta t)\right|^2, \tag{5}$$

where δt stands for the time sampling interval used.

Consequently, we formulate the synchronization task as an optimization problem as follows: find the time-disparity function $\Delta_{n,m}(p)$, which minimizes the synchronization distance between the compared signals $\mathbf{x}_n$ and $\mathbf{x}_m$, i.e.

$$\Delta_{n,m}(p) = \arg\min_{d} \sum_{i=0}^{<P} E_{n,m}(i, d(i)) + \mu \sum_{i=0}^{<P-1} \left|d(i+1) - d(i)\right|. \tag{6}$$

The discrete function $\Delta_{n,m}(p)$ coincides with the optimal path through the DSI trellis. In other words, we must find the path whose sum of cost values plus its weighted length is minimal among all other possible paths. This is solved efficiently by using the Dynamic Programming. The method consists of an step-by-step control and optimization given by the following recurrence relation:

$$S(p,d) = E(p,d) + \min_{k\in 0,\pm 1}\{S(p-1, d+k) + \mu 1 d + k 1\},$$

$$S(0,d) = E(0,d), \qquad (7)$$

where the scope of the minimization parameter is chosen in accordance with $\left|\Delta_{n,m}(p+1) - \Delta_{n,m}(p)\right| \le 1$. By using that recurrence relation, the minimal value of the objective function in Eq.(6) can be found at the last step of optimization. Next, the algorithm works in reverse order and recovers a sequence of optimal steps (stored in a lookup table K(p,d) for the values of the index k in the recurrence relation given by Eq. (7)) and eventually the optimal path, given by

$$d(p-1) = d(p) + K(p, d(p)),$$

$$d(P-1) = 0,$$

$$\Delta(p) = d(p). \qquad (8)$$

Finally, having found $\Delta_{n,m}(p)$, the synchronised version of $\mathbf{x}_m$(t) to a base rate sequence $\mathbf{x}_n(t)$ might be calculated by

$$\mathbf{x}_{n,m}(p\delta t) = \mathbf{x}_m(p\delta t + \Delta_{n,m}(p)\delta t). \qquad (9)$$

Summarizing, the dense matching algorithm that synchronises two arbitrary human motion sequences $\mathbf{x}_n$ (t) and $\mathbf{x}_m$ (t) is as follows:

1. Prepare a 2D DSI matrix, and set initial cost values E_o using Eq. (5)
2. Find the optimal path trough the DSI using recurrence Eqs. (7), (8).
3. Synchronise $\mathbf{x}_m(t)$ to the rate of $\mathbf{x}_n(t)$ using Eq. (9).

Our algorithm assumes that a particular sequence is chosen to be a time scale pattern for all other sequences. In order to make an optimal choice of the sequence that will be used as the pattern for synchronizing the rest, a statistically proven rule according to some appropriate criterion is desirable. Towards this end, we define the synchronisation distance between a pair of sequences (*n*, *m*) as

$$D_{n,m} = \sum_{i=0}^{P}\left|\mathbf{x}_n(i\delta t) - \mathbf{x}_m(i\delta t + \Delta_{n,m}(i)\delta t)\right|^2$$

$$+\mu\sum_{i=0}^{P-1}\left|\Delta_{n,m}(i+1)\delta t - \Delta_{n,m}(i)\right|, \qquad (10)$$

Then, we can compute the global distance of the full synchronization of all the sequences m relative to the pattern sequence *n* as

$$D_n = \sum_{m\in A_k} D_{n,m}. \qquad (11)$$

We thus choose the synchronizing pattern sequence with minimal global distance D_n: in a statistical sense, such signal can be considered as a median value over all the performances that belong to the set of A_k or can be referred to as *median* sequence.

Finally, after running the algorithm on all our training performances $\tilde{\Psi}_i$ all the walking cycles have been synchronised and will be denoted as $\hat{\Psi}_i = \{\hat{\psi}_i^1, ..., \hat{\psi}_i^F\}$.

Figure 2(a) shows the first 4 dimensions of the input walking sequences represented in the PCA space without performing any synchronisation. Figure 2(b) shows the same situation after applying the synchronization algorithm proposed in this work. Notice that a common motion pattern arises after the synchronisation step.

