VOXEL BASED ANNEALED PARTICLE FILTERING FOR MARKERLESS 3D ARTICULATED MOTION CAPTURE

C. Canton-Ferrer, J.R. Casas, M. Pardàs

Image Processing Group, Technical University of Catalonia (Spain)

ABSTRACT

This paper presents a view-independent approach to markerless human motion capture in low resolution sequences from multiple calibrated and synchronized cameras. Redundancy among cameras is exploited to generate a 3D voxelized representation of the scene and a human body model (HBM) is introduced towards analyzing these data. An annealed particle filtering scheme where every particle encodes an instance of the pose of the HBM is employed. Likelihood between particles and input data is performed using occupancy and surface information and kinematic constrains are imposed in the propagation step towards avoiding impossible poses. Test over the HumanEva annotated dataset yield quantitative results showing the effectiveness of the proposed algorithm.

Index Terms— Human motion capture, multi-camera analysis, particle filtering, voxel processing

1. INTRODUCTION

Automatic human motion capture (HMC) has been studied extensively [1] basically fostered by the number of potential applications and its inherent complexity. This research area contains a number of hard and often 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 [2], unusual behavior detection in security applications [3] or tele-conferencing [4].

Recovering the pose of a human body model (HBM) involves estimating highly dimensional and multimodal statistic distributions associated to the defining parameters of this HBM. In this field, some contributions employed linear techniques such as Kalman filtering [5] although being prone to loose track. However, due to the multimodal shape of the involved distributions, techniques based on the particle filtering algorithm (PF) proved to be efficient to tackle this problem as done in [6]. Finally, state-of-the-art annealed particle filtering (APF) introduced by [7] in the field of HMC noticeably improved the performance and robustness of the obtained results. In some cases, annotated data allowed analyzing specific actions [6] yielding to a more efficient exploration of these distributions when tracking motion patterns present in the training corpus.

Processing multiple images separately exploiting calibration information have been a common research direction [7] but it 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 [5] and colored [6] voxels. Efficient implementations of the shape from silhouette algorithms required to generate the voxel reconstruction proved this 3D representation appropriate towards real-time applications [8].

In this paper, a multi-camera algorithm for markerless human motion tracking is presented using an extension of the APF with 2D image measurements to 3D using a dynamic binary voxel reconstruction. In this way, the presented algorithm gets rid of perspective and clutter issues while exploiting the APF properties through a particle likelihood evaluation based on volume overlap and surface distances. Kinematic restrictions applied in the particle propagation step allows avoiding impossible poses. Finally, effectiveness of the proposed algorithm is assessed by means of objective metrics.

2. HBM BASED ANNEALED PARTICLE FILTERING

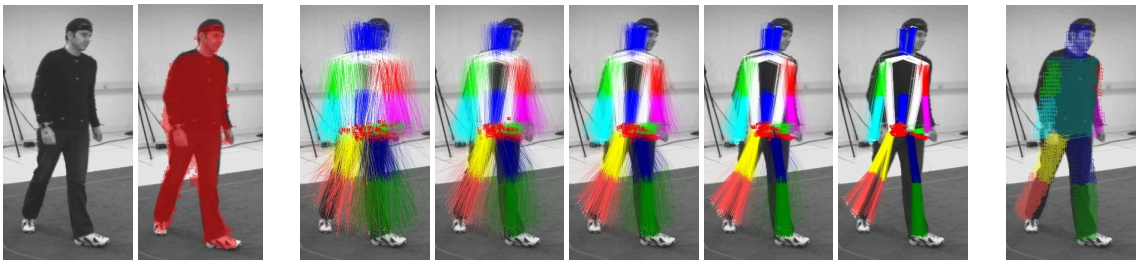
Let us define a state space $\mathcal{X} \subset \mathbb{R}^D$ formed by the D defining parameters of an articulated HBM, in our case, the angles at the joints and the global translation and rotation of the model with respect to the real world, adding up to $D = 27$ (see an example in Fig.1). Estimating the optimal pose of this HBM at time t, $\widehat{\mathcal{X}}_t$, given a set of noisy observations $\mathbf{z}_{1:t}$ up to time t, involves computing a representation of the posterior likelihood $p(\mathcal{X}|\mathbf{z}_{1:t})$, that usually exhibits a multimodal shape. Particle filtering [9] has been found suitable to tackle such problems but, due to the high dimensionality of the state space, the number of particles required to efficiently explore $\mathcal{X}$ turns out to be computationally unfeasible.

