Computer Vision and Image Understanding 115 (2011) 1363–1374

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Human motion capture using scalable body models

Cristian Canton-Ferrer *, Josep R. Casas, Montse Pardàs

Technical University of Catalonia, Barcelona, Spain

ARTICLE INFO

Article history:
Received 25 January 2010
Accepted 6 June 2011
Available online 17 June 2011

Keywords:
Motion capture
Monte Carlo techniques
Particle filtering
Scalability
Human motion capture
Monte Carlo filtering
Scalable analysis
Robust analysis

ABSTRACT

This paper presents a general analysis framework towards exploiting the underlying hierarchical and scalable structure of an articulated object for pose estimation and tracking. Scalable human body models are introduced as an ordered set of articulated models fulfilling an inclusive hierarchy. The concept of annealing is applied to derive a generic particle filtering scheme able to perform a sequential filtering over the set of models contained in the scalable human body model. Two annealing loops are employed, the standard likelihood annealing and the newly introduced *structural annealing*, leading to a robust, progressive and efficient analysis of the input data. The validity of this scheme is tested by performing markerless human motion capture in a multi-camera environment employing the standard HumanEva annotated datasets. Finally, quantitative results are presented and compared with other existing HMC techniques.

1. Introduction

Automatic human motion capture (HMC) has been studied extensively [1,2], basically fostered by the number of potential applications and its inherent complexity. This research area contains a number of ill-posed problems such as inferring the pose and motion of a highly articulated and self-occluding 3D object from a set of images. Applications that benefit from the obtained information are, for instance, human computer interfaces [3,4], unusual behavior detection in security applications [5] or teleconferencing [6].

Although some applications only need a coarse human body model [7], many others require to work with an articulated structure. Recovering the pose of an articulated structure such as the human body involves estimating highly dimensional and multimodal statistic distributions. In this field, some contributions employ linear techniques such as Kalman filtering [8,9] or the EM algorithm [10] although being prone to loose track. Monte Carlo based techniques [11] have been thoroughly applied due to its ability to cope with multimodal distributions with an affordable computational complexity. Particle filtering [12] has been the seminal idea to develop specific systems aiming at recovering human body pose such as the annealed particle filter [13], the hierarchical sampling [14] or the partitioned sampling [15] among others. Another common approach to HMC is based on learning pose appearances from a large annotated data set and then classifying the new input data accordingly [16,17]. In other words, HMC is no longer posed as a tracking problem. However, despite the high accuracy produced by these techniques, they are constrained to track beforehand learnt motion patterns, hence their applicability to analyze unknown gestures or poses is limited.

From a data processing point of view, HMC capture algorithms can be divided into two groups: marker-based and markerless. The first employs a set of optically distinguishable markers placed in some landmarks of the body and are able to produce highly accurate results, mostly employed by the cinema [18] and medical [19] industries. However, the hardware required to acquire these markers constrains the analysis to a setup scenario and the marker placement might be intrusive and/or uncomfortable for the user [20]. On the other hand, markerless approaches grant more freedom of movement and a more natural HMC scenario. These methods are based on obtaining a number of features derived from the video input such as edges and silhouettes [13,14] and estimating the pose afterwards. In both marker-based and markerless approaches, multi-camera acquisition systems are commonly used in order to obtain some different points of view of the same scene in order to cope with occlusions and perspective issues [21,22].

A common approach is model-based motion capture, where a human body model (HBM) is selected beforehand and then fitted to the input data. This paper presents a general analysis framework that exploits the underlying hierarchical structure of an articulated object by using a scalable human body model mimicking the bottom-up and multi-resolution concepts into the HMC field. A sequential Monte Carlo fitting is performed over a set of human body models with increasing level of detail by applying the newly

* Corresponding author. Fax: +44 1865 240527.
E-mail address: cristian.canton@gmail.com (C. Canton-Ferrer).

1077-3142/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2011.06.001

introduced concept of *structural annealing* denoted as the Hierarchical Structure based Annealed Particle Filter. This technique allows an analysis of data robust to noisy inputs with a high efficiency. This paper extends our preliminary work [23] where these concepts were sketched; here, the necessary theoretical framework is provided as well as a detailed description of the several processing modules involved. Particularly, a noteworthy contribution is the adaptive sorting and resampling technique employed to merge state spaces with different dimension and number of sampling particles. Finally, additional experimental tests have been conducted employing HumanEva I and II datasets [24] to corroborate the effectiveness of the presented technique.

The presented algorithm is intended for any HMC system, regardless of the input data (marker-based or markerless). We test its validity by using it in a multi-camera markerless HMC scenario using a 3D voxelized reconstruction of the scene as input [21]. The standard HumanEva dataset [24] is employed to both quantitatively assess the accuracy of the proposed technique and compare its performance with other existing systems.

This paper first presents the concept of scalable human body models in Section 2 and then, in Section 3, the Hierarchical Structure based Annealed Particle Filter is generally described. Details on its implementation are given in Section 4 and a markerless HMC multi-camera application is presented in Section 5. Results and comparisons are presented in Section 6 and some conclusions are drawn in Section 7.

2. Scalable human body models

Let us define a human body model (HBM) as the set $\mathcal{H}$ formed by a root segment (torso) denoted as $\mathcal{T}$ and a set of $N_{\mathcal{L}}$ open kinematic chains modeling the head, arms and legs. Each limb is formed by a variable number of elements (links in this kinematic chain) denoted as $\mathcal{P}$. Hence,

$$\mathcal{H} = \{\mathcal{T}, \mathcal{P}_{i,j}, J_{i,j}\}, \quad 1 \leqslant i \leqslant N_{\mathcal{L}}, \quad 1 \leqslant j \leqslant N_{\mathcal{P}(i)}, \tag{1}$$

where $N_{\mathcal{P}(i)}$ stands for the number of parts in the ith limb. The torso, limbs and their sub-parts are connected to one another by means of joints, $J_{i,j}$ mathematically modeled using exponential maps [25]. In order to constrain the possible poses that a HBM may adopt, we define the number of degrees of freedom and the legal angular ranges at each joint.

The concept of scalability has been widely adopted in many topics within the image processing community. For instance, multiresolution analysis has been largely employed in image coding to exploit similarity of an image across scale changes [26] or in motion estimation [27]. Despite some scalability contributions have been presented for HMC, they are tailored to a specific application or in a very ad hoc fashion.

Exploiting the scalability potential of a HBM can be addressed from two perspectives: model-based and algorithm-based. In the first, the employed HBM is modified along the analysis of a given input data to progressively improve the fitting process. In [28], several HBMs are used in a scheme to heuristically search body parts in 2D. In an attempt to simplify the inverse kinematics problem, [29] proposed a 2-layer kinematic model where a coarse and unconstrained layer is first fitted to the tracked body parts and then, a second layer containing some kinematic constraints, is adjusted onto the data from the first layer.

Scalability within the fitting process can be addressed from an algorithmic point of view, using the same HBM along all analysis process. These algorithms take advantage of the topology of the HBM towards performing a progressive fitting of it. In [14], the fitting process is performed progressively through the limbs of the body in a layered scheme as similarly done in [30]. However, due to the assumption of a certain degree of statistical independence among state space variables required by such scalable methods, inter-relations among limbs are not usually considered. In this paper, a framework for a joint model and algorithm-based scalable analysis of the input data is provided.

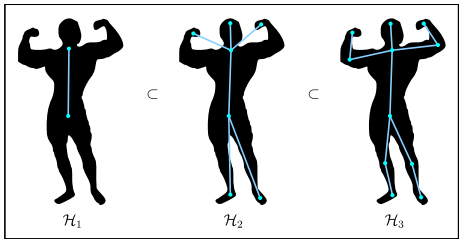


Fig. 1. Example of inclusive scalable human body model in terms of model detail. Most detailed model $\mathcal{H}_{i+1}$ is understood as a refinement of $\mathcal{H}_i$ since it adds further elements of the body model to the previous level.

2.1. Definition

A scalable human body model (SHBM) can be defined as a set of HBMs:

$$\mathcal{M} = \{\mathcal{H}_1, \mathcal{H}_2, \cdots, \mathcal{H}_M\}. \tag{2}$$

To achieve scalability, a certain hierarchy among the elements of $\mathcal{M}$ must be defined. Among all possible criteria, the inclusion one has been selected in this work. Under this rule, the relationship among different $\mathcal{H}_i$ is defined as:

$$\mathcal{H}_i \subset \mathcal{H}_j, \quad i < j, \tag{3}$$

where the inclusion operation is understood in terms of the scalability criterion. This criterion is a design parameter and can be defined, for instance, as the number of elements in the body parameter, the information encoded in every model, etc. An example of detail scalability of the human body is depicted in Fig. 1. The inclusion condition establishes an ordering among the $\mathcal{H}_i$.

3. Hierarchical Structure based Annealed Particle Filter

Assuming that a SHBM $\mathcal{M}$ with a given hierarchy has been stated, we define a sequential fitting process over the several HBMs $\mathcal{H}_i \in \mathcal{M}$. In order to carry out this task, we propose an extension of the annealing in [13] where particles are placed around the peaks of the likelihood function by means of a recursive search over a set of decreasingly smoothed versions of this function. The main concept is to use the set of progressively refined HBMs contained in $\mathcal{M}$ instead of a set of smoothed versions of the likelihood function. This process mimics the annealing strategy of the coarse-to-fine analysis of the likelihood function thus leading to what we denote as a *structural annealing* process.

