

# Articulated Model Based People Tracking Using Motion Models

Huazhong Ning, Liang Wang, Weiming Hu and Tieniu Tan  
National Laboratory of Pattern Recognition

Institute of Automation, Chinese Academy of Sciences, Beijing, P. R. China, 100080  
{hzning, lwang, wmhu, tnt}@nlpr.ia.ac.cn

## Abstract

*This paper focuses on acquisition of human motion data such as joint angles and velocity for applications of virtual reality, using both articulated body model and motion model in the CONDENSATION framework. Firstly, we learn a motion model represented by Gaussian distributions, and explore motion constraints by considering the dependency of motion parameters and represent them as conditional distributions. Then both of them are integrated into the dynamic model to concentrate factored sampling in the areas of state-space with most posterior information. To measure the observing density with accuracy and robustness, a PEF (Pose Evaluation Function) modeled with a radial term is proposed. We also address the issue of automatic acquisition of initial model posture and recovery from severe failures. A large number of experiments on several persons demonstrate that our approach works well.*

## 1 Introduction

In virtual reality, temporal data of human motions are required for animation. But previously, such data were obtained by hand, which was very boring and lacked accuracy. Human location and tracking using computer vision will eliminate these difficulties. Currently, the main approaches to locating and tracking people can be divided into two categories: contour based and articulated model based [1, 2]. Although very fast and sometime real-time, contour based methods [3, 4] cannot recover high-level information such as joint angles because only the region information is considered. In contrast, articulated model based approaches, often using motion models, can achieve this robustly and accurately. Also, such approaches can be used in human-machine interactions. Here we will mainly concentrate on locating and tracking people walking parallel to image plane in monocular image sequences.

Due to the complex nature of human body, tracking human in video sequences is a very difficult task and involves a number of hard issues such as detection, occlusion and high dimensions. To alleviate these difficulties, many human body models and motion models were introduced as priors in previous work. As far as human body models are concerned, they vary widely in

their levels of detail, ranging from stick figure model [6], cylinder model [7, 8], truncated-cone model [9, 10] to super-quadratics model [11], and recently to hierarchical human model [12]. As a general rule, the more complex the human body model, the more accurate tracking results may be expected but at the expense of higher computational complexity.

Also, motion models of body limbs and joints were widely used in the tracking process. They serve as prior knowledge to predict motion parameters [5, 14], to interpret and recognize human dynamics [15], or to constrain the estimation of low-level image measurements [13]. For instance, Bregler [15] decomposed human dynamics into multiple abstractions, and represented the high-level abstraction by HMM (Hidden Markov Model) as successive phases of simple movements. Then it was used for tracking and recognition. Zhao [5] learnt a highly structured motion model similar to a FSM (Finite State Machine). The popular method, MPCA (Multivariate Principal Component Analysis), was also used to train a walking model in Sidenbladh et al [13].

In this paper, we present an approach to tracking human based on both body model and motion model in the CONDENSATION framework [17]. As a trade-off of accuracy and complexity, the human body model is represented by articulated truncated cones and a sphere, and its degrees of freedom are reduced from more than 40 to 12 under the assumption of walking parallel to the image plane (see our previous work [19] for more information). Additionally, a motion model that is definitely different from that in Sidenbladh et al [13] is learnt from semi-automatically acquired training data. Using the motion model, we explore the dependency between shoulder and elbow joint, thigh and knee joint to discover and describe the motion constraints. These constraints, together with the motion model, are integrated into the dynamic model to concentrate the factored sampling areas. To compute the observation density, we use the PEF that combines both boundary and region information to make it accurate and robust, and model it with a radial term to improve the efficiency of factored sampling. Also, the tracker can be automatically initialized and recovered from severe failures.

There are two main contributions in our paper. One is the learnt motion model and motion constraints that

enforce our dynamic model for CONDENSATION algorithm and result in low computational cost. The other is the automatic initialization process that is usually done manually in previous work [5, 13, 8]. Cheng [14] also provided an initialization method that searches the entire motion model to locate the first frame by finding the dominant peak of a cost function. However, this approach evaluates the cost function for many times, in turn leading to high computational cost. In contrast, our method is much faster due to the spatio-temporal information that avoids computing the PEF.

## 2 Motion Model and Motion Constraints

A motion model, encoding much information about the dynamics of the human body, can be used to greatly reduce the computational cost while achieving better results. As a highly constrained activity, the patterns of human walking are symmetric, periodical and of little variation in a wide range of people, according to Murray [16]. So it is relatively easy to learn a compact and effective motion model for human walking from limited training data. In this paper, our own motion model is learnt from semi-automatically acquired training data and formulated as Gaussian distributions. Also the dependency of joint angles is analyzed to explore the motion constraints that, together with the motion model, are integrated into the dynamical model to focus on the heavy weighted samples in the CONDENSATION framework.