Annealed particle filtering [7] was presented in the context of HMC as a technique to efficiently estimate $p(\mathcal{X}|\mathbf{z}_{1:t})$ requiring far less particles than PF, hence allowing affordable implementations. APF introduces a layered posterior estimation where a set of N_{P} weighted particles $\left\{\mathbf{y}_t^j \in \mathcal{X}, w_t^j \in \mathbb{R}\right\}_{j=1}^{N_{\mathrm{P}}}$ are evaluated and propagated through a set of N_{L} progressively smoothed versions of the likelihood function (also called annealing layers) thus avoiding getting trapped in local maxima. Finally, once reaching the last annealing layer, pose $\widehat{\mathcal{X}}_t$ is computed as the weighted average of all particles. An example of the APF operation is depicted in Fig.1.

Following the standard APF algorithm, some factors are to be taken into account when implementing it: the measurement generation, the likelihood evaluation and the propagation model.

2.1. 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 [10] and these data is used to generate a 3D voxel reconstruction of the scene using a shape-from-silhouette process [8] (see an example in Fig.1a). Let us denote this set as $\boldsymbol{\mathcal{V}}^{\mathrm{R}}$ and its associated surface voxels as $\boldsymbol{\mathcal{V}}^{\mathrm{S}}$.



(a) Image and voxels input (b) Annealing PF (c) Result

Fig. 1. APF operation example. In (a), the input image and the voxel reconstruction projection. In (b), the progressive fitting of particles driven by the annealing process and, in (c), the final pose estimation $\widehat{\mathcal{X}}_t$.

2.2. Likelihood Evaluation

Likelihood measure is computed for each particle estimating how well the encoded body pose fits with the input data $\mathbf{z}_t$. We extend the commonly used silhouette overlap and edge distance measures employed in 2D measurements [7] to 3D as volume intersection and surface distance measures.

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

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

where $N_{\mathcal{P}(i)}$ stands for the number of parts in the i-th limb. The torso, limbs and their sub-parts are connected to one another by means of joints, $J_{i,j}$. In order to constrain the possible poses that this HBM may adopt to be valid, we define a number of degrees of freedom (DoF) and a legal angular range at each joint.

The proposed model has its root position in the pelvis and adds up to 21 DoF distributed as follows: 3 DoF at each shoulder, 1 DoF at the elbows, 3 DoF at the hips, 1 DoF at the knees and 3 DoF at the waist. Apart of these DoF, we must consider the translation and rotation of the root with respect to the world, resulting in the 27 DoF associated to $\mathcal{H}$ (see Fig.2a).

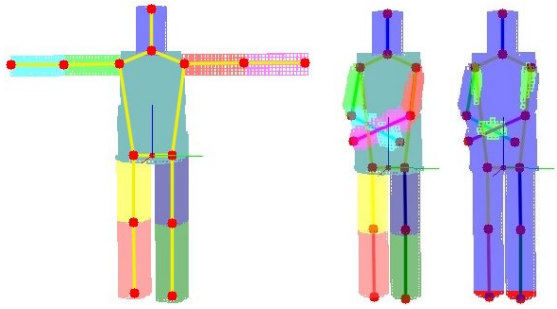
Volume based Likelihood

In order to define a meaningful measure between the pose encoded by a given particle $\mathbf{y} \in \mathcal{X}$ and the available data $\mathbf{z}_t = \{\boldsymbol{\mathcal{V}}^{\mathrm{R}}, \boldsymbol{\mathcal{V}}^{\mathrm{S}}\}$, we have to establish a relation between $\mathbf{y}$ and the 3D voxelized space. 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. In our particular case, we will use truncated cones in a the 3D discretized space. Let us define the voxel representation of this fleshed HBM as the set $\boldsymbol{\mathcal{V}}_{\mathbf{y}}^{\mathrm{HBM}}$ related with the pose described by $\mathbf{y}$; Fig.2a depicts an example.

