

# Tracking Human Body Parts Using Particle Filters Constrained by Human Biomechanics

J. Martínez<sup>1</sup>, J.C. Nebel<sup>2</sup>, D. Makris<sup>2</sup> and C. Orrite<sup>1</sup>

<sup>1</sup>Computer Vision Lab, University of Zaragoza, Spain

<sup>2</sup>Digital Imaging Research Centre, Kingston University, UK

## Abstract

In this paper, a novel framework for visual tracking of human body parts is introduced. The presented approach demonstrates the feasibility of recovering human poses with data from a single uncalibrated camera using a limb tracking system based on a 2D articulated model. It is constrained only by biomechanical knowledge about human bipedal motion, instead on relying on constraints linked to a specific activity or camera view. These characteristics make our approach suitable for real visual surveillance applications. Experiments on HumanEva I & II datasets demonstrate the effectiveness of our method on tracking human lower body parts. Moreover, a detail comparison with current tracking methods is presented.

## 1 Introduction

Human motion modelling is one of the most active areas in computer vision. It can be defined as the ability to estimate, at each frame of a video sequence, the position of each joint of a human figure which is represented by an articulated model. Because of the high dimensional space of human motion, tracking methods based on 3D anthropomorphic articulated models have proved the most effective [12, 13, 14, 15, 16]. Their applications include analysis of human activity, entertainment, ambient intelligence and medical diagnosis to name a few. However, their main drawback is they generally rely on data capture synchronously by several cameras which have been accurately calibrated. Therefore, these techniques are unpractical for applications targeting unconstrained environments such as video-surveillance. The alternative is usage of tracking methods based on 2D models which cannot deal by themselves with the intrinsic ambiguity of projected 3D postures, self occlusions and distortions introduced by camera perspective. Therefore, they are usually restricted to well defined motions and specific camera views; however these constraints reduce their value in many real applications.

We propose a novel framework based on a set of particle filters to track human body parts. It relies on a generative approach based on a 2D model constrained only by human biomechanics. The inclusion of biomechanical knowledge about bipedal motion significantly reduces the complexity of the problem. This reduction of complexity is achieved by the detection of the pivot foot - i.e. the foot which is static during a step - and its

---

<sup>0</sup>This work was partially supported by the EPSRC sponsored MEDUSA and PRoCeSS projects (Grant No. EP/E001025/1 and EP/E033288 respectively), TIN2006-11044 and Feder from Spanish Ministry of Education, and FPI grant (BES-2004-3741) from MEyC.

trajectory during a whole step. Then using our 2D model, tracking of human body parts is achieved using a set of particle filters [17, 12], which iteratively refine their solution.

In this work, we concentrate our effort on tracking a subject legs since the other body parts do not benefit from biomechanics constraints. Our results are evaluated against the HumanEva data set which is becoming the standard for assessing human body tracking algorithms [5]. After a brief description of the state of the art in human body part tracking, we present an overview of our methodology. Then we detail the biomechanics constraints we use and key algorithms. Finally, after presentation and evaluation of our results, conclusions are drawn.

## 2 Related work

Tracking complexity increases exponentially with the number of targets when their motion is not independent from each other. Articulated models have been shown to be essential tools to handle tracking and detection tasks by reinforcing motion constraints in either the 2D [21] or 3D space [11] so that motions of subparts are interrelated. Several approaches have been investigated to alleviate this challenge, such as dynamic programming [1], annealed sampling [18], partitioned sampling [15], eigenspace tracking [10], hybrid Monte Carlo filtering [19] and bottom-up [6] approaches.