3. LEARNING THE MOTION MODEL

Once all the walking sequences share the same time pattern, we learn an action specific model for walking which is accurate without loosing generality, and suitable for many applications such as gait analysis, gait recognition and tracking. Thus, we want to learn where the postures lie in the space used for representation, how do they change over time as the action goes by, and what characteristics do the different performances have in common which can be

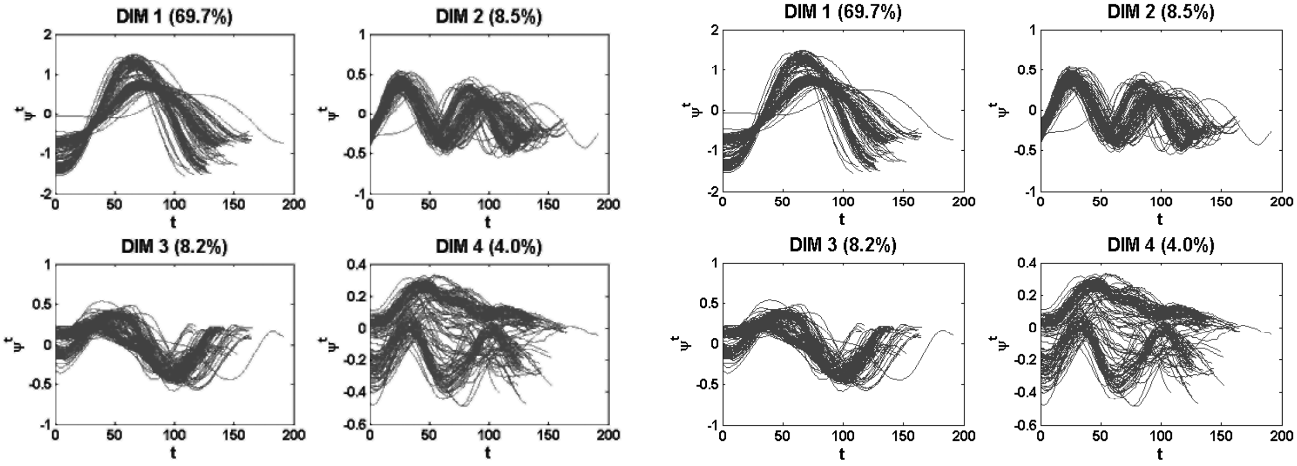

Figure 2: The first b = 4 dimensions within the PCA space before (a) and after (b) synchronization of the training set.

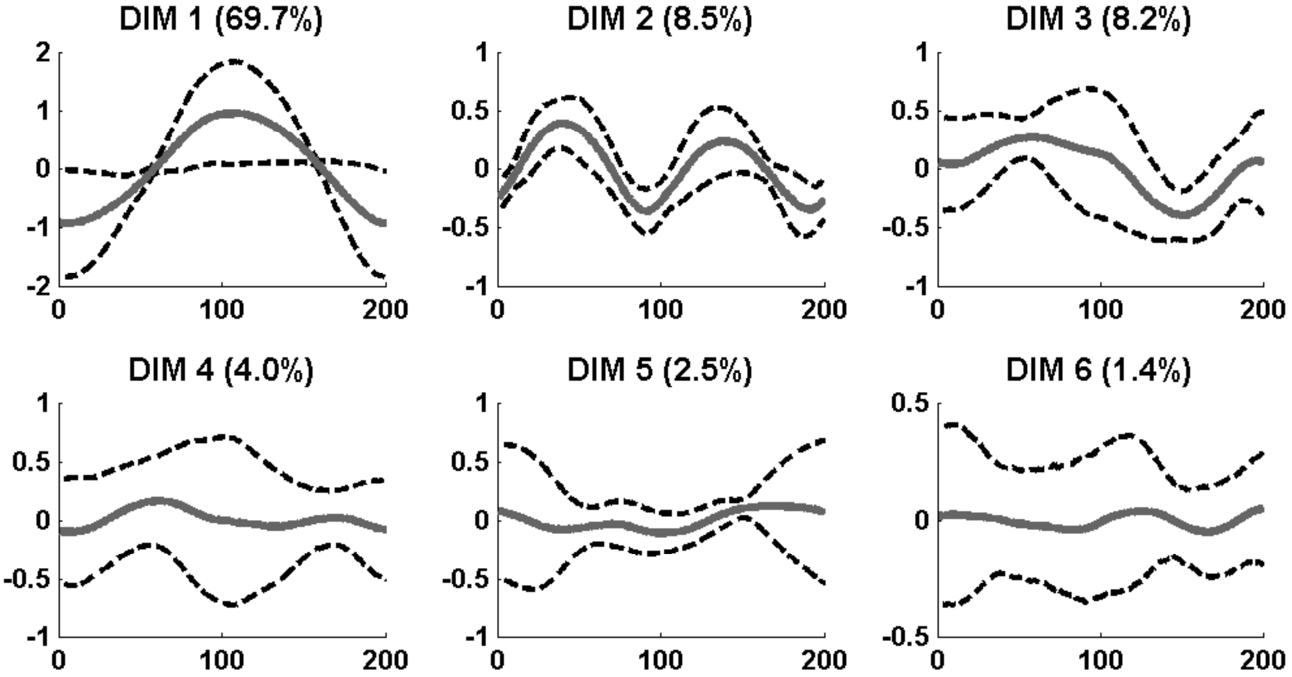

Figure 3: Learned mean performance $\bar{\Psi}$ and standard deviation σ_t for the walking action.

exploited for enabling the aforementioned tasks. In other words, we aim to characterize the *shape* of the synchronised version of the training set for the walking action in the PCA-like space. The process is as follows.