The set $\boldsymbol{\mathcal{V}}_{\mathbf{y}}^{\mathrm{HBM}}$ will allow us measuring the fitness of a given pose $\mathbf{y}$ with respect to the input data $\mathbf{z}_t$. This set will be constructed by performing an union (with addition) among the individual volumes of the torso, $\boldsymbol{\mathcal{V}}_{\mathcal{T}}$, and all limbs, $\boldsymbol{\mathcal{V}}_{\mathcal{P}_{i,j}}, \forall i, j$, that is:

$$\boldsymbol{\mathcal{V}}_{\mathbf{y}}^{\mathrm{HBM}} = \biguplus_{\boldsymbol{\mathcal{Y}} \in \{\boldsymbol{\mathcal{V}}_{\mathcal{T}}, \boldsymbol{\mathcal{V}}_{\mathcal{P}_{i,j}}\}} \boldsymbol{\mathcal{Y}}, \qquad 1 \leq i \leq N_{\mathcal{L}}, 1 \leq j \leq N_{\mathcal{P}(i)}. \quad (2)$$

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.2c.



(a) (b) (c)

Fig. 2. HBM analysis based on the voxel set $\boldsymbol{\mathcal{V}}_{\mathbf{y}}^{\mathrm{HBM}}$. In (a), an example of the employed HBM. In (b), an invalid pose depicted with false colors to differentiate body parts and, in (c), the set $\boldsymbol{\mathcal{V}}_{\mathbf{y}}^{\mathrm{HBM}}$. 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. Three limbs occupying the same region is an odd case and may be produced with very awkward poses. Red regions denote those voxels falling out of the scene.

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

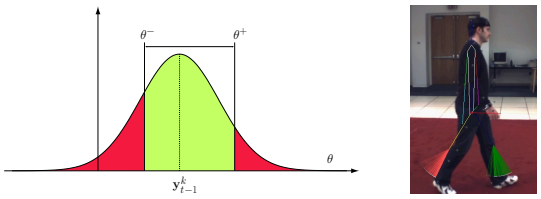
$$\rho_{\boldsymbol{\mathcal{Y}}}^{\mathrm{Out}} = \frac{|\{\mathcal{V} \in \boldsymbol{\mathcal{Y}} | \mathcal{V} \notin \text{Analysis scene}\}|}{|\boldsymbol{\mathcal{Y}}|}, \quad (3)$$

$$\rho_{\boldsymbol{\mathcal{Y}}}^{\mathrm{DO}} = \frac{|\{\mathcal{V} \in \boldsymbol{\mathcal{Y}} | \boldsymbol{\mathcal{V}}_{\mathbf{y}}^{\mathrm{HBM}}(\mathcal{V}) > 1\}|}{|\boldsymbol{\mathcal{Y}}|}, \quad (4)$$

$$\rho_{\boldsymbol{\mathcal{Y}}}^{\mathrm{Occ}} = \frac{|\{\mathcal{V} \in \boldsymbol{\mathcal{Y}} | \boldsymbol{\mathcal{V}}_{\mathbf{y}}^{\mathrm{HBM}}(\mathcal{V}) \geq 1 \& \boldsymbol{\mathcal{V}}^{\mathrm{R}}(\mathcal{V}) \neq 0\}|}{|\boldsymbol{\mathcal{Y}}|}, \quad (5)$$