While 3D articulated models are the most suitable to human tracking, they can only be applied in dedicated environments [12, 13, 14, 15, 16]. Therefore, many 2D models have been proposed. In [1], tracking of an articulated object is performed in two steps. First, each limb is tracked separately with a dynamic Markov network. Secondly, positions are refined by introducing constraints between the different subparts with mean field Monte Carlo where a set of low dimensional particle filters interact with each other. In [2], robust tracking in planar patches is performed by incorporating feature weights in RANSAC and maximum likelihood consensus algorithms. In [3], authors propose a planar-patch articulated model where the relationship between the model and the image is estimated using one particle filter for each limb and the constraints between limbs are represented by interaction potentials. The main drawback of all these 2D models is their usage is restricted to specific types of motions which are usually linear and seen from a fixed point of view.

Unlike the previous models whose constraints restrict the type of sequences they can handle, our approach is based on an advance 2D model designed to tackle 3D motion patterns such as changes in the pose of the object with respect to the camera. Therefore, we will be able to deal with variations in rotation and depth. To achieve this without introducing strong motion constraints which would restrict the application of our system, we propose to use some specific knowledge about biomechanics and human gait analysis.

## 3 Methodology

### 3.1 General Principle

Human motion is highly multi-modal and this non-gaussianity is amplified in the image plane by the camera perspective. Therefore, a tracking framework capable of working with non-linear distribution is required. Since particle filter has been successfully applied

for this purpose [17], we will apply this algorithm within our tracking framework. Our scheme is based on a set of particle filters to fit a 2D articulated model on each frame of a video sequence. In addition, we take advantage of a biomechanics constraint inherent in human bipedal motion: during a ‘step’, one leg pivots around a single point. This allows us dealing with much more activities than other techniques which rely on training on a specific activity. Since we are able to detect the position of this point, this constraint is integrated in an asymmetrical 2D model where the two legs are treated differently. Finally, model fitting is performed after different trackers have been applied successively.

Initially, a ‘standard’ particle filter process operates to track lower limb locations until the end of the ‘step’. Due to the high dimensionality of the problem and the ill-conditioned model, it may not be able to produce satisfactorily tracking. In order to refine the tracking of the articulated model, two assistant particle filters are then launched in parallel using information intrinsic to the ‘step’ of interest. The main reason for using two trackers instead of one is to handle the degradation and potential divergence of tracking over time.

To take advantage of the ‘pivot’ point constraint and trajectory information, we propose to rely on data captured during a full ‘step’ before completing the tracking task. While a short delay is introduced - typically around 10 frames (i.e. 0.5s) - in a real time system, this allows processing a wide range of human activities without loss of accuracy. Although some actions, such as running or jumping, break the ‘pivot’ constraint during short periods of time and the ‘pivot’ point can be momentarily occluded, this can be detected and handled without affecting significantly the proposed tracking framework.

### **3.2 Biomechanics constraints for human motion tracking**

Most human motion tracking methods rely on constraints such as specific activity, constant velocity, linear or periodic motion which critically impact on their accuracy and/or their genericity. Study of human biomechanics, however, reveals that human motion itself provides some explicit constraints. In this section, we show they can be utilised to simplify the task of tracking human body parts. Walking is a very common human activity whose many other motions, such as loitering, balancing and dancing, can be seen as derivatives and where the underlying mechanics of walking can be applied. All these bipedal motions are based on a series of ‘steps’ defined as one leg ‘swinging’ around a ‘support’ leg whose foot, or ‘pivot’, stays in contact with the ground at any instant [20].

Therefore, the detection of this pivot point from a video sequence permits a significant reduction of the complexity of the tracking task without important restriction regarding the types of motions which can be processed by the tracker. Knowledge of the precise position of the pivot foot also allows using different strategies for tracking either the ‘support’ or the ‘swinging’ leg, which enhances the power of our 2D model. Moreover, positions of consecutive striking feet provide some information about the subject’s trajectory in the image plane which supplies clues regarding the relative camera-subject position.

In addition to the ‘pivot’ foot constraint, the ‘support’ leg has another property: upper and lower legs are supposed to be aligned during the pivot motion around the static foot. Therefore, estimate of the locations of the associated knee and hip is refined if they do not form a straight line with the pivot foot.