First, we extract from the training set $\hat{A}_k = \{\hat{\Psi}_1, \dots, \hat{\Psi}_{I_k}\}$ a mean representation of the action by computing the mean performance $\bar{\Psi}^{A_k} = \{\bar{\psi}^1, \dots, \bar{\psi}^F\}$, where each mean posture $\bar{\psi}^t$ is defined as

$$\bar{\psi}^t = \sum_{i=1}^{I_k} \frac{\hat{\psi}_i^{\,t}}{I_k}, \quad t = 1, \dots F. \tag{12}$$

I_k is the number of training performances for the action A_k, $\hat{\psi}_i^{\,t}$ corresponds to the *t*-th posture from the *i*-th training performance, and finally, *F* denotes the total number of postures of each synchronised performance.

Then, we want to quantify how much the training performances $\hat{\Psi}_i$ vary from the computed mean performance $\bar{\Psi}^{A_k}$ of Eq. (12). Therefore, for each time step *t*, we compute the standard deviation σ_t of all the postures $\hat{\psi}^t$ that share the same time stamp *t*, i.e.

$$\sigma_t = \sqrt{\frac{1}{I_k} \sum_{i=1}^{I_k} (\hat{\psi}_i^{\,t} - \bar{\psi}^t)}. \tag{13}$$

Figure 3 shows the learned mean performance $\bar{\Psi}$ (red solid line) and ±3 times the computed standard deviation σ_t (dashed black line) for the walking action. We used b = 6 dimensions for building the PCA space representation explaining the 93.3% of total variation of training data.

On the other hand, we are also interested in characterising the temporal evolution of the action. Therefore, we compute the main direction of the motion $\bar{\mathbf{v}}_t$ for each subsequence of *d* postures from the mean performance $\bar{\Psi}^{A_k}$, i.e.

$$\mathbf{v}_t = \frac{\sum_{j=t}^{t-d+1} \frac{(\bar{\psi}^j - \bar{\psi}^{j-1})}{\left\|(\bar{\psi}^j - \bar{\psi}^{j-1})\right\|}}{d}; \quad \bar{\mathbf{v}}_t = \frac{\mathbf{v}_t}{\|\mathbf{v}_t\|}, \tag{14}$$

where $\bar{\mathbf{v}}_t$ is a unitary vector representing the observed direction of motion averaged from the last *d* postures at a particular time step *t*. In Figure 2, the first 3 dimensions of the mean performance are plotted together with the direction vectors computed in Eq. (14). Each black arrow corresponds to the unitary vector $\bar{\mathbf{v}}_t$ computed at time *t*, scaled for visualization purposes. Hence, each vector encodes the mean observed motion's direction from time *t* – *d* to time *t*, where *d* stands for the length of the motion window considered. Additionally, selected postures from the mean performance have been sampled at times t = 1, 30, 55, 72, 100, 150 and 168 and overlaid in the graphic.

As a result, the action model Γ^{A_k} is defined by

$$\Gamma^{A_k} = \{\Omega^{A_k}, \bar{\Psi}^{A_k}, \sigma_t, \bar{\mathbf{v}}_t\}, \ t = 1..F, \tag{15}$$

where Ω^{A_k} is the PCA space definition for action A_k, $\bar{\Psi}^{A_k}$ is the mean performance, and $\sigma_t, \bar{\mathbf{v}}_t$ correspond to the computed standard deviation and mean direction of motion at each time step *t*, respectively.

Finally, to handle the cyclic nature of the waking action, we concatenate the last postures in each cycle with the initial