3.1. Filter description

Let us have a SHBM $\mathcal{M}$ whose elements $\mathcal{H}_i$ fulfill the inclusive hierarchy criteria and denote the state space associated to HBM $\mathcal{H}_i$ as $\mathcal{X}_{\mathcal{H}_i} = [\theta_1 \cdots \theta_{K_{\mathcal{H}_i}}] \subseteq \mathbb{R}^{K_{\mathcal{H}_i}}$, where $K_{\mathcal{H}_i}$ is the associated dimension of the model $\mathcal{H}_i$. If the SHBM is properly defined, its elements will fulfill that:

$$\mathcal{X}_{\mathcal{H}_1} \subset \mathcal{X}_{\mathcal{H}_2} \subset \cdots \subset \mathcal{X}_{\mathcal{H}_M}, \tag{4}$$

$$K_{\mathcal{H}_1} < K_{\mathcal{H}_2} < \cdots < K_{\mathcal{H}_M}. \tag{5}$$

These conditions state the relation between two HMBs $\mathcal{H}_i$ and $\mathcal{H}_j, i < j$, as $\mathcal{H}_i$ being a subset of $\mathcal{H}_j$ with strict lower dimension. Concretely, SHBMs will be designed following the rule:

$$\mathcal{X}_{\mathcal{H}_i} = \left[f(\mathcal{X}_{\mathcal{H}_{i-1}}) \theta_{K_{\mathcal{H}_{i-1}}+1} \cdots \theta_{K_{\mathcal{H}_i}} \right]. \quad (6)$$

The state space $\mathcal{X}_{\mathcal{H}_i}$ is designed to contain information directly related to the variables from the previous model $\mathcal{H}_{i-1}$ and, recursively, from all previous models. Function $f(\cdot)$ is intended to perform a mapping among variables between two consecutive state spaces and typically involves a linear (or trivial) operation. If we define the incremental state space as $\mathcal{X}^{\Delta}_{\mathcal{H}_i} = \{\theta_m \in \mathcal{X}_{\mathcal{H}_i} | \theta_m \notin \mathcal{X}_{\mathcal{H}_{i-1}}\}$, Eq. (6) can be rewritten as:

$$\mathcal{X}_{\mathcal{H}_i} = \left[f(\mathcal{X}_{\mathcal{H}_{i-1}}) \mathcal{X}^{\Delta}_{\mathcal{H}_i} \right], \quad (7)$$

where the associated dimension of $\mathcal{X}^{\Delta}_{\mathcal{H}_i}$ is:

$$\dim\left(\mathcal{X}^{\Delta}_{\mathcal{H}_i} \right) \equiv K^{\Delta}_{\mathcal{H}_i}. \quad (8)$$

Let us have a particle filter associated to each state space $\mathcal{X}_{\mathcal{H}_i}$, with its associated particle set $\{(\mathbf{y}^j_t, \pi^j_t)\}^{\mathcal{H}_i}$, containing $N_{\mathcal{H}_i}$ particles. The overall operation of the proposed Hierarchical Structure based Annealed Particle Filter (HS-APF) scheme is to filter the initial distribution obtained at instant $t-1$ associated to the simplest HBM $\mathcal{H}_1$ and then combine the resulting particle set with the initial particle set of the following model, $\mathcal{H}_2$. In this way, information from already filtered variables of $\mathcal{H}_1$ improves the initial particle set associated to $\mathcal{H}_2$. This process is performed for all the models in the SHBM until reaching the last one. Information contained in the particle set of the last model is back-propagated to the models with a lower hierarchy rank thus refining their associated particle sets and closing the information filtering loop. The scheme of the proposed technique is depicted in Fig. 2 for $M = 3$.

When a new measurement $\mathbf{z}_t$ is available, a structural annealing iteration is performed. The HS-APF can be summarized as follows:

1. Starting from model $\mathcal{H}_1$, its associated particle set $\{(\mathbf{y}^j_{t-1}, \pi^j_{t-1})\}^{\mathcal{H}_1}$ is resampled with replacement. Then the filtered state $\{(\tilde{\mathbf{y}}^j_t, \tilde{\pi}^j_t)\}^{\mathcal{H}_1}$ is constructed by applying a propagation model $P(\mathbf{y}^j_t, \Sigma_{\mathcal{H}_1})$ and the likelihood function $w_{\mathcal{H}_1}(\tilde{\mathbf{y}}^j_t, \mathbf{z}_t)$ to every resampled particle as:
$$\tilde{\mathbf{y}}^j_t = P(\mathbf{y}^j_{t-1}, \Sigma_{\mathcal{H}_1}), \quad (9)$$
$$\tilde{\pi}^j_t = w_{\mathcal{H}_1}(\tilde{\mathbf{y}}^j_t, \mathbf{z}_t). \quad (10)$$
Weights are normalized such that $\sum_j \tilde{\pi}^j_t = 1$. At this point, the output estimation of this model $\mathcal{Y}^{\mathcal{H}_1}_t$ can be computed by applying
$$\mathcal{Y}^{\mathcal{H}_1}_t = \sum_{j=1}^{N_{\mathcal{H}_1}} \tilde{\pi}^j_t \tilde{\mathbf{y}}^j_t. \quad (11)$$
2. For the following HBMs, $i > 1$, the filtered particle set of the previous model in the hierarchy, $\{\left(\tilde{\mathbf{y}}^j_t, \tilde{\pi}^j_t\right)\}^{\mathcal{H}_{i-1}}$, is combined through the operator G_{forward} with the particle set associated to model $\mathcal{H}_i$, $\{(\mathbf{y}^j_{t-1}, \pi^j_{t-1})\}^{\mathcal{H}_i}$. State space variables associated to $\mathcal{H}_i$ contain information from model $\mathcal{H}_{i-1}$ due to the imposed inclusive hierarchy relation. Since these common variables have been already filtered, the updated information can be transferred to the particles of model $\mathcal{H}_i$ in order to generate an improved initial particle set (a further review of G_{forward} is presented in Section 4). Then, the filtered state $\{(\tilde{\mathbf{y}}^j_t, \tilde{\pi})^j_t\}^{\mathcal{H}_i}$ is constructed applying Eqs. (9) and (10). At this point, the output estimation of this model, $\mathcal{Y}^{\mathcal{H}_i}_t$, can be computed.
3. Once reaching the highest hierarchy level, that is the most detailed HBM, the information contained in the particle set $\{(\tilde{\mathbf{y}}^j_t, \tilde{\pi}^j_t)\}^{\mathcal{H}_M}$ is back-propagated to the other models in the hierarchy by means of the operator $G_{\text{backwards}}$. This operator adaptively replaces the state of the particles in lower hierarchy models by the particles associated to model $\mathcal{H}_M$. In this way, the particle sets of every model are refined thus closing the filtering loop.

An example of the execution of this scheme is depicted in Fig. 3.

In order to refine the particle set $\{(\tilde{\mathbf{y}}^j_t, \tilde{\pi}^j_t)\}^{\mathcal{H}_i}$ that will be transferred to layer $i+1$, we added a standard annealing loop represented as a dashed line in the overall scheme. This will concentrate the particles around the main modes of the likelihood function at layer i before delivering them to layer $i+1$. It must be noted that an accurate likelihood estimation at a lower layer will benefit the subsequent estimation layers. For further discussion, let us denote as $L_{\mathcal{H}_i}$ the number of annealing layers associated to the filtering thread associated to model $\mathcal{H}_i$. The presented configuration can be seen as a filtering scheme with a double annealing loop: one in the model complexity, benefiting from the hierarchical properties of the SHBM, and one in the filtering

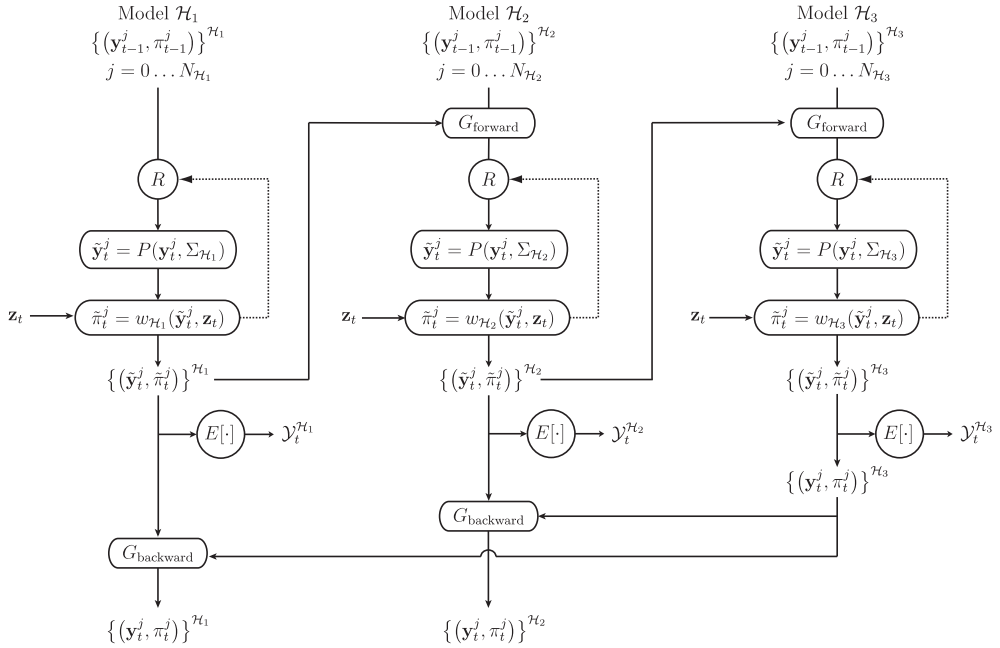


Fig. 2. Hierarchical Structure based Annealed Particle Filter (HS-APF) scheme for $M = 3$ elements in the SHBM fulfilling the inclusive criteria.