In our framework, the static foot is detected using the algorithm proposed in [4]. It is based on the biomechanics of gait motion. During the strike phase, the foot of the striking leg stays at the same position for half a gait cycle, whilst the rest of the human

body moves. The pivot foot is detected using a low-level feature: corners produced by the Harris corner detector. Corners associated to the pedestrian of interest are accumulated across several frames (i.e. 20 in our implementation). The region where the leg strikes the ground must have a high density of corners. Although this approach is usually efficient, motions towards or away from the camera produce many points seen as static on the body. We deal with this by apply a double filtering process to remove outliers by maintaining both temporal and spatial coherences of the ‘pivot’ point.

### 3.3 Multiple particle filter tracking based on 2D articulated model

#### 3.3.1 2D asymmetrical articulated model informed by trajectory information

Our model aims to track simultaneously the global position of the subject in the image as well as the relative position of the different parts of the limbs. Thus, the tracker state vector is composed of the image coordinates of the hip points and the parameters which model the relative motions and positions, such as angles and lengths in the image plane. In order to introduce the biomechanics constraints, which rely on a relative independence between both legs, both hip points are employed as references and the angles of both legs with respect to the hips are included in the state vector. The state vector of each leg is described by the following equation:

$$X_{leg} = [x_{hip}, y_{hip}, \dot{x}_{hip}, \dot{y}_{hip}, \theta_{thigh}, \theta_{leg}, \dot{\theta}_{thigh}, \dot{\theta}_{leg}, l_{thigh}, l_{leg}, \dot{l}_{thigh}, \dot{l}_{leg}] \quad (1)$$

where  $x$  and  $y$  are the coordinates in pixels,  $\theta$  is the angle between a limb and the  $x$  axis and  $l$  is the length of the limb (see figure 2a).

Once the ‘support’ leg is estimated, the hip point of the ‘swinging’ leg is constrained by the distance between the two hips, which is set at a fixed anthropometric value,  $D_0$ . Moreover, the two hips points are supposed to share the same  $y$  coordinate.

Due to its nature, 2D tracking allows a higher flexibility and simplicity of use and initialisation than 3D tracking. However, in 2D it is not possible to introduce traditional constraints, such as motion dynamic or kinematics. Instead we include 3D properties in the 2D world. In the 3D world, the distance between the hips remains constant over time. However, when this fixed distance is projected in the camera plane, its value is changed by two different parameters: the location and the orientation. Whereas the location introduces a factor of scale which is estimated with the global size of the legs, the orientation distort this distance in a non-linear way which depends on the view point.

Because of the stochastic nature of our tracking algorithm, the exact value of this distance is not required. Given the poses of the hips at the beginning and the end of a step, values of the hips between these two frames are estimated. In fact, the distance is correlated to the angle of the step trajectory in non-linear manner as shown on Figure 1a. We approximate this correlation function using a function which models a S-curve.

$$D(\theta) = D_0 \cdot \frac{1 - e^{-\alpha\theta}}{1 + e^{-\alpha\theta}} \quad (2)$$

where  $D_0$  is the maximum size of the hip distance with respect to the size of the leg (in our implementation, it is half the value of the sum of the thigh widths),  $\theta$  is the angle between the trajectory and the  $x$  axis in the image plane and  $\alpha$  is an empirical factor which controls the speed of the curve descent.

Therefore, we infer hip distances by estimating at each frame the trajectory angle. This is performed by fitting cubic splines to all pivot points (see figure 1a,b).

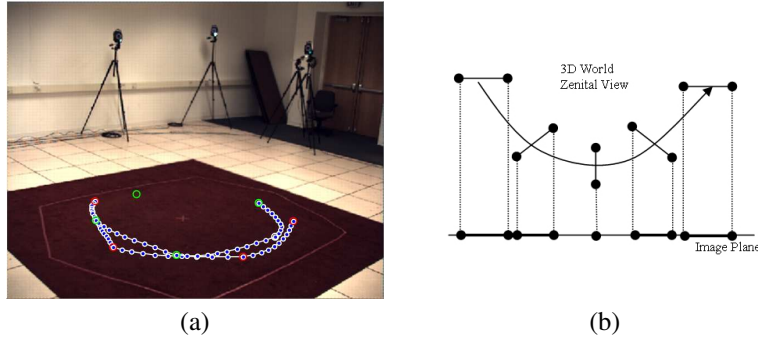


Figure 1: (a) Interpolated trajectory (blue dots) of the pivot points (red and green dots). (b) Correspondences during a turn between the hip distances in a zenithal view and in the image plane.