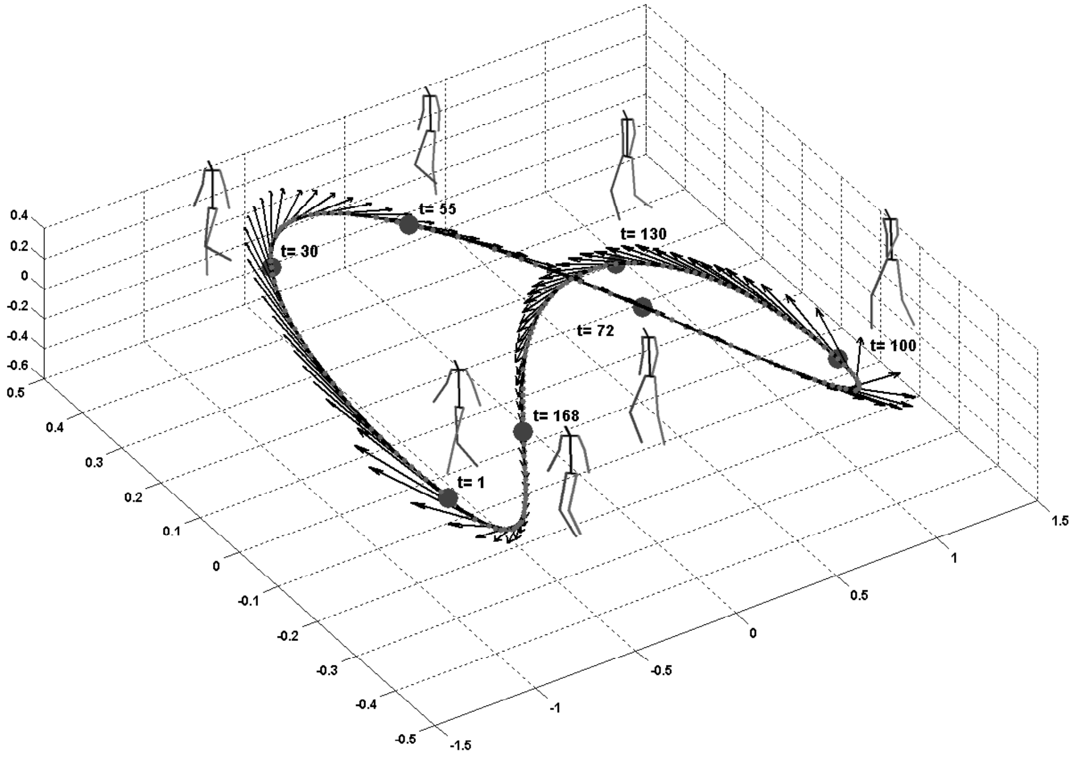

Figure 4: Sampled postures at different time steps, and learnt direction vectors $\overline{\mathbf{v}}_t$ from the mean performance for the walking action.

postures of the most close performance according to a Euclidean distance criterion within the PCA space. Additionally, the first and last (d/2) postures from the mean performance (where d is the length of the considered subsequences) are resampled using cubic spline interpolation in order to soft the transition between walking cycles. As a result, we are able to compute σ_t and $\overline{\mathbf{v}}_t$ for the last postures of a full walking cycle.

4. APPLICATIONS AND EXPERIMENTAL RESULTS

In this section we use the action specific model Γ^{A_k} in different application scenarios. A similar model was successfully used within a Bayesian 3D tracking framework in [18], and here its applicability for gait analysis and gait identification is presented and some experimental results are shown.

4.1 Gait Analysis

Given the synchronisation of different performances to the same time pattern, the angle variation between different performances can be quantified and analysed at any particular moment of the action.

We took three performances from different subjects, namely S2, S5 and S7, in order to analyse how different they perform on a walking cycle. The first performance corresponds to subject #2, 1st walking cycle from performance #1, and will be denoted as Ψ_{S2}. The second one, Ψ_{S5}, corresponds to the 1st cycle of the 1st performance of subject #5, and finally, the first cycle from performance #2 from subject #7 was compared and will be denoted as Ψ_{S7}. Figure 3 shows the evolution of absolute direction cosine angles from 4 limbs of the body model, namely the hip, the shoulders, the right upper arm and the right upper leg, respectively. It is worth saying that subjects S2 and S7 were males, while subject S5 corresponds to a female. By comparing the depicted angle variation values between the three walkers, one can observe several differences. In the first place, there are not substantial differences between hip's motion between the two male subjects. However, the hip's angles w. r. t. the X and Y axes from subject S5, corresponding to the elevation and rotation parallel to the floor according to Fig. 1, are a lot different from the other tested subjects. Thus, the swing movement of the hip is more emphasized in the female subject performance. Contrarily, when comparing the angle variation of the right arm and leg between male walkers and the female, few dissimilarities can be derived except that the female walker exhibits a less emphasized swing movement in the whole walking cycle. On the other hand, the shoulder movement is slightly different specially concerning the angle w.r.t. Y axis, corresponding to the elevation of the limb. In general, while subjects S2 and S7 show some differences, they share a very similar walking style compared to subject S5. These results confirm the conclusions stated in [10] about differences between walking styles between male and female actors.