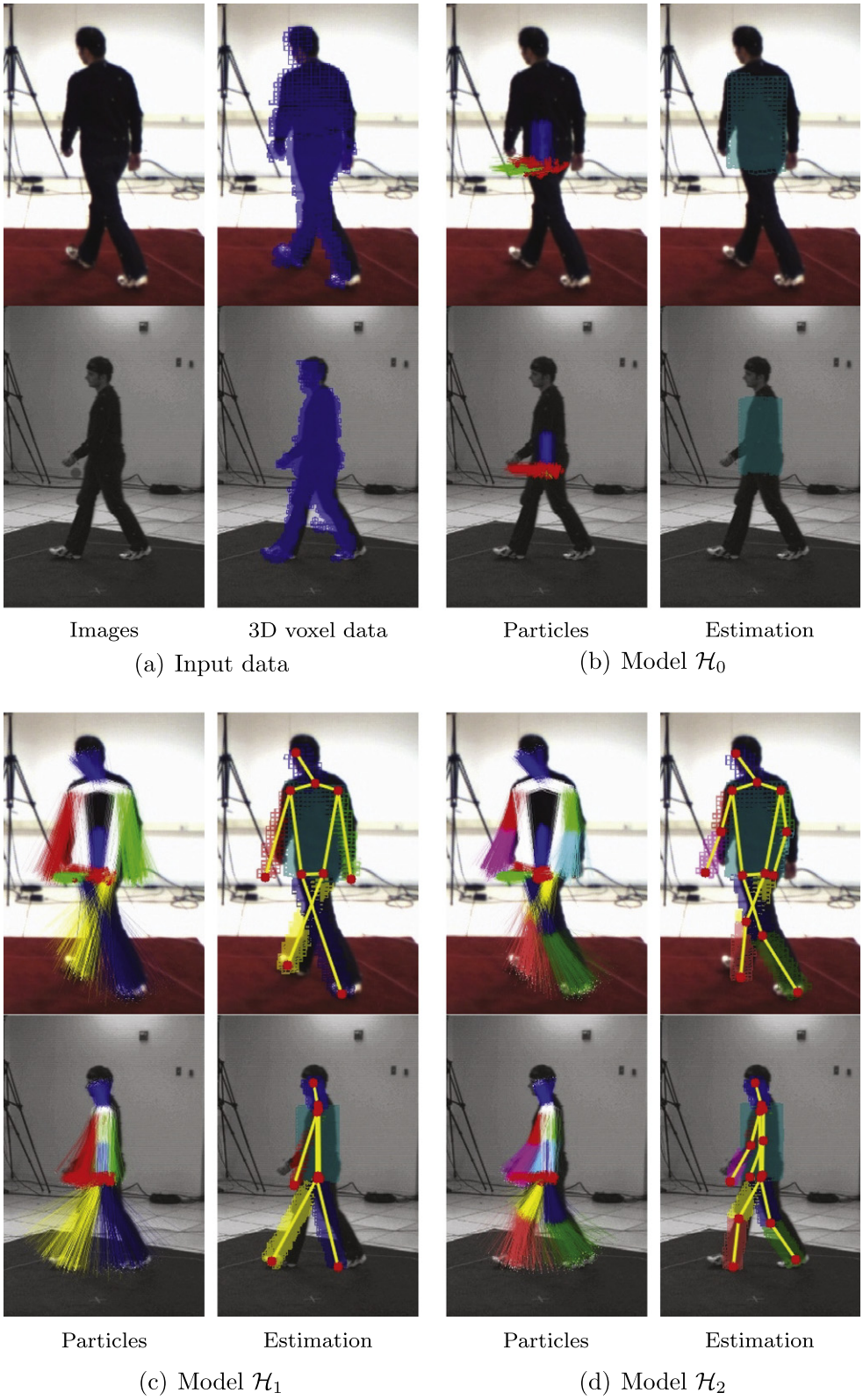


Fig. 3. Example of the HS-APF algorithm operation from two camera views. In (a), the input data (two views): the original images and, overlayed in blue, the projection of the 3D reconstruction obtained with voxels of size $s_v = 2$ cm. In (b–d), the successive filtering threads associated to every HBM $\mathcal{H}_i$. Particles $\{(\tilde{\mathbf{y}}_t^j, \tilde{\pi}_t^j)\}^{\mathcal{H}_M}$ and estimation $\mathcal{Y}_t^{\mathcal{H}_i}$ for every HBM are displayed. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

branch associated to each model, benefiting from the annealing effects of particles concentration around the main peaks of the likelihood function. Let us denote as α_S the variance decrease rate among two consecutive HBMs and $\alpha_{\mathcal{H}_i}$ as the variance decrease rate in the inner likelihood annealing loop associated to a given HBM $\mathcal{H}_i$.

4. Filter implementation

Once the general overview of the HS-APF is presented, some further details on its modules are provided. However, since this system is intended for any HMC, the elements related to the data fed to the system, that is the input data, $\mathbf{z}_t$, and the likelihood function, $w_{\mathcal{H}_i}(\tilde{\mathbf{y}}_t^j, \mathbf{z}_t)$, are described in the specific markerless implementation presented in Section 5.

4.1. SHBM model

Two proposals of inclusive SHBM are presented, depicted in Fig. 4. In model $\mathcal{M}_1$, the overall limbs orientation is first estimated and then the remaining joints whereas, in model $\mathcal{M}_2$, the limbs orientation is progressively estimated. In motions where limbs are mostly straight, i.e. walking or running, model $\mathcal{M}_1$ is more adequated, capturing the overall orientation of each limb (hips and shoulders joints) and then refining the estimation (knees and elbows). Other types of motion like gesturing are better captured using model $\mathcal{M}_2$.

The associated state spaces to both models are identical:

$$\mathcal{X}_{\mathcal{H}_1} = [\mathbf{xR}], \tag{12}$$

$$\mathcal{X}_{\mathcal{H}_2} = \left[\mathcal{X}_{\mathcal{H}_1} \theta_x^{\text{Neck}} \theta_y^{\text{Neck}} \theta_x^{\text{R.Should.}} \theta_z^{\text{R.Should.}} \theta_x^{\text{L.Should.}} \theta_z^{\text{L.Should.}} \theta_x^{\text{R.Hip}} \theta_y^{\text{R.Hip}} \theta_x^{\text{L.Hip}} \theta_y^{\text{L.Hip}}\right], \tag{13}$$

$$\mathcal{X}_{\mathcal{H}_3} = \left[\mathcal{X}_{\mathcal{H}_2} \theta_y^{\text{R.Should.}} \theta_z^{\text{R.Elbow}} \theta_y^{\text{L.Should.}} \theta_z^{\text{L.Elbow}} \theta_z^{\text{R.Hip}} \theta_y^{\text{R.Knee}} \theta_z^{\text{L.Hip}} \theta_y^{\text{L.Knee}}\right], \tag{14}$$

where $\mathbf{x}$ and $\mathbf{R}$ stand for the global translation and rotation w.r.t. origin of coordinates, respectively. Note that the mapping functions $f(\cdot)$ in our case are trivial. Regarding the associated dimensions $K_{\mathcal{H}_i}$ introduced in Eq. (5), we have that:

$$K_{\mathcal{H}_1} = 6, K_{\mathcal{H}_2} = 16, K_{\mathcal{H}_3} = 22, \tag{15}$$

with the following associated incremental state space dimensions:

$$K^{\Delta}_{\mathcal{H}_1} = 6, K^{\Delta}_{\mathcal{H}_2} = 10, K^{\Delta}_{\mathcal{H}_3} = 6. \tag{16}$$

Although a selected SHBM is employed to track any person, the size of the limbs must be adequate to the particular subject under study. For the majority of people, there is a strong quasi-linear correlation between the height of a person and the length of the limbs [9] thus allowing a proper scaling of these magnitudes after automatically measuring the height directly from the input images as done in [31].

4.2. Particle assignment

In order to set a criterion to set the number of particles per model thread, $N_{\mathcal{H}_i}$, and the number of layers per model, $L_{\mathcal{H}_i}$, the effective number of particles associated to a given state space is introduced:

$$N^{\text{eff}}_{\mathcal{H}_i} = \sum_{j=i}^{M} N_{\mathcal{H}_j} \cdot L_{\mathcal{H}_j}. \tag{17}$$

Fig. $N^{\text{eff}}_{\mathcal{H}_i}$ quantizes the real amount of particles that will contribute to the estimation of a given state space variable. From this equation, it can be seen that variables from a given HBM, $\mathcal{H}_i$, are filtered by all following structural annealing layers. For instance, variables from the first model $\mathcal{H}_1$ associated to the global position and orientation of the torso will be filtered by all structural annealing layers since they are a set of very relevant variables. Variables associated to the last model of the hierarchy, being less important, are only filtered by their layers.

We devised the following method to set $N_{\mathcal{H}_i}$ proportional to the increment of dimensionality between state spaces:

$$N_{\mathcal{H}_i} \propto K^{\Delta}_{\mathcal{H}_i}. \tag{18}$$

Factor $L_{\mathcal{H}_i}$ is set through empirical validation.

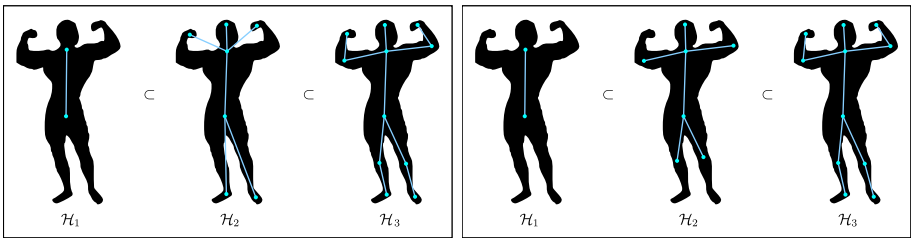
4.3. Propagation

Kinematic restrictions imposed by the angular limits at each joint of a given HBM may produce a more robust tracking output. In this field, some methods employ large volumes of annotated data to accurately model the angular cross-dependencies among joints [32] or to learn dynamic models associated to a given action [33]. In our case, these angular constraints will be enforced in the propagation step $P(\mathbf{y}^j_{t-1}, \Sigma_{\mathcal{H}_i})$ presented in Eq. (9). Typically, propagation consists in adding a random component to the state vector of a particle as

$$\tilde{\mathbf{y}}^j_t = P(\mathbf{y}^j_{t-1}, \Sigma_{\mathcal{H}_i}) = \mathbf{y}^j_{t-1} + \mathcal{N}(\mathbf{0}, \Sigma_{\mathcal{H}_i}) = \mathcal{N}(\mathbf{y}^j_{t-1}, \Sigma_{\mathcal{H}_i}). \tag{19}$$

That is to generate samples from a multi-variate Gaussian distribution centered at $\mathbf{y}^j_{t-1}$ with covariance matrix $\Sigma_{\mathcal{H}_i}$. However, this may lead to poses out of the legal angular ranges of the joints of the HBM. In order to avoid such effect, some works [34] add a term into the likelihood function that penalizes particles that do not fulfill the angular constraints. The following alternative is proposed: to take into account angular constrains and draw samples from a truncated Gaussian distribution [35], denoted as $\mathcal{N}^{\star}$ as shown in Fig. 5. In this way, particles are always generated within the allowed ranges thus avoiding the evaluation of particles that encode impossible poses and therefore increasing the performance of the sampling set.