### 3.3.2 Multiple particle filter tracking

One of the most challenging problems of 2D tracking is to deal with the perspective effect which amplifies changes in trajectories and, therefore, can create major variations in the target's size. Therefore, the usage of a simple first order models does not allow representing size dynamics adequately. Since our tracking framework is based on a full 'step' where heel strike positions are known, the final position of a step is partially reinitialised. Consequently, information is available to define the trajectory of the target during each step. Moreover, new tracking constraints are derived regarding maximum and minimum apparent limb sizes and distances between the hip points during the step. This last constraint provides a reference point for the 'swing' leg similarly as the pivot point restricts the location of the 'support' leg. These new constraints, which were not initially available when the standard tracker operated, would reduce significantly the complexity of the tracking problem. Furthermore, when using a particle filter based tracker, the probability of divergence increases after each prediction: the closer a frame is to the initialisation frame the more accurate is the estimation is likely to be.

In order to take advantages of these new constraints and tackle this inherent tracker weakness, we propose that once the standard tracker has processed a full 'step', two new trackers should be launched in parallel (see Figure 2b). These trackers have the same configuration and dynamical models enhanced by the constraints extracted from the output of the standard tracker. Whereas the 'forward' tracker starts from the first frame of the step, the 'backward' tracker begins at the last frame and tracks targets backwards. Consequently, for each frame two estimates are available. However, since estimate reliability depends on its distance from the initialisation frame, a criterion is designed to decide at which frame the backward tracker is more likely to provide more accurate estimates than the forward tracker. Although the particle filter does not provide for each frame an actual estimation, a weighted mean estimation is extracted combining all the hypotheses. By using this temporal estimation, the measurement of its likelihood function is obtained.

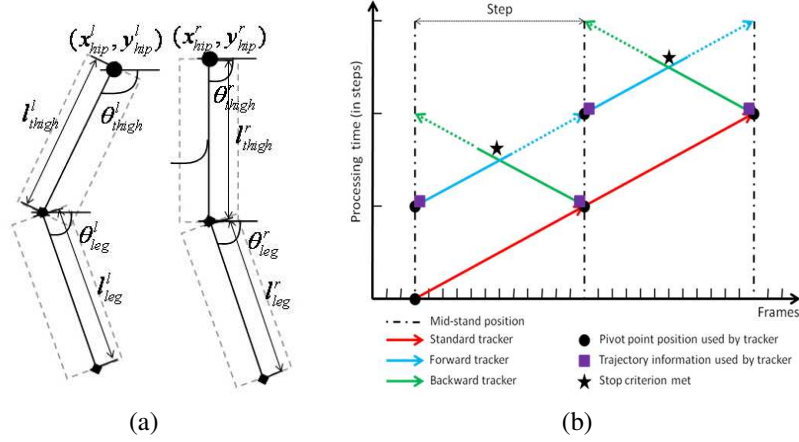


Figure 2: (a) 2D articulated model. (b) Multi tracker framework.

Therefore, comparison between the forward and backward tracking is made, and a stopping criterion is established. The reduction of reliability is introduced as an exponential decreasing function that multiplies the estimation likelihood.

In our framework, the likelihood function relies on colour and edge features. Since we assume these features are independent from each other, we can combine them to obtain the observation probability:

$$p(z_t|x_t) = p(z_t^1|x_t) \cdot p(z_t^2|x_t) \quad (3)$$

where  $z_t^1$  and  $z_t^2$  are the colour and edge observations respectively.