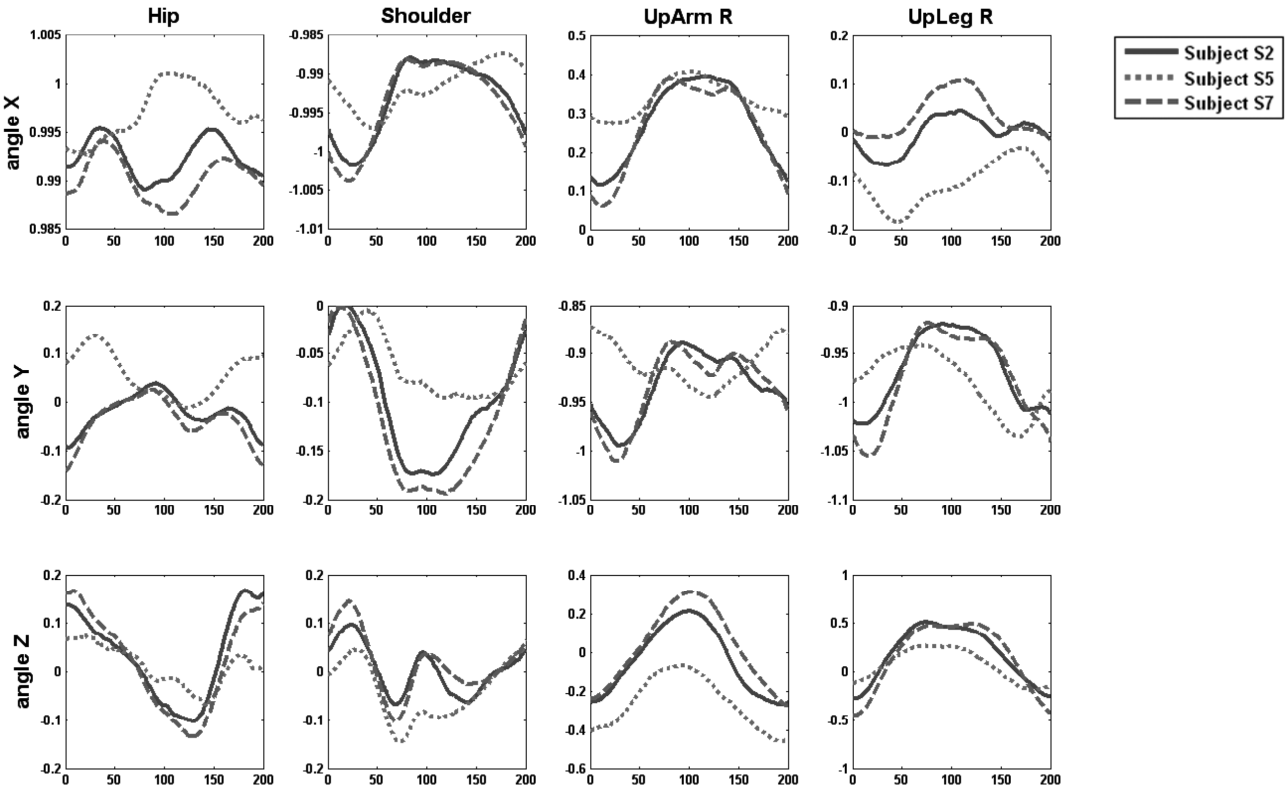

Figure 5: Absolute direction cosines computed for subjects S2, S5 and S7 for the hip, shoulder, right upper arm and upper leg limbs.

4.2 Gait Identification

To test the suitability of our action model for gait recognition applications, we aim to identify which subject is performing an action by analysing the observed motion from a particular test subject. Hence, we trained an specific model for each subject Si, where i identifies the subject according to Table 1. As a result, we learned 11 different action models, namely Γ^{S2}, Γ^{S5}, Γ^{S7}, Γ^{S8}, Γ^{S12}, Γ^{S16}, Γ^{S35}, Γ^{S38}, Γ^{S39}, Γ^{S43}, and Γ^{S49}. All subject-dependent action models share the same PCA space representation Ω^{A_k} so all the postures are represented in a common space. Notice that subject $S55$ was not considered in this experiment since we had only 1 walking cycle available from this subject.