Propagation of particles basically depends on the initial variances associated to each model filtering thread, $\Sigma_{\mathcal{H}_i}$, the structural annealing variance reduction, α_S, and the annealing variance reduction within the same filtering thread, $\alpha_{\mathcal{H}_i}$. It must be noted that the variance associated to a given HBM variable should always decrease as it is processed through either an inner or a structural annealing loop. Indeed, the propagation step assigns a higher drift



(a) Model $\mathcal{M}_1$ (b) Model $\mathcal{M}_1$

Fig. 4. Two SHBM analysis models employed in the HS-APF algorithm.

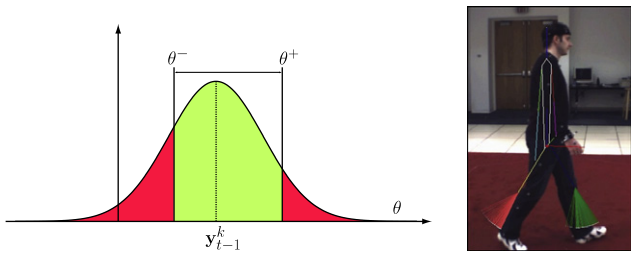


Fig. 5. Angular constraints enforcement by propagating particles within the allowed angular ranges $[\theta^-, \theta^+]$. In (a), samples are propagated following a truncated Gaussian distribution $\mathcal{N}^{\theta}$ centered at $\mathbf{y}_{t-1}{}^j$ with covariance matrix $\Sigma = \sigma$ bounded between θ^- and θ^+ (green zone). In (b), an example of particle propagation in the knee angle displaying how propagated particles never fall out the legal ranges ($\theta < 0$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

to the newly added variables of model $\mathcal{H}_i$ while assigning a lower drift to those that have been more recently filtered in the previous layers. Hence, for the filtering example presented in Fig. 2, we can state the initial $\Sigma_{\mathcal{H}_i}$ variances as:

$$\Sigma_{\mathcal{H}_1} = \mathrm{diag}\{\sigma_{\mathcal{H}_1}\}, \tag{20}$$

$$\Sigma_{\mathcal{H}_2} = \mathrm{diag}\left\{\left(\alpha_S \alpha_{\mathcal{H}_1}^{L_{\mathcal{H}_1}}\right)\sigma_{\mathcal{H}_1}, \sigma^{\Delta}_{\mathcal{H}_2}\right\}, \tag{21}$$

$$\Sigma_{\mathcal{H}_3} = \mathrm{diag}\left\{\left(\alpha_S^2 \alpha_{\mathcal{H}_1}^{L_{\mathcal{H}_1}} \alpha_{\mathcal{H}_2}^{L_{\mathcal{H}_2}}\right)\sigma_{\mathcal{H}_1}, \left(\alpha_S \alpha_{\mathcal{H}_2}^{L_{\mathcal{H}_2}}\right)\sigma^{\Delta}_{\mathcal{H}_2}, \sigma^{\Delta}_{\mathcal{H}_3}\right\}. \tag{22}$$

Or, in a more general way:

$$\Sigma_{\mathcal{H}_i} = \mathrm{diag}\left\{\left(\alpha_S^{i-1}\prod_{p=1}^{i-1}\alpha_{\mathcal{H}_j}^{L_{\mathcal{H}_j}}\right)\sigma_{\mathcal{H}_1}, \underbrace{\left(\alpha_S^{q-1}\prod_{p=q}^{i-1}\alpha_{\mathcal{H}_j}^{L_{\mathcal{H}_j}}\right)\sigma^{\Delta}_{\mathcal{H}_q}}_{q=1\ldots(i-1)}, \sigma^{\Delta}_{\mathcal{H}_i}\right\} \tag{23}$$

4.4. $G_{forward}$ operator

When the particle set associated to the HBM $\mathcal{H}_{i-1}, \{(\tilde{\mathbf{y}}_t^j, \tilde{\pi}_t^j)\}^{\mathcal{H}_{i-1}}$, has been filtered, the encoded *pdf* is transferred to the next model, $\mathcal{H}_i$, to improve its associated initial particle set $\{(\mathbf{y}_t^j, \pi_t^j)\}^{\mathcal{H}_i}$. The information delivery is performed by the operator G_{forward} that has to deal with two problems: the difference of the number of particles associated to each filtering thread ($N_{\mathcal{H}_i} \neq N_{\mathcal{H}_{i-1}}$) and the dimension difference between $\mathcal{X}_{\mathcal{H}_i}$ and $\mathcal{X}_{\mathcal{H}_{i-1}}$ ($K_{\mathcal{H}_i} > K_{\mathcal{H}_{i-1}}$). The first issue is addressed by a new technique called *sorting and adaptive resampling* while the second is tackled by means of *genetic crossing*.

4.4.1. Adaptive resampling

A first step to combine information from two different particle sets is to have the same number of elements in each set. Moreover, these two sets, $\{(\tilde{\mathbf{y}}_t^j, \tilde{\pi}_t^j)\}^{\mathcal{H}_{i-1}}$ and $\{(\mathbf{y}_{t-1}^j, \pi_{t-1}^j)\}^{\mathcal{H}_i}$, are both weighted (that is $\pi_t^j \neq N_{\mathcal{H}_i}^{-1}, \forall j, i$). We address this problem following the sorting and adaptive resampling procedure described as follows:

1. *Sorting:* Internal order within the particle set has not been considered previously since it does not affect the filtering operation. For the presented strategy, this order will be taken into account hence particles will be sorted decreasingly according to their associated weights. This operation will be applied to both particle sets as shown in Fig. 6a.
2. *Adaptive resampling:* Sampling Importance Resampling (SIR) [36] is the usual resampling strategy employed in PF systems due to its linear complexity and good performance. It is designed to produce an output set of $N_{\text{p}}^{\text{output}}$ resampled particles from an input set of the same number of elements, $N_{\text{p}}^{\text{input}}$. However, the problem of generating an output set with $N_{\text{p}}^{\text{output}} \neq N_{\text{p}}{}^{\text{input}}$ has not been addressed in the literature, as it is a very unusual requirement. For our purposes, we designed a variant of SIR algorithm able to cope with this requirement, reported in Algorithm 1 and depicted in Fig. 6(b). Note that resampling does not alter the order of the input and output vector. According to the previously applied sorting, particles with lower index in the output vector are resampled from particles with a higher weight in the input vector. Hence, despite all particles have the same weight, we can still distinguish which ones are the most relevant, according to their index value.

Merging information from the two particle sets once they have the same number of elements requires a criterion to combine their defining state space variables, $\mathcal{X}_{\mathcal{H}_{i-1}}$ and $\mathcal{X}_{\mathcal{H}_i}$. According to the SHBM construction stated in Eqs. (12)–(14), model $\mathcal{H}_i$ includes all variables from the preceding models. Hence, since variables from model $\mathcal{H}_{i-1}$ have been already filtered, it is proposed to combine these two particle sets by disregarding information from model $\mathcal{H}_{i-1}$ in model $\mathcal{H}_i$ and replacing these variables with the already filtered ones from the preceding filtering thread. That is:

$$\tilde{\mathbf{y}}_t^j \in \mathcal{X}_{\mathcal{H}_{i-1}}, \quad \mathbf{y}_{t-1}^k \in \mathcal{X}_{\mathcal{H}_i}, \quad \mathbf{y}_{t-1}^{k,\Delta} \in \mathcal{X}^{\Delta}_{\mathcal{H}_i}, \quad \mathbf{y}_{t-1}^k = \left[\tilde{\mathbf{y}}_t^j \mathbf{y}_{t-1}^{k,\Delta}\right]. \tag{24}$$

Note that indices i and k are not set to be the same. The procedure to associate these two indices will conform the particle combination algorithm.