Colour features are obtained by sampling each region by a grid and expressing the colour information by RGB values subsampled to 4 bits to filter out noise and small variations. The colour density is measured by comparing the colour feature of each region of the articulated model with its corresponding colour model. It is evaluated by estimating the Bhattacharyya coefficient between their histograms.

$$p(z_t^1|x_t) = \prod_{r \in X_t} \left( \sum_{h=1}^H \sqrt{r(h) \cdot q(h)} \right)^{\alpha_c} \quad (4)$$

where  $r$  is a body part from the articulated model,  $H$  are all the histogram bins,  $r(h)$  is the current histogram,  $q(h)$  is the reference model and  $\alpha_c > 0$  is an empirical factor to increase the discriminative power of the feature.

A gradient detector is used to detect edges, and the result is thresholded to eliminate spurious edges. The Canny algorithm is applied for this purpose. The result is smoothed with a Gaussian filter and normalised between 0 and 1. The resulting density image  $P^e$  assigns a value to each pixel according to its proximity to an edge using an adaptation of the Euclidean distance transform.

$$p(z_t^2|x_t) = \prod_{r \in X_t} \frac{1}{N} \sum_{i=1}^N P^e(I, i)^{\alpha_e} \quad (5)$$

where  $r$  represents each of the regions which compose the articulated model,  $N$  are all the pixels which compose the region and  $\alpha_e > 0$  is an empirical factor similar to  $\alpha_c$ .

## 4 Results

The algorithm has been tested over the HumanEva (HE) dataset I and II. As motion capture and video data were collected synchronously, motion capture data provides the groundtruth. Since cameras are calibrated, ground truth data points can be projected on the 2D sequences in order to evaluate quantitatively 2D pose estimates. A standard set of error metrics [5] is defined that can be used for evaluation of both pose estimations and tracking algorithms. To ensure fair comparison between algorithms that use different numbers of parts, only predicted joints are included in the error metric.

The algorithm has been tested with 3 sequences from these datasets: *S2\_Walking\_1\_(C1)*, *S2\_Combo\_1\_(C1)* from HE I and *S2\_Combo\_1\_(C1)* from HE II. Since the latest sequence is especially long, we divided it in two parts: part 1, which is at the beginning of the sequence and corresponds to walking, and part 2, which is at the end and shows some balancing, see Table 1 for details. These sequences were chosen to include a variety of movements (walking a complete circle and balancing) seen from different points of view and happening mainly outside the camera plane (see Figures 4 and 5). Although experiments were performed with a number of particles in the Particle Filter ranging from 200 to 500, their number did not affect tracking accuracy.

The pivot point detector can produce erroneous locations: the average error is 20.4 pixels. Therefore, tracking can be affected negatively by this initial process. To analyse independently the tracking algorithm, results are also provided where manual annotation was used to define pivot points (see Table 1). The mean error is around 15 pixels. It increases to 18 pixel when tracking is combined with automatic pivot point detection.

Figure 3 depicts the error of our tracking framework compared with a version using only the standard tracker. Not only does our system perform significantly better, but this chart also highlights one of the strength of our proposition: tracking is able to recover from serious divergence because of the partial reinitialisation provided by detection of pivot points and trajectory constraints. Although the tracker diverges around frame 200, where limbs reach their apparent maximum size and are self-occluded, legs are accurately labelled on frames 219 and 242 (see Figures 3 and 4).