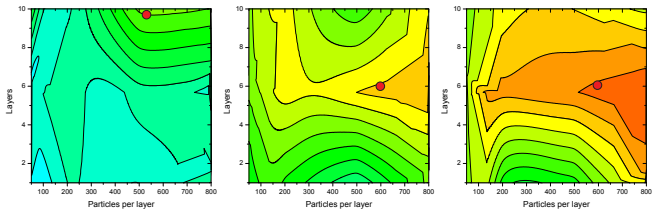
where $\boldsymbol{\mathcal{V}}(\mathcal{V})$ stands for the content of $\boldsymbol{\mathcal{V}}$ at the position of voxel $\mathcal{V}$ and $|\boldsymbol{\mathcal{Y}}|$ for the number of non-zero elements in set $\boldsymbol{\mathcal{Y}}$. These set of measures will allow assessing the fitness between the pose $\mathbf{y}$ and the data $\boldsymbol{\mathcal{V}}^{\mathrm{R}}$. Output score $\rho_{\boldsymbol{\mathcal{Y}}}^{\mathrm{Out}}$ will quantize the amount of voxels of a given body part that fall out of the analyzed scene. Interpenetration among limbs may occur even when a valid pose is evaluated. In this case, score $\rho_{\boldsymbol{\mathcal{Y}}}^{\mathrm{DO}}$ measures the degree of double occupancy or interpenetration. These two figures will determine those regions of the state space $\mathcal{X}$ to be avoided since poses resulting in high values of $\rho_{\boldsymbol{\mathcal{Y}}}^{\mathrm{Out}}$ and/or $\rho_{\boldsymbol{\mathcal{Y}}}^{\mathrm{DO}}$ are likely to be invalid.

Finally, the occupancy score $\rho_{\boldsymbol{\mathcal{Y}}}^{\mathrm{Occ}}$ measures the fraction of the body part that is occupied. Ideally, a good match will yield low values of $\rho_{\boldsymbol{\mathcal{Y}}}^{\mathrm{Out}}$ and $\rho_{\boldsymbol{\mathcal{Y}}}^{\mathrm{DO}}$ and high values of $\rho_{\boldsymbol{\mathcal{Y}}}^{\mathrm{Occ}}$, for every body part.



(a) (b)

Fig. 3. Angular constraints enforcement. In (a), particles are propagated using a truncated Gaussian distribution $\mathcal{N}^\star$ centered at $\mathbf{y}^k_{\mathbf{t}-1}$ with covariance matrix Σ 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 ($\theta < 0$).



$s_\mathcal{V} = 3$ cm, $N_\mathrm{C} = 3$ $s_\mathcal{V} = 3$ cm, $N_\mathrm{C} = 7$ $s_\mathcal{V} = 2$ cm, $N_\mathrm{C} = 7$

Fig. 4. Data quality influence. For every pair of voxel size $s_\mathcal{V}$ and number of cameras N_C, we display the *MMTA* scores for different number of annealing layers N_L (y axis) and particles per layer N_p (x axis). Red dots mark the optimal operation point with minimal absolute number of particles $N_\mathrm{L}N_\mathrm{p}$.

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 $\widetilde{\mathcal{V}}^\mathrm{S}$, in which each voxel is assigned a value related to its proximity to a surface. Finally, the surface measurement is defined as:

$$\rho^{\text{Surf}}_{\mathcal{Y}} = \frac{1}{|\mathcal{Y}|}\sum_{\mathcal{V}\in\mathcal{Y}}\left(1-\widetilde{\mathcal{V}}^\mathrm{S}\left(\mathcal{V}\right)\right), \mathcal{Y}\in\{\mathcal{V}^\mathrm{S}_\mathcal{T}, \mathcal{V}^\mathrm{S}_{\mathcal{P}_{i,j}}\}, \forall i,j. \quad (6)$$