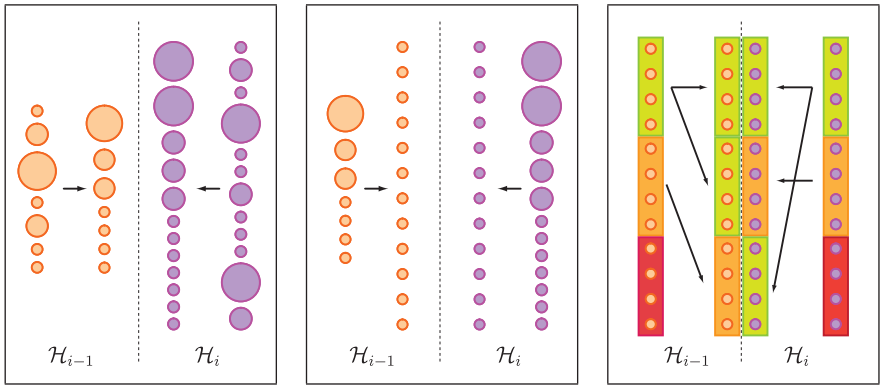
Algorithm 1. Systematic Adaptive Resampling Algorithm

$c_1 = \pi_t^1$
for $j = 2$ **to** $N_{\text{p}}^{\text{input}}$ **do**
  $c_j = c_{j-1} + \pi_t^j$
end
Draw a starting point $u_1 \sim \mathbf{U}[0, 1/N_{\text{p}}^{\text{output}}]$
$j = 1$
for $i = 1$ **to** $N_{\text{p}}^{\text{output}}$ **do**
  $u_i = u_1 + (i-1)/N_{\text{p}}^{\text{output}}$
  while $u_i > c_j$ & $j < N_{\text{p}}^{\text{input}}$ **do**
    $j = j + 1$
  end
  $\{\mathbf{x}_t^i, \pi_t^i\} = \{\mathbf{x}_t^j, 1/N_{\text{p}}^{\text{output}}\}$
end

4.4.2. Genetic crossing

Once the adaptively resampled particles belonging to two consecutive HBM have been generated (see Fig. 6b), it is required to design a combination rule in order to produce a resulting particle set, $\{(\mathbf{y}_{t-1}^j, \pi_{t-1}^j)\}^{\mathcal{H}_i}$, benefiting from this already filtered information. A direct association, $i \to j$, combining the best particles of each model has been tested and founded unable to cope with fast and unexpected motion since weak hypothesis from both models were rarely considered. Instead, we defined a more versatile particle combination method, inspired on genetic or biologically inspired algorithms [37], to develop the following cross-over technique.

Let us first define a partition over our sorted and resampled particle set of equal size denoted as $\mathcal{S}_{\mathcal{H}_i}^n, 1 \leqslant n \leqslant N$, being N the number of partitions. Again, the lower the n index, the more relevant the partition (in terms of resample particles originated from particles with higher weights). Then we define an association rule be-



(a) Sorting (b) Adaptive resampling (c) Genetic crossing

Fig. 6. Combination process of particles from two different state spaces corresponding to two different HBMs following the sorting, adaptive resampling and genetic crossing methodology.

tween the sets $\mathcal{S}^n_{\mathcal{H}_{i-1}}$ and $\mathcal{S}^m_{\mathcal{H}_i}$ based on generating combinations of sets with high indices but also allowing combinations of sets with high and low indices. In this way, some variability is introduced thus becoming more robust to rapid motion and sudden pose changes. The rule to generate such index correspondence has been set empirically as shown in Fig. 6c.

4.5. $G_{backward}$ operator

Once reaching the last HBM model, $\mathcal{H}_M$, the filtered particle set $\{(\mathbf{y}^j_t, \pi^j_t)\}^{\mathcal{H}_M}$ contains the most accurate and detailed estimation of the HBM. Taking into account that state variables of $\mathcal{H}_M$ are also represented in $\mathcal{H}_i, i < M$, by means of Eq. (7), we might use this information to update the already filtered particle sets $\{(\tilde{\mathbf{y}}^j_t, \tilde{\pi}^j_t)\}^{\mathcal{H}_i}$. In this way, the initial particle set to be filtered at time $t+1$ at each HBM analysis thread will be derived from the best estimation at time t. Basically, operator G_{backward} will first sort and generate $M-1$ adaptively resampled sets from $\{(\mathbf{y}^j_t, \pi^j_t)\}^{\mathcal{H}_M}$ with input dimension $K_{\mathcal{H}_M}$ and output dimensions $K_{\mathcal{H}_i}$, $1 \leqslant i < M$ (note that filtered sets $\{(\tilde{\mathbf{y}}^j_t, \tilde{\pi}^j_t)\}^{\mathcal{H}_i}$ are already ordered from the G_{forward} operation). Then the variables associated to the state space subset $\mathcal{X}^{\Lambda}_{\mathcal{H}_i}$ of each HBM are replaced with the values from the adaptively resampled set derived from $\{(\mathbf{y}^j_t, \pi^j_t)\}^{\mathcal{H}}_M$.

5. Markerless multi-camera HMC with HS-APF

Processing multiple images separately exploiting calibration information for HMC has been a common research direction [1]. However, this strategy turned out to be very sensitive to perspective and occlusion issues and cluttered backgrounds thus requiring setup scenarios. As a solution, data fusion towards generating a 3D representation of the scene unifying information from several camera views allowed fitting a HBM using binary [8] and colored [22] voxels. Efficient implementations of the shape from silhouette algorithms required to generate the voxel reconstruction proved this 3D representation appropriate towards real-time applications [21].

In order to test the efficiency of the proposed HS-APF method, a voxel-based markerless approach is presented. Basically, two elements have to be defined: the employed input data, $\mathbf{z}_t$, and the evaluation of the likelihood function, $w_{\mathcal{H}_i}(\tilde{\mathbf{y}}^j_t, \mathbf{z}_t)$.

5.1. Measurement generation

For a given frame in the video sequence, a set of N_C images are obtained from the N_C cameras. Each camera is modeled using a pinhole camera model based on perspective projection with camera calibration information available. Foreground regions from input images are obtained using a standard background learning and substraction technique [38] and these data is used to generate a 3D voxel reconstruction of the scene using a shape-from-silhouette process [21,39] (see an example in Fig. 3a), with a pre-defined resolution of the data, that is the voxel size $s_{\mathcal{V}}$. Let us denote this set as $\mathcal{V}$ and its associated surface voxels as $\mathcal{V}^S$.

5.2. Likelihood evaluation

Likelihood $w_{\mathcal{H}_i}(\tilde{\mathbf{y}}^j_t, \mathbf{z}_t)$ is computed for every particle estimating how measurement $\mathbf{z}_t = \{\mathcal{V}, \mathcal{V}^S\}$ fits with the pose of a given HBM $\mathcal{H}_i$ encoded in particle $\tilde{\mathbf{y}}^j_t$. The commonly used silhouette overlap and edge distance measures employed in 2D measurements [13,30] are extended to 3D as volume intersection and surface distance measures.

5.2.1. Volume based likelihood

In order to define a meaningful measure between the pose encoded in particle $\tilde{\mathbf{y}}^j_t$ and the available data $\mathbf{z}_t$, a relation between $\tilde{\mathbf{y}}^j_t$ and the 3D voxelized space must be stated. This can be achieved by defining an appearance model of the HBM, that is to "flesh out" the HBM skeleton with a volumetric model of the limbs, torso and head [8]. In our particular case, truncated cones in the 3D discretized space are used. Another approach is to obtain a more accurate representation by using a 3D surface mesh model computed with a 3D scan of the specific subject [40,41]. We have chosen to use truncated cones in order to make the system more general. However, HS-APF does not depend on the "flesh out" method and could also be used for mesh-based models.

Let us define the voxel representation of the fleshed HBM as the set $\mathcal{V}^{\text{HBM}}_{\mathbf{y}}$ related with the pose described by $\tilde{\mathbf{y}}^j_t$ as shown in Fig. 7a.

The set $\mathcal{V}^{\text{HBM}}_{\mathbf{y}}$ is constructed by performing an union (with addition) among the individual volumes of the torso, $\mathcal{V}_{\mathcal{T}}$, and all limbs, $\mathcal{V}_{\mathcal{P}_{ij}}, \forall i,j$, that is:

$$\mathcal{V}^{\text{HBM}}_{\mathbf{y}} = \biguplus_{\mathcal{Y} \in \{\mathcal{V}_{\mathcal{T}}, \mathcal{V}_{\mathcal{P}_{ij}}\}} \mathcal{Y}, \quad 1 \leqslant i \leqslant N_{\mathcal{L}}, 1 \leqslant j \leqslant N_{\mathcal{P}(i)}. \tag{25}$$

Operator $\uplus$ refers to the operation that assigns to each voxel of the 3D space the number of intersections among all body parts in that position, as shown in Fig. 7c.

According to the representation $\mathcal{V}^{\text{HBM}}_{\mathbf{y}}$ and the available raw voxel data $\mathcal{V}$, we may define the output, double occupancy and occupancy scores for every body part $\mathcal{Y} \in \left\{\mathcal{V}_{\mathcal{T}}, \mathcal{V}_{\mathcal{P}_{ij}}\right\}, \forall i,j$, as:

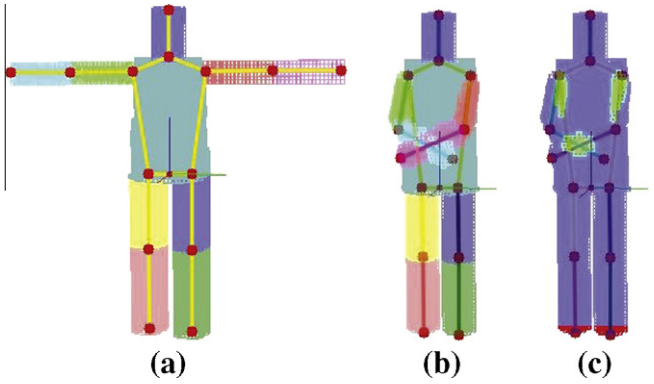


Fig. 7. HBM analysis based on the voxel set $\mathcal{V}_{\mathbf{y}}^{\mathrm{HBM}}$. In (a), an example of the appearance of the employed HBM. In (b), an invalid pose depicted with false colors to distinguish body parts and, in (c), the set $\mathcal{V}_{\mathbf{y}}^{\mathrm{HBM}}$. Blue voxels stand for places with only one body limb occupying that space while green regions stand for places with two limbs occupying that space. Red regions denote those voxels falling out of the scene. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$\rho_{\mathcal{Y}}^{\mathrm{Out}} = \frac{|\{\mathcal{V} \in \mathcal{Y} | \mathcal{V} \notin \text{Analysis scene}\}|}{|\mathcal{Y}|}, \tag{26}$$

$$\rho_{\mathcal{Y}}^{\mathrm{DO}} = \frac{\left|\left\{\mathcal{V} \in \mathcal{Y} | \mathcal{V}_{\mathbf{y}}^{\mathrm{HBM}}(\mathcal{V}) > 1\right\}\right|}{|\mathcal{Y}|}, \tag{27}$$