Table 1: Numerical results with manual and automatic pivot point detection

Sequence (C1 camera)	Frames	Manual pivot point		Automatic pivot point	
		Absolute Mean Error [in pixels]	Standard Deviation [in pixels]	Absolute Mean Error [in pixels]	Standard Deviation [in pixels]
S2_Walking1, HE I	[6, 418]	17.1	8.7	25.9	10.9
S2_Combo1, HE I	[1661, 2054]	9.3	5.8	9.1	6.2
S2_Combo1, HE II	[1, 307]	25.1	12.4	25.8	12.3
S2_Combo1, HE II	[747, 1202]	9.7	2.3	10.0	3.1
Total	1570	<b>14.6</b>	<b>9.9</b>	<b>17.6</b>	<b>14.7</b>

Table 2 shows how our results compare with other techniques used to recover either 2D or 3D poses from the HumanEva data sets. When authors only provided mean errors for 3D poses, they were converted in pixels using approximate relationships between pixel

Table 2: Comparison with state of the art

Algorithm	Dataset	Pixel error	Constraints	Training	Initialised
Manual pivot	HE I	13.2	Bipedal motion	No	Yes
	HE II	15.9	Bipedal motion	No	Yes
Automatic pivot	HE I	17.5	Bipedal motion	No	Yes
	HE II	17.7	Bipedal motion	No	Yes
Lee et al. [10]	HE I	5-7*	Activity specific & cyclic	Yes	No
Howe [7, 8]	HE I	12.5	Activity specific	Yes	No
	HE II	18.5	Activity specific	Yes	No
Pope et al. [9]	HE I	10-14*	View and activity specific	Yes	No
	HE II	17-20*	View and activity specific	Yes	No
Husz et al. [11]	HE I	33	Single calibrated camera	Yes	Yes
	HE I	14.8	Multiple calibrated cameras	Yes	Yes
	HE II	19	Multiple calibrated cameras	Yes	Yes

\* Pixel error estimated from 3D error

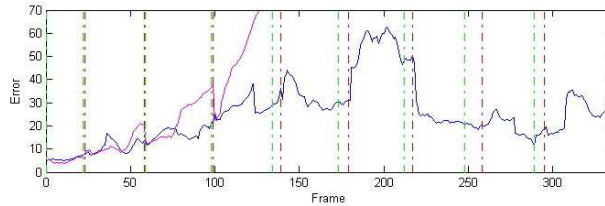


Figure 3: Tracking error for each frame of the part 1 of S2\_Combo.1\_(C1) (HE II) sequence. Green and red lines are respectively the manual and automatic detection of the beginning/end of a 'step'. Magenta line is the error using only the standard tracker. Blue line shows the error using the whole framework.

and object lengths for each of the HumanEva datasets.

Most methods perform similarly on the HumanEva datasets, i.e. a pixel error in the 12-15 and 17-20 ranges for respectively HE I and HE II. The only exception is Lee and Elgammal's [10] work which relies on a manifold whose topology is learned using a training set. Their technique performs extremely well: joint mean accuracy is 31 mm, i.e. 5 to 7 pixels. The main drawback of their approach is it relies on a walking scenario or more generally on cyclic activities that have to be learnt explicitly.

The hierarchical particle filter proposed by Husz et al. [11] relies on a motion model based on action primitives which predicts the next pose in a stochastic manner. Although their tracker performs similarly to ours when 2 or more camera sequences are available, its performances degrade significantly when processing a single sequence. Both Howe [7, 8] and Poppe et al. [9] present example-based approaches to pose recovery. They use very different image descriptors, respectively silhouettes and histograms of oriented gradients, but their results appear to be quite similar. The main drawback of these methods is they