Table 2
Confusion Matrix in Percentages for Full Cycle Recognition

	S2	*S5*	*S7*	*S8*	*S12*	*S16*	*S35*	*S38*	*S39*	*S43*	*S49*
S2	66.7	0	0	0	0	0	0	0	33.3	0	0
S5	0	100	0	0	0	0	0	0	0	0	0
S7	0	0	80	13.3	0	0	0	6.7	0	0	0
S8	0	0	0	77.8	0	0	0	0	22.2	0	0
S12	0	0	0	0	100	0	0	0	0	0	0
S16	0	0	0	0	0	40	33.4	13.3	0	13.3	0
S35	0	0	0	0	0	7.14	92.86	0	0	0	0
S38	0	0	0	0	0	25	0	75	0	0	0
S39	15.4	0	0	19.2	0	0	0	0	65.4	0	0
S43	0	0	0	0	0	0	0	0	0	100	0
S49	0	0	0	0	0	0	0	0	0	0	100

The approach is as follows: given an input motion sequence of length d, we compute the similarity S to all the subsequences of the same length from the 11 learned mean performances. Then, the subsequence which best matched a subject's mean performance according to our measure determines the identity of the subject.

Hence, the similarity measure used for gait identification between 2 subsequences of length d, namely $\Psi^a = \{\psi_a^1, \ldots, \psi_a^d\}$ and $\Psi^b = \{\psi_b^1, \ldots, \psi_b^d\}$ is defined as follows:

$$S(\Psi^a, \Psi^b) = \exp\left(D_M(\Psi^a, \Psi^b)\right)\left[\frac{(\bar{\mathbf{v}}_a \bullet \bar{\mathbf{v}}_b) + 1}{2}\right]^{\alpha}, \quad (16)$$

where $\bullet$ stands for the dot product between vectors $\bar{\mathbf{v}}_a$ and $\bar{\mathbf{v}}_b$ corresponding to the average direction of motion computed following Eq. (14). D_M is the sum of the Mahalanobis distance within the PCA space Ω^{A_K} between each posture ψ_a^j and ψ_b^j from the subsequences, $j = 1..d$. Our similarity measure is decomposed in two terms. The exp term accounts for the similarity between postures within the PCA space, while the dot product term expresses similarity between directions of motion across time, regardless the body postures exhibited. Finally the exponent α controls the importance given to the latter term for computing the final similarity. In other words, high values for α will provide high similarity values to sequences following the same direction of motion, while low values will take more into account the position of their postures within the PCA space. Therefore, this similarity metric defines a trade off from one hand between sequences that exhibit similar motion directions, and from the other hand sequences with close postures within the PCA space according to their Mahalanobis distances. As a result, only close sequences which follow the same direction will get high scores, while sequences that do not match in motion direction or position are given low similarity scores.

In our first experiment, we took a full walking cycle of each individual for testing the identification approach. We chose b = 10 dimensions for the PCA space representation of human postures. Subsequently, the similarity of the full test cycle to each specific action model's mean performance was computed according to Eq. (16). The tested walking cycle was removed from the training set in the learning stage. Then, this experiment was repeated for each cycle of the database, resulting in a total of 126 identification tests. The confusion matrix explaining the recognition performance can be seen in Table 2. Several miss classifications occur due to different reasons. On the one hand, results obtained for subjects S2, S38, S43 and S49 are not statistically confident since less than 5 cycles are provided in the training database. On the other hand, looking at the miss classification obtained between subjects S16 and S35 we discovered that indeed they correspond to the same actor who performed the recording. Despite of the fact that in the specification of the CMU database, these subjects are defined as different, the authors of this paper recognised that the same person performed the recordings for both subjects datasets by subsequently checking the video recordings from those sessions.

Afterward, we ran another experiment taking d = 10 as the length of the subsequences considered for performing gait identification. All the testing walking cycles have a total length of F = 200 postures. Then, for each subject, we selected a random test walking cycle from the database. Thus, each tested cycle is composed of a total of (*F*–*d*+1) overlapping motion subsequences. Hence, we ran the gait identification experiment for each possible motion subsequence of each tested subject and computed its confusion matrix. The same experiment was repeated a total of 10 times. The average of the obtained confusion matrices can be seen in Table 3. One can observe that the performance obtained is comparable with the full cycle experiment, but using only 1/20 of a walking cycle. Although some miss classifications occur between subjects that did not appear in the previous experiment, in some cases the performance is even better. This can be explained because of the better statistical robustness of this experiment, since we performed an identification test for each of the (*F* – *d* + 1) = (200–10 + 1) = 191 subsequences belonging to a full tested cycle. This results in a total of 191 ∗ *nSubjects* ∗ *timesRepeated* = 191 ∗ 11 ∗ 10 identification tests as opposed to the 126 identification tests from the previous experiment. The results are very encouraging, since they show that we are able to recognise which subject is performing an action by observing only a very reduced motion portion from it.