$$\rho_{\mathcal{Y}}^{\mathrm{Occ}} = \frac{\left|\left\{\mathcal{V} \in \mathcal{Y} | \mathcal{V}_{\mathbf{y}}^{\mathrm{HBM}}(\mathcal{V}) \geqslant 1 \& \boldsymbol{\mathcal{V}}(\mathcal{V}) \neq 0\right\}\right|}{|\mathcal{Y}|}, \tag{28}$$

where $\boldsymbol{\mathcal{V}}(\mathcal{V})$ stands for the content of V at the position of voxel $\mathcal{V}$ and $|\mathcal{Y}|$ for the number of non-zero elements in set $\mathcal{Y}$. Output score $\rho_{\mathcal{Y}}^{\mathrm{Out}}$ will quantize the amount of voxels of a given body part that fall outside of the analyzed scene. Interpenetration among limbs may occur even when a valid pose is evaluated. In this case, score $\rho_{\mathcal{Y}}^{\mathrm{DO}}$ measures the degree of double occupancy or interpenetration. These two figures will determine those regions of the state space $\mathcal{X}_{\mathcal{H}_i}$ to be avoided since poses resulting in high values of $\rho_{\mathcal{Y}}^{\mathrm{Out}}$ and/or $\rho_{\mathcal{Y}}^{\mathrm{DO}}$ are likely to be invalid. Finally, the occupancy score $\rho_{\mathcal{Y}}^{\mathrm{Occ}}$ measures the fraction of the body part that is occupied. Ideally, a good match will yield low values of $\rho_{\mathcal{Y}}^{\mathrm{Out}}$ and $\rho_{\mathcal{Y}}^{\mathrm{DO}}$ and high values of $\rho_{\mathcal{Y}}^{\mathrm{Occ}}$, for every body part.

5.2.2. Surface based likelihood

Surface data is smoothed with a Gaussian mask and the obtained voxel values are re-mapped between 0 and 1. This produces a voxel map $\widetilde{\boldsymbol{\mathcal{V}}}^{\mathrm{S}}$, in which each voxel is assigned a value related to its proximity to a surface. Finally, the surface measurement is defined as:

$$\rho_{\mathcal{Y}}^{\mathrm{Surf}} = \frac{1}{|\mathcal{Y}|}\sum_{\mathcal{V}\in\mathcal{Y}}(1 - \widetilde{\boldsymbol{\mathcal{V}}}^{\mathrm{S}}(\mathcal{V})), \quad \mathcal{Y} \in \{\boldsymbol{\mathcal{V}}_{\mathcal{T}}^{\mathrm{S}}, \boldsymbol{\mathcal{V}}_{\mathcal{P}_{ij}}^{\mathrm{S}}\}, \forall i,j. \tag{29}$$

Low values of this score indicate a proper alignment of the body part with the input data.

5.2.3. Joint likelihood function

A common assumption in Monte Carlo based HMC algorithms [13,34] is to consider a statistical independence among limbs. Therefore, likelihood function $w(\mathbf{z}_t, \mathbf{y}_t^j)$ can be constructed as:

$$w(\mathbf{z}_t, \mathbf{y}) \propto p\left(\{\boldsymbol{\mathcal{V}}_t, \boldsymbol{\mathcal{V}}_t^{\mathrm{S}}\} | \mathcal{V}_{\mathbf{y}}^{\mathrm{HBM}}\right) = \prod_{\mathcal{Y}} \in \{\boldsymbol{\mathcal{V}}_{\mathcal{T}}, \boldsymbol{\mathcal{V}}_{\mathcal{P}_{ij}}\} p(\{\boldsymbol{\mathcal{V}}_t, \boldsymbol{\mathcal{V}}_t^{\mathrm{S}}\} | \mathcal{Y}). \tag{30}$$

Assuming that the involved errors follow a Gaussian distribution [42], an accurate way to define the likelihood function for individual body parts is

$$p(\{\boldsymbol{\mathcal{V}}_t, \boldsymbol{\mathcal{V}}_t^{\mathrm{S}}\} | \mathcal{Y}) \propto \exp\left\{-\frac{1}{2}(\mathbf{d}-\boldsymbol{\mu})^{\top}\Sigma_{\mathcal{Y}}^{-1}(\mathbf{d}-\boldsymbol{\mu})\right\}, \tag{31}$$

where parameters $\mathbf{d}$,$\boldsymbol{\mu}$ and $\Sigma_{\mathcal{Y}}$ are defined as:

$$\mathbf{d} = \left[\rho_{\mathcal{Y}}^{\mathrm{Out}}, \rho_{\mathcal{Y}}^{\mathrm{DO}}, \rho_{\mathcal{Y}}^{\mathrm{Empty}}, \rho_{\mathcal{Y}}^{\mathrm{Surf}}\right], \boldsymbol{\mu} = \mathbf{0}, \tag{32}$$

$$\rho_{\mathcal{Y}}^{\mathrm{Empty}} = 1 - \rho_{\mathcal{Y}}^{\mathrm{Occ}}, \tag{33}$$

$$\Sigma_{\mathcal{Y}} = \mathrm{diag}(\sigma_{\mathrm{Out}}^2, \sigma^2{}_{\mathrm{DO}}, \sigma_{\mathrm{Empty}}^2, \sigma^2{}_{\mathrm{Surf}}). \tag{34}$$

Values of variances were empirically set to $\sigma^2{}_{\mathrm{Out}} = \sigma_{\mathrm{DO}}^2 = 0.01, \sigma_{\mathrm{Empty}}^2 = \sigma^2{}_{\mathrm{Surf}} = 0.1$, leading to satisfactory results.

6. Experiments and results

In order to test the proposed algorithm, HumanEva data sets [24] have been selected since they provide synchronized and calibrated data from both several cameras and a professional motion capture system to produce ground truth data. Two metrics are suggested in [24], the mean, μ, and the standard deviation of the estimation error, σ, towards providing quantitative and comparable results. Moreover, in this article, metrics proposed in [43] for 3D

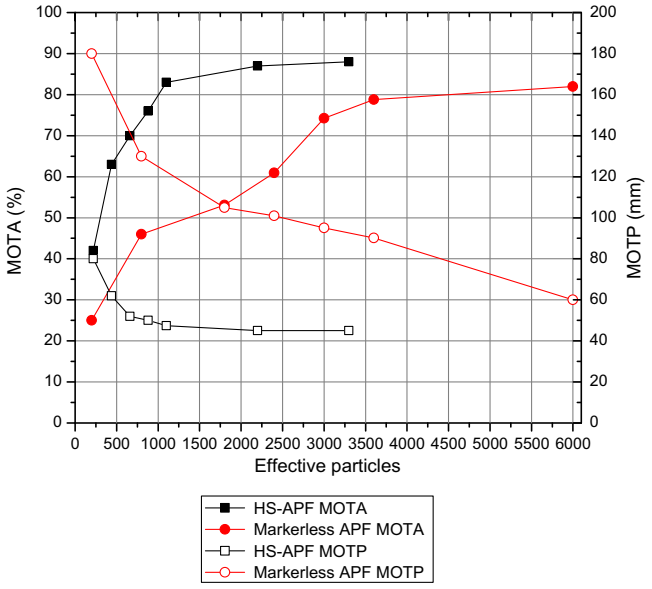


Fig. 8. Comparison of the performance scores *MMTA* and *MMTP* of the HS-APF and a fixed HMC-APF algorithm. Computational complexity is compared through the number of effective particles.

Table 1
HS-APF tracking results on HumanEva-I dataset. *MMTP* and *MMTA* scores are computed using ϵ = 5 cm.

	Model $\mathcal{M}_1$			
	μ (mm)	σ	*MMTP* (mm)	*MMTA* (%)
Walking	42.11	24.95	39.27	83.19
Jog	46.90	26.71	42.62	75.08
Throw/Catch	64.22	32.17	51.84	68.11
Gesture	53.55	30.81	50.40	71.77
Box	58.53	27.96	48.89	72.16
Average	*54.06*	*31.71*	*47.46*	*76.85*
Walking	43.07	26.12	40.21	82.53
Jog	46.51	27.18	43.09	73.85
Throw/Catch	58.85	26.79	50.05	72.52
Gesture	48.10	29.12	42.91	76.14
Box	55.19	26.54	45.31	77.21
Average	*51.34*	*28.51*	*45.31*	*76.42*

human pose tracking evaluation are also provided and employed, namely the **M**ultiple **M**arker **T**racking **A**ccuracy (*MMTA*), defined as the percentage of 3D body landmarks positions whose estimation error is below a threshold ϵ, and the **M**ultiple **M**arker **T**racking **P**recision (*MMTP*), defined as the average of the estimation error of those landmarks considered by the *MMTA*.