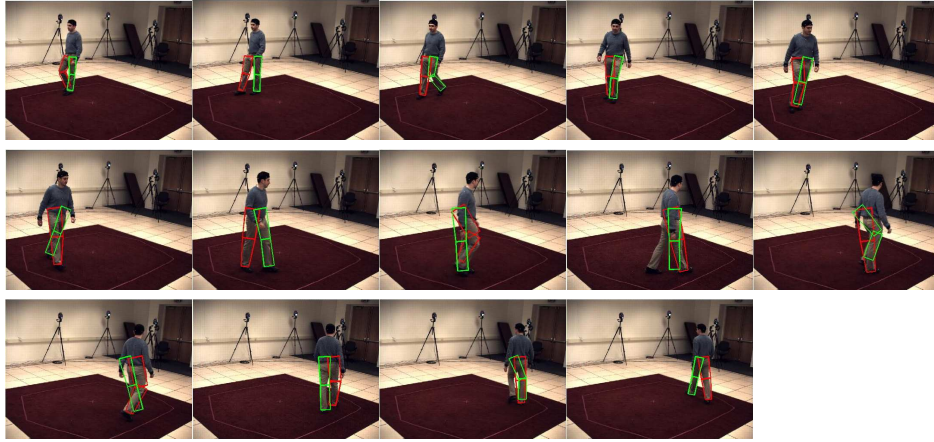


Figure 4: Results for *S2\_Combo\_1\_(C1)* (HumanEva II) sequence. Frames: 1, 10, 29, 45, 81, 95, 125, 154, 175, 194, 219, 242, 272, 290.

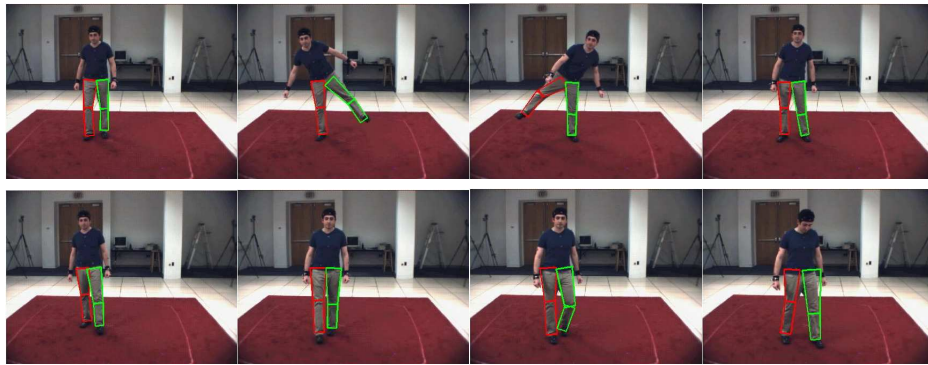


Figure 5: Results for *S2\_Combo\_1\_(C1)* (HumanEva I) sequence. Frames: 1661, 1730, 1853, 1920, 1967, 2021, 2045, 2073

are action specific and therefore they may not be able to track individuals which display either unexpected motions or a combination of motions. It is also important to notice that like many 3D algorithms, Poppe's assume that the location of the silhouette in the image is provided by an auxiliary image tracker. Therefore, the results they present do not take into account the lack of accuracy such a tracker would produce. One strength of our methodology is that all processing steps are fully integrated. Finally, since our framework is based on a generative approach, no training phase is required. Thus, our system can recover human poses of unusual movements as shown in Figure 5.

## 5 Conclusion and Future Work

This paper introduces a novel framework based on a set of particle filters to track human body parts from a single camera. The results presented here demonstrate the feasibility

of recovering human pose using a 2D limb tracking system on the basis of articulated models constrained only by human biomechanics. Not only does the use of a 2D model reduce the computation complexity of tracking human body parts, but also simplifies the tracker initialization. The presented approach has been successfully applied to walking and balancing sequences which include changes of view in the 3D world. Moreover, its accuracy is comparable to other systems relying on constraints incompatible with most video surveillance applications. Therefore, our framework yields potential for tracking human body segments in those applications where motions are generally bipedal.

In future work, we want to tackle the initialisation of our tracker. We will build on the already existing methods which have been proposed to detect automatically limbs from an individual either from still or sequence images. In particular, we will focus on bottom up strategies which have the advantage of not relying on a training stage. We will also extend our framework so that tracking can be performed when short ‘pivot’ point occlusions and temporary loss of foot contact, e.g. in running, occur.