Joint likelihood function

Likelihood function $w(\mathbf{z}_t, \mathbf{y}^j_t)$ is defined assuming a statistical independence among limbs as:

$$w(\mathbf{z}_t,\mathbf{y}) \propto p\left(\{\mathcal{V}^\mathrm{R}_t,\mathcal{V}^\mathrm{S}_t\}|\mathcal{V}^{\text{HBM}}_\mathbf{y}\right) = \prod_{\mathcal{Y}\in\{\mathcal{V}_\mathcal{T},\mathcal{V}_{\mathcal{P}_{i,j}}\}} p\left(\{\mathcal{V}^\mathrm{R}_t,\mathcal{V}^\mathrm{S}_t\}|\mathcal{Y}\right). \quad (7)$$

As done previously in [7], likelihood function for individual body parts is approximated to:

$$p\left(\{\mathcal{V}^\mathrm{R}_t,\mathcal{V}^\mathrm{S}_t\}|\mathcal{Y}\right) \propto \exp\left\{-\frac{1}{2}(\mathbf{d}-\boldsymbol{\mu})^\top\Sigma^{-1}_\mathcal{Y}(\mathbf{d}-\boldsymbol{\mu})\right\}, \quad (8)$$

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

$$\begin{aligned}\mathbf{d} &= \left[\rho^{\text{Out}}_\mathcal{Y}, \rho^{\text{DO}}_\mathcal{Y}, \rho^{\text{Empty}}_\mathcal{Y}, \rho^{\text{Surf}}_\mathcal{Y}\right], \quad \rho^{\text{Empty}}_\mathcal{Y} = 1-\rho^{\text{Occ}}_\mathcal{Y}, \quad (9)\\ \boldsymbol{\mu} &= \mathbf{0}, \quad \Sigma_\mathcal{Y} = \text{diag}(\sigma^2_{\text{Out}},\sigma^2_{\text{DO}},\sigma^2_{\text{Empty}},\sigma^2_{\text{Surf}}).\end{aligned}$$

Values of variances were empirically set by analyzing a part of the analyzed corpus to $\sigma^2_{\text{Out}} = 0.01, \sigma^2_{\text{DO}} = 0.01, \sigma^2_{\text{Empty}} = 0.1, \sigma^2_{\text{Surf}} = 0.1$ leading to satisfactory results.

2.3. Particle Propagation

Kinematic restrictions imposed by the angular limits at each joint may produce a more robust tracking output. Employing a previously learnt motion model in the particle propagation step can improve tracking results if annotated data is available [6]. However, these methods are constrained to deal with motions present in the training corpus thus being not suitable for unconstrained motion tracking. In this paper, angular constraints are enforced in the propagation step of the APF scheme. Usually, the propagation step consists in adding a random component to the state vector of a particle as:

$$\mathbf{y}^k_t = \mathbf{y}^k_{t-1} + \mathcal{N}(\mathbf{0},\Sigma) = \mathcal{N}(\mathbf{y}^k_{t-1},\Sigma), \quad (10)$$

where $\mathcal{N}(\boldsymbol{\mu},\Sigma)$ stands for a random multivariate Gaussian distribution of mean $\boldsymbol{\mu}$ and covariance matrix Σ. Diagonal elements of covariance matrix Σ are set to half of the maximum expected variation of each variable of the state space over one time step. However, this propagation may lead to poses out of the joint legal angular ranges. In our work, we present the following technique: when propagating particles, angular constraints are taken into account and samples of a truncated Gaussian distribution, denoted as $\mathcal{N}^\star$, are generated instead of a complete Gaussian distribution, as shown in Fig.3. In this way, particles are always generated within the allowed ranges.

3. EXPERIMENTS AND RESULTS

In order to test the proposed algorithm, the standard HumanEva [11] dataset is employed. This dataset contains a set of 5 actions performed by 3 different subjects at a resolution of 640x480 pixels and a frame-rate of 25 fps. Ground truth data is available and two metrics are defined (mean, μ, and standard deviation, σ, of the estimation error) towards providing quantitative and comparable results. The initial dataset was partitioned in a training dataset containing the 10% of the total data to set up the tracking variables and the rest was used for testing.