Table 3
Confusion Matrix in Percentages for Subsequences of d = 10 postures

	S2	*S5*	*S7*	*S8*	*S12*	*S16*	*S35*	*S38*	*S39*	*S43*	*S49*
S2	91.28	0	0.05	0.47	0	0	0	0	7.04	1.16	0
S5	0	97.21	0	0	1.92	0	0	0.35	0	0	0.52
S7	0.35	1.80	89.88	0.12	0	0	0.06	2.50	0.12	1.92	3.25
S8	0.47	0	0.29	91.86	0	0	0	0.12	7.26	0	0
S12	0	0	0	0	99.83	0	0	0	0	0	0.17
S16	0	0	0	0	3.20	64.17	19.37	6.34	0	2.04	4.88
S35	0	1.34	0.06	0	4.19	19.09	69.28	3.84	0	1.10	1.10
S38	0	1.28	0	0	1.86	6.17	2.91	76.85	0	0	10.93
S39	6.51	0	0.06	3.49	0	0	0	0	88.66	1.28	0
S43	0	0	0	0	0	0	0	0	0	100	0
S49	0	0	0	0	0.17	0	0	0	0	0	99.83

5. CONCLUSIONS AND FUTURE WORK

We have presented an action-specific model suitable for gait analysis, gait identification and tracking applications. The model is tested for the walking action, and is automatically learnt from the public CMU motion capture database. A methodology for synchronising the original human motion input sequences is detailed, which uses Dynamic Programming techniques. As a result, we learnt the parameters of our action model which characterise the pose variability observed within a set of walking performances used for training.

The resulting action model consists of a representative manifold for the action, namely the mean performance, the standard deviation from the mean performance, and the mean observed direction vectors from each motion subsequence of a given length. The action model can be used to classify which postures belong to the action or not. Moreover, the trade off between accuracy and generality of the model can be tuned using more or less dimensions for building the PCA space representation of human postures. Hence, using this coarse-to-fine representation, the main modes of variation correspond to meaningful natural motion modes. Thus, for example, we found that the main modes of variation for the walking action obtained from PCA, explain the combined motion of both the legs and the arms, while in the bending action they mainly correspond to the motion of the torso.

Subsequently, the learnt action model was used in combination with the synchronisation algorithm for gait analysis applications. This enabled us to compare and quantify the difference between different performances of the same action. Furthermore, the computed mean observed direction vectors for a performance allow the formulation of a similarity measure *S* between motion subsequences of the same length. The measure combines similarity in the direction of the performed motion and distance within the PCA space. Its usefulness for gait identification has been presented, and experimental results point out that we are able to recognise the 11 tested subjects using a very reduced number of motion samples.

Future research lines rely on obtaining the joint positions directly from image sequences. Previously, the action model has been successfully used in a probabilistic tracking framework for estimating the parameters of our 3D model from a sequence of 2D images. In [19], the action model improved the efficiency of the tracking algorithm by constraining the space of possible solutions only to the most feasible postures while performing a particular action, thus avoiding estimating postures which are not likely to occur during an action. However, we need to develop robust image-based likelihood measures which evaluate the predictions from our action model according to the measurements obtained from images. Work based on extracting the image edges and the silhouette from the tracked subject is currently in progress. Hence, the pursued objective is to learn a piece-wise linear model which evaluates the fitness of segmented edges and silhouettes to the 2D projection of the stick figure from our human body model. Methods for estimating the 6DOF of the human body within the scene, namely 3D translation and orientation, also need to be improved. Lastly, a method for automatically initialising the tracker is also being studied, since the Bayesian inference framework used to face the tracking problem does not provide any clue for the initial state of the tracked object.

Finally, even using tracking approaches, recovering all joints' positions from images accurately is specially difficult in the presence of occlusions and when all joints are not directly observable due to 2D projection effects. Therefore, we aim to explore and extend the gait analysis and identification facilities of the action model presented here in case that not all the joints' positions are available or correctly estimated by the tracking algorithm.

ACKNOWLEDGMENTS

This work has been supported by the Catalan Research Agency (AGAUR), by EC grants IST-027110 for the HERMES project and IST-045547 for the VIDI-Video project, and by the Spanish MEC under projects TIN2006-14606 and DPI-2004-5414. Jordi Gonzàlez also acknowledges the support of a Juan de la Cierva Postdoctoral fellowship from the Spanish MEC. The database used in this project was obtained from mocap.cs.cmu.edu which was created with funding from NSF EIA-0196217.

REFERENCES

[1] J. K. Aggarwal and Q. Cai. Human motion analysis: A review. *Computer Vision and Image Understanding*, 73(3): 428440, 1999.

[2] A. F. Bobick and J. W. Davis. The recognition of human movement using temporal templates. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(3): 257267, 2001.

[3] M. Z. Brown, D. Burschka, and G. D. Hager. Advances in computational stereo. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(8): 9931008, 2003.

[4] R. Cutler and L. S. Davis. Robust real-time periodic motion detection, analysis, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8): 781796, 2000.