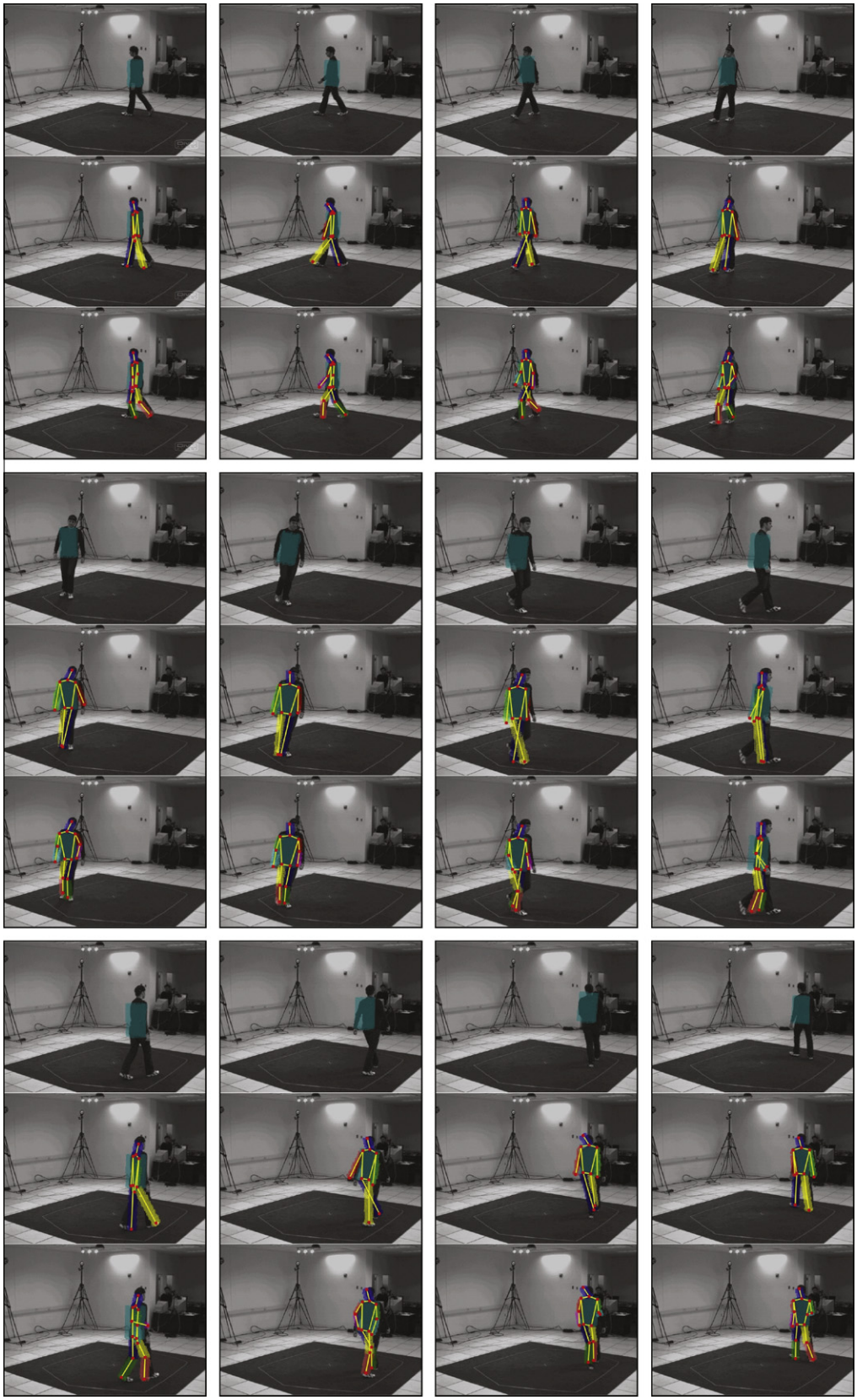


Fig. 9. HS-APF operation example for action walking. The three involved HBM are stacked for every frame.

Particle assignment has been done following Eq. (17) as:

$$N_{\mathcal{H}_i} = \gamma \cdot K^{\Delta}_{\mathcal{H}_i}. \tag{35}$$

The internal annealing has been set to $L_{\mathcal{H}_i} = 2$ and an exploratory analysis over a fraction of the dataset has been conducted for the two proposed SHBMs in Section 4.1, model $\mathcal{M}_1$ and $\mathcal{M}_2$ towards finding the optimal working point as $\gamma = 50$. In this case, we obtain

$$\begin{aligned} &N_{\mathcal{H}_1} = 300, \quad N_{\mathcal{H}_2} = 500, \quad N_{\mathcal{H}_3} = 300, \\ &N^{\text{eff}}_{\mathcal{H}_1} = 2200, \quad N^{\text{eff}}_{\mathcal{H}_2} = 1600, \quad N^{\text{eff}}_{\mathcal{H}_3} = 600. \end{aligned} \tag{36}$$

Initial variance values employed in Eq. (23), $\sigma_{\mathcal{H}_i}$, are set to be half of the maximum variation expected in each joint angle. Variance reduction among HBMs, has been set to $\alpha_S = 0.8$ in order to keep an smooth transition among HBMs while likelihood annealing variance reduction rate is set to $\alpha_{\mathcal{H}_i} = 0.5$.

Regarding input data $\mathbf{z}_t$, the 3D reconstruction has been done with voxels of size $s_v = 2$ cm. Larger voxel sizes allow a faster processing of volumes at a cost of losing resolution and distinguishability among body parts. It must be noted that, for the type of analyzed sequences, the obtained input data $\mathbf{z}_t = \{\mathcal{V}, \mathcal{V}^S\}$ is noisy due to faulty foreground/background segmentation.

Results for the HS-APF algorithm are reported in Table 1 and a visual example is shown in Fig. 9. The comparison of the performance of the HS-APF using the two aforementioned analysis models, shows that motions where limbs are mostly straight are well captured when using model $\mathcal{M}_1$ (Fig. 4a) as in the case of walking or jogging. Activities with a high flexion of limbs such as gesturing or boxing are better captured using model $\mathcal{M}_2$ (Fig. 4b). It has been tested that, for a large value of γ, that is a large value of overall particles, both methods tend to converge to the same performance level.

Another effect observed in the operation of the HS-APF algorithm is its ability to deal with corrupted data. In cases where there is a sudden missing of a part of the data (typically, in the legs segment), the simplest model, $\mathcal{H}_0$, is able to keep tracking the torso part regardless of the poor accuracy of the system in the affected limbs. When the data quality is back to normal, the system is able to adapt. An example of the execution of the HS-APF algorithm can be found in http://www.cristiancanton.org.

6.1. Scalable vs. non-scalable comparison

In order to prove the performance gain obtained when employing a scalable HBM in comparison with the usage of a fixed HBM (the highest model in the hierarchy of $\mathcal{M}_1$ or $\mathcal{M}_2$), the proposed technique is contrasted with a voxel-based HMC system employing an APF with the same likelihood evaluation and propagation implementation described in Sections 4.3, 5.1 and 5.2 [44]. In Fig. 8 we plot the *MMTP* and *MMTA* scores related with the number of effective particles of both algorithms. It can be seen how, by exploiting the hierarchical structure of the human body, we can obtain better results with a lower computational complexity. For example, for a fixed computational load, $N_p = 2200$, there is an improvement of $\Delta(MMTP, MMTA) = (58.0, 51.2)\%$.

When comparing the obtained results with our first approach to SHBM-based motion capture [23], a noticeable performance increase is achieved both by employing surface information yielding to a more discriminative likelihood function and by the more efficient exploration of the SHBM associated state spaces through the presented genetic crossing strategy (in comparison with the naïve one mentioned in [23]).

6.2. State-of-the-art comparison

A number of algorithms in the literature have been evaluated using HumanEva data sets and their results are reported in Table 2. However, some algorithms presented results only using a fraction of HumanEva-I or alternatively used the HumanEva-II database, which is significantly smaller (only two sequences) and involves a very reduced set of motions (walking and jogging).

Markerless HMC has been addressed from two points of view: as a tracking problem or as a learning/classification problem. Whithin the tracking ones, those applying linear techniques [10] are prone to loose track, clearly outperformed by those relying on a Monte Carlo approach [34,44,45]. All the compared algorithms analyzed the images from each camera separately except the markerless voxel-based APF presented in [44]. Some algorithms are based on learning motion patterns in order to efficiently drive the particles on the state space, as done in [34,45,46], hence being

Table 2
Result comparisons with state of the art methods evaluated over HumanEva datasets. The presented score corresponds to the mean of the error estimation μ, as reported by the compared authors in their respective contributions. Standard deviation of the error σ is provided between parenthesis when available.

Method	Walk	Jog	Box	Gesture	Throw/Cach
HumanEva-I					
Hierarchical Partitioned PF [34]	101.9	–	–	–	–
PF+Dynamic model [45]	100.4 (59.6)	–	–	–	–
ICP+Naïve classification [47]	53.1 (20.9)	–	45.4 (16.8)	–	–
Example-based pose estimation [16]	44.29 (20.7)	42.74 (14.7)	90.74(41.7)	50.87(13.3)	75.2 (27.7)
Example-based pose estimation [48]	37.98	–	–	–	–
Sparse probabilistic regression [17]	32.7	31.2	38.5	–	–
Bayesian mixture of experts [49]	25.6	26.7	30.4	24.3	46.2
Twin Gaussian Processes [46]	27.6	30.5	89.46	40.1	57.3
Markerless APF [44]	96.52 (41.6)	130.34 (62.0)	145.22 (42.6)	124.87 (45.6)	122.27 (52.1)
Structured output-associative regression [50]	33.0	34.2	81.4	44.5	75.0
HS-APF (first version) [23]	115.21 (20.3)	–	–	–	–
HS-APF (Model $\mathcal{M}_1$)	42.11(24.9)	49.90 (26.7)	58.53 (27.9)	53.55(30.8)	64.22 (32.1)
HS-APF (Model $\mathcal{M}_2$)	43.07 (26.1)	46.51 (27.1)	55.19 (26.5)	48.10 (29.1)	58.85 (26.7)

HumanEva-II Method	S2	S4
EM+Kinematically constrained GMM [10]	137.00 (152.0)	177.00 (196.0)
Hierarchical Partitioned PF [34]	149.40	156.60
Example-based pose estimation [16]	170.40 (105.8)	179.47(91.2)
Action priors [51]	44.90 (9.5)	45.20 (13.4)
HS-APF (Model $\mathcal{M}_1$)	57.71 (31.2)	61.86 (33.6)
HS-APF (Model $\mathcal{M}_2$)	63.30 (35.0)	65.22 (36.7)

unsuitable to track gestures not present in the training corpus. A similar problem is found in the classification based HMC algorithms [16,17,47–51] that achieve high performance when tracking previously learnt motion patterns.

When comparing the proposed technique with some of these state-of-the-art algorithms, it can be seen that HS-APF outperforms the reviewed tracking-based approaches. However, for those techniques relying on learning and classification, the performance is within the same order in most of the cases.