Let $X = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_M\}$, $\mathbf{x}_m \in \mathbb{R}^3$, denote the M landmark positions of the HBM (typically, the body joints and the end of the limbs) corresponding to the pose described by the state variable $\mathbf{y}$ computed using forward kinematics [12]. Assuming that landmark positions $\hat{\mathbf{x}}_m$ associated to particle $\mathbf{y}^j_t$ are available, we can define a *matched* marker estimation $\hat{\mathbf{x}}_m$ with respect to the ground truth position $\mathbf{x}_m$ as the one fulfilling $\epsilon = \|\mathbf{x}_m - \hat{\mathbf{x}}_m\| < \delta$. This stands for those estimations that fall δ-close to the ground truth position. Then, the **M**ultiple **M**arker **T**racking **A**ccuracy (*MMTA*), is defined as the percentage of markers $\mathbf{x}_m \in X$ fulfilling the $\epsilon < \delta$ condition, and the **M**ultiple **M**arker **T**racking **P**recision (*MMTP*), as the average of the metric error between $\hat{\mathbf{x}}_m$ and $\mathbf{x}_m$, of all pairs fulfilling $\epsilon < \delta$. Finally, these scores are averaged for all frames in the sequence.

Training dataset was evaluated using 3D reconstructions made with different voxel sizes $s_\mathcal{V}$ and using a different number of cameras N_C towards deciding the optimal number of layers N_L and particles per layer N_p and the influence of the quality of the input data into the overall performance. *MMTA* score has been chosen as the most significant figure to assess the APF performance as shown in Fig.4 where high values of N_C and low values of $s_\mathcal{V}$ lead to better performance since these magnitudes have a crucial influence in the quality of the 3D reconstruction. The chosen working with point for $s_\mathcal{V} = 2$ cm and $N_\mathrm{C} = 7$ is found to be $N_\mathrm{L} = 6$ and $N_\mathrm{p} = 600$, leading to a total of 3600 efficient particles.

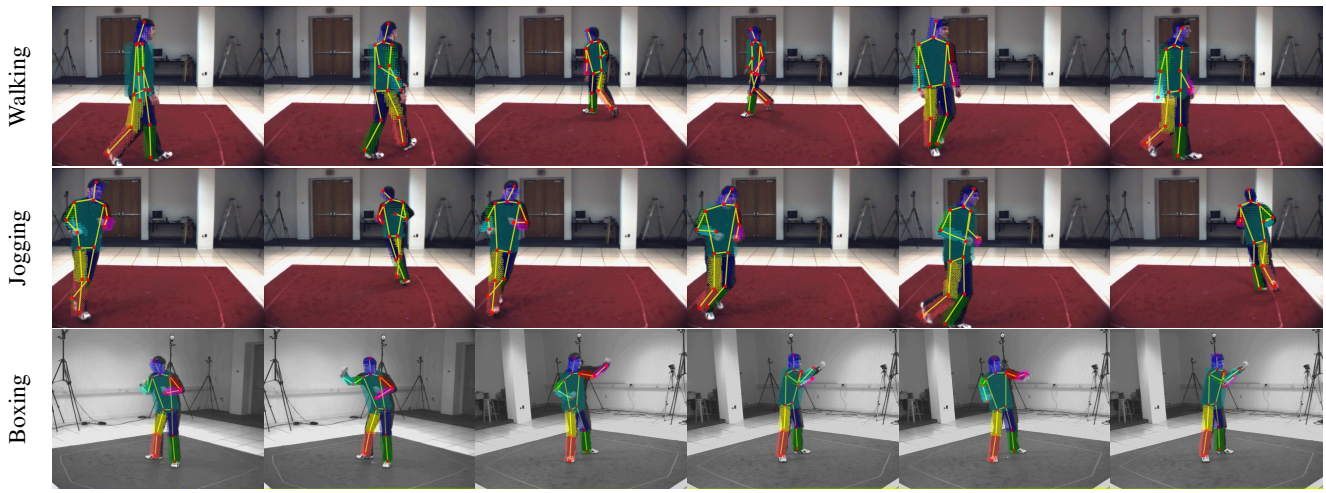


Fig. 5. Human motion capture examples over the HumanEva dataset.