## References

- [1] Y. Wu, G. Hua, and T. Yu, “Tracking Articulated Body by Dynamic Markov Network”, in *ICCV*, pp. 1094-1101 (2003).
- [2] G. McAllister, S.J. McKenna, and I.W. Ricketts, “MLESAC-Based Tracking with 2D Revolute-Prismatic Articulated Models”, in *ICPR*, vol. 2, pp. 725-728, Quebec (2002).
- [3] P. Noriega, and O. Bernier, “Multicues 2D Articulated Pose Tracking Using Particle Filtering and Belief Propagation on Factor Graphs”, in *ICIP*, vol. 5(2), pp. 725-728 (2007).
- [4] I. Bouchrika, and M. S. Nixon, “People detection and recognition using gait for automated visual surveillance”, in *IET Conf. on Crime and Security*, pp. 576-581 (2006).
- [5] L. Sigal and M. J. Black, “HumanEva: Synchronized video and motion capture dataset for evaluation of articulated human motion”, in *Technical Report CS-06-08, Brown Univ.* (2006).
- [6] L. Sigal and M. J. Black, “Predicting 3D People from 2D Pictures”, in *AMDO* (2006).
- [7] N. R. Howe, “Evaluating Lookup-Based Monocular Human Pose Tracking on the HumanEva Test Data”, in *Workshop on Evaluation of Articulated Human Motion and Pose Estimation (EHuM2)*, (2007).
- [8] N. R. Howe, “Recognition-Based Motion Capture and the HumanEva II Test Data”, in *Workshop on Evaluation of Articulated Human Motion and Pose Estimation (EHuM2)*, (2007).
- [9] R. Poppe, “Evaluating Example-based Pose Estimation: Experiments on the HumanEva Sets”, in *Workshop on Evaluation of Articulated Human Motion and Pose Estimation (EHuM2)*, (2007).
- [10] C.S. Lee, and A. Elgammal, “Body pose tracking from uncalibrated camera using supervised manifold learning”, in *Workshop on Evaluation of Articulated Human Motion and Pose Estimation (EHuM)*, Whistler, Canada, (2006).
- [11] Z. L. Husz, A. M. Wallace, and P. R. Green, “Evaluation of a Hierarchical Partitioned Particle Filter with Action Primitives”, in *Workshop on Evaluation of Articulated Human Motion and Pose Estimation (EHuM2)*, (2007).
- [12] J. Deutscher, and I. D. Reid, “Articulated body motion capture by stochastic search”, in *Int. Jour. of Computer Vision*, vol. 61(2), pp. 185-205 (2005).
- [13] P. F. Felzenszwalb and D. P. Huttenlocher, “Pictorial structures for object recognition”, in *Int. Jour. of Computer Vision*, vol. 61(2), pp. 55-79 (2005).
- [14] M. W. Lee and I. Cohen, “Human body tracking with auxiliary measurements”, in *IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures (AMFG)*, pp. 112-119 (2003).
- [15] J. MacCormick and M. Isard, “Partitioned sampling, articulated objects, and interface-quality hand tracking”, in *ECCV*, vol. 2, pp. 3-19 (2000).
- [16] C. Sminchisescu and B. Triggs, “Covariance scaled sampling for monocular 3D body tracking”, in *CVPR*, vol. 1, pp. 447-454 (2001).
- [17] M. Isard and A. Blake, “Condensation: conditional density propagation for visual tracking”, in *Int. Jour. of computer vision*, 29 (1) pp. 5-28 (1998).
- [18] J. Deutscher, A. Blake and Ian Reid, “Articulated Body Motion Capture by Annealed Particle Filtering”, in *CVPR* vol. 2, pp. 126-133 (2000).
- [19] K. Choo and D. Fleet, “People Tracking Using Hybrid Monte Carlo Filtering”, in *ICCV*, pp. 321-328 (2001).
- [20] C. M. Fryer, “Biomechanics of the lower extremity”, in *Instruct Course Lect*, vol. 20, pp. 124-130 (1971).
- [21] S. X. Ju, M. J. Black, and Y. Yacoob, “Cardboard People: A Parameterized Model of Articulated Image Motion”, in *Int. Conf. on Automatic Face and Gesture Recognition*, pp. 38-44 (1996).