7. Conclusions

This paper presents a general framework to address estimation and tracking problems where a scalable hierarchy can be defined within the analysis model, that is the scalable human body model. An extension of the annealing concept to exploit the structure of the SHBM is presented, the *structural annealing*, and a Monte Carlo based algorithm is proposed. Information from different models within the hierarchy is transferred through the adaptive resampling and genetic crossing techniques. Finally, an implementation of the HS-APF operation is described for markerless HMC using a 3D voxel reconstruction of the scene as the system input. Finally, the validity of this approach is proved using the standard HumanEva HMC datasets and the results are compared with other state-of-the-art available techniques.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.cviu.2011.06.001.

References

[1] R. Poppe, Vision-based human motion analysis: an overview, Computer Vision and Image Understanding 108 (1–2) (2007) 4–18.
[2] Q. Cai, J. Aggarwal, Human motion analysis: a review, Journal of Computer Vision and Image Understanding 73 (3) (1999) 428–440.
[3] L. Ding, A. Martinez, Modelling and recognition of the linguistic components in american sign language, Image and Vision Computing 27 (12) (2009) 1826–1844.
[4] CHIL – Computers in the Human Interaction Loop, FP-6 European Integrated Project. <http://chil.server.de> (2004–2007).
[5] W. Hu, T. Tan, L. Wang, S. Maybank, A survey on visual surveillance of object motion and behaviors, IEEE Transactions on Systems, Man, and Cybernetics 34 (3) (2004) 334–352.
[6] P. Kauff, O. Schreer, An immersive 3D video-conferencing system using shared virtual team user environments, in: Proceedings of International Conference on Collaborative Virtual Environments, 2002, pp. 105–112.
[7] Q. Cai, J. Aggarwal, Tracking human motion in structured environments using a distributed-camera system, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (11) (2002) 1241–1247.
[8] I. Mikič, M. Trivedi, E. Hunter, P. Cosman, Articulated body posture estimation from multi-camera voxel data, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2001, pp. 455–460.
[9] S. Dockstader, M. Berg, A. Tekalp, Stochastic kinematic modeling and feature extraction for gait analysis, IEEE Transactions on Image Processing 12 (8) (2003) 962–976.
[10] S. Cheng, M. Trivedi, Articulated body pose estimation from voxel reconstructions using kinematically constrained Gaussian mixture models: algorithm and evaluation, in: Proceedings of 2nd Workshop on Evaluation of Articulated Human Motion and Pose Estimation, 2007.
[11] M. Arulampalam, S. Maskell, N. Gordon, T. Clapp, A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, IEEE Transactions on Signal Processing 50 (2) (2002) 174–188.
[12] M. Isard, A. Blake, CONDENSATION – conditional density propagation for visual tracking, International Journal of Computer Vision 29 (1) (1998) 5–28.
[13] J. Deutscher, I. Reid, Articulated body motion capture by stochastic search, International Journal of Computer Vision 61 (2) (2005) 185–205.
[14] J. Mitchelson, A. Hilton, Simultaneous pose estimation of multiple people using multiple-view cues with hierarchical sampling, in: Proceedings of British Machine Vision Conference, 2003.
[15] J. MacCormick, M. Isard, Partitioned sampling, articulated objects, and interface-quality hand tracking, in: Proceddings of European Conference on Computer Vision, 2000, pp. 3–19.
[16] R. Poppe, Evaluating example-based pose estimation: experiments on the humaneva sets, in: Proceedings of 2nd Workshop on Evaluation of Articulated Human Motion and Pose Estimation, 2007.
[17] R. Urtasun, T. Darrell, Sparse probabilistic regression for activity-independent human pose inference, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.
[18] VICON. <http://www.vicon.com>.
[19] P. Cerveri, A. Pedotti, G. Ferrigno, Robust recovery of human motion from video using Kalman filters and virtual humans, Human Movement Science 22 (2003) 377–404.
[20] A. Kirk, J. O'Brien, D. Forsyth, Skeletal parameter estimation from optical motion capture data, in: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, vol. 2, 2005, pp. 782–788.
[21] G. Cheung, T. Kanade, J. Bouguet, M. Holler, A real time system for robust 3D voxel reconstruction of human motions, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, 2000, pp. 714–720.
[22] F. Caillette, A. Galata, T. Howard, Real-time 3D human body tracking using variable length Markov models, in: Proceedings of British Machine Vision Conference, vol. 1, 2005, pp. 469–478.
[23] C. Canton-Ferrer, J.R. Casas, M. Pardàs, Exploiting structural hierarchy in articulated objects towards robust motion capture, in: Proceedings of International Conference on Articulated Motion and Deformable Objects, Lecture Notes on Computer Science, vol. 5098, 2008, pp. 82–91.
[24] L. Sigal, A. Bălan, M. Black, HumanEva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion, International Journal Computer Vision 87 (1–2) (2010) 4–27.
[25] C. Bregler, J. Malik, Tracking people with twists and exponential maps, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, No. 1, 1998, pp. 8–15.
[26] S. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (7) (1989) 674–693.
[27] J. Bergen, P. Anandan, K. Hanna, R. Hingorani, Hierarchical model-based motion estimation, in: Proceedings of European Conference on Computer Vision, Lecture Notes on Computer Science, vol. 588, 1992, pp. 237–252.
[28] T. Foures, P. Joly, Scalability in human shape analysis, in: Proceedings of IEEE International Conference on Multimedia and Expo, 2006, pp. 2109–2112.
[29] C. Theobalt, M. Magnor, P. Schuler, H. Seidel, Combining 2D feature tracking and volume reconstruction for online video-based human motion capture, in: Proceedings of 10th Pacific Conference on Computer Graphics and Applications, 2002, pp. 96–103.
[30] L. Raskin, M. Rudszky, E. Rivlin, Using hierarchical models for 3D human body-part tracking, in: Proceedings of the 16th Scandinavian Conference on Image Analysis, 2009, pp. 11–20.
[31] C. Canton-Ferrer, J.R. Casas, M. Pardàs, Towards a Bayesian approach to robust finding correspondences in multiple view geometry environments, in: Proceedings of 4th International Workshop on Computer Graphics and Geometric Modelling, Lecture Notes on Computer Science, vol. 3515, 2005, pp. 281–289.
[32] L. Herda, R. Urtasun, P. Fua, Hierarchical implicit surface joint limits for human body tracking, Computer Vision and Image Understanding 99 (2) (2005) 189–209.
[33] R. Urtasun, D. Fleet, P. Fua, 3D people tracking with Gaussian process dynamical models, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2006, pp. 238–245.
[34] Z. Husz, A. Wallace, Evaluation of a hierarchical partitioned particle filter with action primitives, in: Proceedings of 2nd Workshop on Evaluation of Articulated Human Motion and Pose Estimation, 2007.
[35] J. Kotecha, P. Djuric, Gibbs sampling approach for generation of truncated multivariate Gaussian random variables, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 3, 1999.
[36] N. Gordon, D. Salmond, A. Smith, Novel approach to nonlinear/non-Gaussian Bayesian state estimation, IEE Proceedings on Radar and Signal Processing 140 (2) (1993) 107–113.
[37] M. Mitchel, An Introduction to Genetic Algorithms, MIT Press, 1998.
[38] C. Stauffer, W. Grimson, Adaptive background mixture models for real-time tracking, in: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, 1999, pp. 252–259.
[39] J. Landabaso, M. Pardàs, J. Casas, Shape from inconsistent silhouette, Computer Vision and Image Understanding 112 (2) (2008) 210–224.
[40] S. Corazza, L. Mündermann, E. Gambaretto, G. Ferrigno, T. Andriacchi, Markerless motion capture through visual hull, articulated ICP and subject specific model generation, International Journal of Computer Vision 87 (1) (2010) 156–169.
[41] J. Gall, B. Rosenhahn, T. Brox, H. Seidel, Optimization and filtering for human motion capture, International Journal of Computer Vision 87 (1) (2010) 75–92.
[42] J. Lichtenauer, M. Reinders, E. Hendriks, Influence of the observation likelihood function on particle filtering performance in tracking applications, in: Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition, 2004, pp. 767–772.
[43] C. Canton-Ferrer, J. Casas, M. Pardàs, E. Monte, Towards a fair evaluation of 3D human pose estimation algorithms, Tech. Rep., Technical University of Catalonia, 2009. <www.cristiancanton.org>.
[44] C. Canton-Ferrer, J.R. Casas, M. Pardàs, Voxel based annealed particle filtering for markerless 3D articulated motion capture, in: Proceedings of IEEE Conference on 3DTV, 2009, pp. 2644–2647.

[45] M. Brubaker, D. Fleet, A. Hertzmann, Physics-based human pose tracking, in: Proceedings of Workshop on Evaluation of Articulated Human Motion and Pose Estimation, 2006.
[46] L. Bo, C. Sminchisescu, Twin Gaussian processes for structured prediction, International Journal of Computer Vision 87 (1–2) (2010) 25–52.
[47] L. Münderman, S. Corazza, T. Andriacchi, Markerless human motion capture through visual hull and articulated ICP, in: Proceedings of 1st Workshop on Evaluation of Articulated Human Motion and Pose Estimation, 2006.
[48] R. Okada, S. Soatto, Relevant feature selection for human pose estimation and localization in cluttered images, in: Proceedings of European Conference on Computer Vision, 2008.
[49] L. Bo, C. Sminchisescu, A. Kanaujia, D. Metaxas, Fast algorithms for large scale conditional 3D prediction, in: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.
[50] L. Bo, C. Sminchisescu, Structured output-associative regression, in: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, 2009, pp. 2403–2410.
[51] J. Gall, A. Yao, L. Van Gool, 2D action recognition serves 3D human pose estimation, in: Proceedings European Conference on Computer Vision, Lecture Notes on Computer Science, vol. 6313, 2010, pp. 425–438.