HumanEva dataset was analyzed using reconstructions made with $s_{\mathcal{V}} = 2$ cm and $N_{\mathrm{C}} = 7$ producing the results in Table 1. Averaged *MMTA* and *MMTP* scores indicates that in 71% of analyzed frames, difference between the estimation and the ground truth is below $\delta = 10$ cm and the commited error in these frames has an average of 9 cm. When comparing the performance for individual actions, it can be seen that those involving fast motion (i.e. boxing or jogging) exhibit a lower tracking performance than the others (i.e. walking). Some examples are depicted in Fig.5.

	μ (mm)	σ	*MMTP* (mm)	*MMTA*
Walking	96.52	41.64	72.05	79.55
Jog	130.34	62.01	92.21	68.24
Throw/Catch	145.22	52.13	94.69	61.30
Gesture	124.87	45.66	90.43	69.17
Box	122.27	42.68	92.77	68.38
Average	**121.18**	**45.92**	**90.17**	**71.36**

Table 1. Quantitative results for the HumanEva dataset using $N_{\mathrm{L}} = 6$ and $N_{\mathrm{p}} = 600$ with $\delta = 100$ mm.

4. CONCLUSIONS AND FUTURE WORK

This paper presents a robust markerless approach for human motion capture using multiple calibrated and synchronized cameras. Spatial redundancy is exploited by generating a 3D voxelized reconstruction of the scene and a HBM is defined to perform the analysis. Progressive fitting of the HBM through the APF algorithm using an occupancy and surface likelihood function and a kinematically constrained particle propagation model allowed an accurate estimation of the body pose. Quantitative evaluation based on HumanEva dataset assessed the robustness of the technique.

Future work aims at exploiting scalability of the HBM towards designing fitting algorithms able to cope with occluded body parts observed when there are occluding elements in the scene. Other research lines aim at gait and motion disorders analysis. A comparison with other existing methods using the same dataset is under study.

5. REFERENCES

[1] R. Poppe, "Vision-based human motion analysis: An overview," *Computer Vision and Image Understanding*, vol. 108, pp. 4–18, 2007.

[2] "CHIL - Computers in the Human Interaction Loop," http://chil.server.de, 2004-2007.

[3] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 34:3, pp. 334–352, 2004.

[4] P. Kauff and O. Schreer, "An immersive 3D video-conferencing system using shared virtual team user environments," in *Proc. Int. Conf. on Collaborative Virtual Environments*, 2002, pp. 105–112.

[5] I. Mikič, M. Trivedi, E. Hunter, and P. Cosman, "Articulated body posture estimation from multi-camera voxel data," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, 2001, vol. 1, pp. 455–460.

[6] F. Caillette, A. Galata, and T. Howard, "Real-time 3D human body tracking using variable length Markov models," in *Proc. British Machine Vision Conference*, 2005, vol. 1, pp. 469–478.

[7] J. Deutscher and I. Reid, "Articulated body motion capture by stochastic search," *Intl. Journal of Computer Vision*, vol. 61:2, pp. 185–205, 2005.

[8] G. Cheung, T. Kanade, J.Y. Bouguet, and M. Holler, "A real time system for robust 3D voxel reconstruction of human motions," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2000, vol. 2, pp. 714–720.

[9] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[10] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, 1999, pp. 252–259.

[11] L. Sigal and M.J. Black, "HumanEva: Synchronized video and motion capture dataset for evaluation of articulated human motion," Tech. Rep. CS-06-08, Brown University, 2006.

[12] G. Guerra-Filho, "Optical motion capture: theory and implementation," *Journal of Theoretical and Applied Informatics*, vol. 12:2, pp. 61–89, 2005.