[5] Q. Delamarre and O. Faugeras. 3D Articulated Models and Multiview Tracking with Physical Forces. *Computer Vision and Image Understanding*, 81(3): 328357, 2001.

[6] J. Deutscher and I. Reid. Articulated Body Motion Capture by Stochastic Search. *International Journal of Computer Vision*, 61(2): 185205, 2005.

[7] A. A. Efros, A. C. Berg, G. Mori, and J. Malik. Recognizing action at a distance. In *Ninth IEEE International Conference on Computer Vision*, pages 726733, Nice, France, 2003.

[8] H. L. Eng, K. A. Toh, A. H. Kam, J. Wang, and W. Y. Yau. An automatic drowning detection surveillance system for challenging outdoor pool environments. In *Ninth IEEE International Conference on Computer Vision*, pages 532539, Nice, France, 2003.

[9] J. Gonzàlez, J. Varona, X. Roca, and J.J. Villanueva. Analysis of human walking based on aSpaces. *In Conference on Articulated Motion and Deformable Objects* (AMDO'04), Palma de Mallorca, Spain, September 2004.

[10] J. Gonzàlez, J. Varona, F. X. Roca, and J. J. Villanueva. A comparison framework for walking performances using aSpaces. *Electronic Letters on Computer Vision and Image Analysis*, 5(3):105-116, August 2005.

[11] R. D. Green, L. Guan, and J. A. Burne. Video analysis of gait for diagnosing movement disorders. *Journal of Electronic Imaging*, 9: 16, 2000.

[12] G. Johansson. Visual motion perception. *Scientific American*, 232(6): 7688, 1975.

[13] A. Kale, A. Sundaresan, A. N. Rajagopalan, N. P. Cuntoor, A. K. Roy-Chowdhury, V. Kruger, and R. Chellappa. Identification of humans using gait. *IEEE Transactions on Image Processing*, 13(9): 11631173, 2004.

[14] H. J. Lee and Z. Chen. Determination of 3d human body posture from a single view. *Computer Vision Graphics*, 30: 148168, 1985.

[15] L. Lee and W. E. L. Grimson. Gait analysis for recognition and classification. In *Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 148-155, Washington D.C., USA, 2002.

[16] Y. Li, T. Wang, and H.Y. Shum. Motion texture: a two-level statistical model for character motion synthesis. In SIGGRAPH, pages 465472, San Antonio, Texas USA, 2002.

[17] J. Little and J. Boyd. Recognizing people by their gait: the shape of motion. *Journal of Computer Vision Research*, 1(2): 132, 1998.

[18] I. Rius, J. Varona, J. Gonzalez, and J. J. Villanueva. Action Spaces for Efficient Bayesian Tracking of Human Motion. In 18th International Conference on Pattern Recognition (ICPR'06), volume 01, pages 472475, Hong Kong, 2006.

[19] I. Rius, J. Varona, F.X. Roca, and J. Gonzàlez. Posture constraints for bayesian human motion tracking. In *IV Conference on Articulated Motion and Deformable Objects* (*AMDO '06*), pages 414-423, Mallorca (Spain), 2006.

[20] M. Roh, B. Christmas, J. Kittler, and S. Lee. Robust Player Gesture Spotting and Recognition in Low-Resolution Sports Video. *European Conference on Computer Vision, Graz, Austria, May*, pages 713, 2006.

[21] D. Scharstein and R. Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision*, 47(1): 742, 2002.

[22] H. Sidenbladh, M. J. Black, and L. Sigal. Implicit Probabilistic Models of Human Motion for Synthesis and Tracking. In *European Conference on Computer Vision*, volume 1, pages 784800, Copenhagen, Denmark, 2002.

[23] L. Sigal and M. J. Black. Measure Locally, Reason Globally: Occlusion-sensitive Articulated Pose Estimation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 20412048, New York, USA, 2006.

[24] L. M. Tanco and A. Hilton. Realistic synthesis of novel human movements from a database of motion capture examples. In *Proceedings of the Workshop on Human Motion (HUMO'00)*, pages 137-142, Austin, Texas USA, 2000.

[25] A. Hilton T.B. Moeslund and V. Kruger. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 104(2-3): 90126, November 2006.

[26] S. Wachter and H. H. Nagel. Tracking persons in monocular image sequences. *Computer Vision and Image Understanding*, 74(3): 174192, June 1999.

[27] L. Wang, W. Hu, and T. Tan. Recent developments in human motion analysis. *Pattern Recognition*, 36(3): 585601, 2003.

[28] V. M. Zatsiorsky. *Kinematics of Human Motion*, chapter 1, pages 2262. Human Kinetics, 1998.