### 2.1 Learning Motion Model

In the learning process, training data ( $m = 9$  examples from 5 different subjects) were semi-automatically acquired by purposefully designed software with friendly interface. The data correspond to the 12-dimensional posture vector  $P = \{x, y, \mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_{10}\}$  where  $(x, y)$  indicates the global position and  $\mathbf{q}_i$  is the  $i$ th joint angle [19]. Several feature points in each frame are marked manually and the motion parameter derived from these features are computed and analyzed automatically. Some data are illustrated in Figure 1(a) which reveal that the temporal curves have different periods and phases. Therefore the walking cycles in each example must be rescaled to the same length and aligned with the same phase before training.

We form matrix  $A_i$  for each training example  $i$  with row index indicating time step and column index indicating motion parameters, in which  $i = 1 \dots m$ . Then the period  $T_i$  of the example  $i$  is computed from the cross-correlation function  $\text{corr}(A_i, A_i)$ . To rescale the walking cycle to the same length  $T$  (in this paper  $T = 100$ ), the B-spline interpolation algorithm is applied to the example  $A_i$  with the scalar  $a_i = T/T_i$ . Given that  $A_i$  is rescaled to  $A_i$ , a specific one from  $A_i$  ( $i = 1 \dots m$ ), e.g.  $A_1$ , is

selected as the reference. Then the phase  $b_i$  of each example  $A_i$  relative to the reference example  $A_1$  is indicated by the predominant peak in the cross correlation function  $\text{corr}(A_i, A_1)$ . In all,

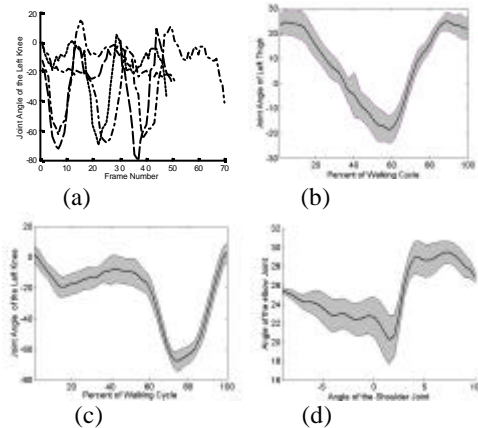
$$B_i(t) = A_i(a_i \times t + b_i), \quad i = 1 \dots m \quad (1)$$

are the normalized examples with the same period and phase. The segments  $B_i(1:T)$ ,  $B_i(T+1:2T)$ , ...,  $i = 1 \dots m$ , renamed as  $W_j$  with  $j = 1 \dots n$ , are exactly all of the normalized walking cycles. Then our motion model is represented by Gaussian distributions  $G_{i,t}(u_{i,t}, \mathbf{s}_{i,t})$  for each joint angle  $i$  ( $i = 1 \dots 10$ ) at any phase  $t$  ( $t = 1 \dots T$ ) in the walking cycle with

$$u_{i,t} = \frac{1}{n} \sum_{k=1}^n W_k(t, i), \quad k = 1 \dots 10, t = 1 \dots T \quad (2)$$

$$\mathbf{s}_{i,t} = \sqrt{\frac{1}{n} \sum_{k=1}^n (W_k(t, i) - u_{i,t})^2}, \quad k = 1 \dots 10, t = 1 \dots T \quad (3)$$

Figure 1(b) and (c) are temporal models of joint angles of left thigh and left knee. Although learnt from limited data, they correspond very well to Murray's results of the medical analysis [16]. In late sections, the motion model is used to estimate the prior distribution of initial pose and to predict new pose in tracking.



**Figure 1.** Motion model and motion constraints. (a) Joint angles of the left knee of 4 different people walking with various periods and phases. (b) and (c) Temporal models of joint angles of left thigh and left knee during a walking cycle. (d) Motion constraints of the elbow joint. In (b), (c) and (d) the dark line and the shaded areas indicate the mean and standard deviation of the corresponding distribution.

### 2.2 Motion Constraints

Obviously, in a walking activity the movements of the lower arm and the upper arm are correlated and regular, so the shoulder joint and the elbow joint are not independent. We assume that the lower arm is driven by the upper arm, and accordingly the elbow joint is determined by the shoulder joint except for some noise. So the motion constraint of the elbow joint can be approximated by the conditional distribution  $p(\mathbf{q}_e | \mathbf{q}_s)$  where  $\mathbf{q}_e$  and  $\mathbf{q}_s$

are the joint angles of the elbow and the shoulder respectively. Using the training data in the previous subsection, the distribution can be easily computed by the following procedure. From each walking cycle  $W_i$  ( $i = 1 \dots n$ ), a series of pairs of the shoulder and elbow joint angles  $(\mathbf{q}_{is}(t), \mathbf{q}_{ie}(t))$  are formed as the time  $t$  varies from 1 to  $T$ . We classify all pairs according to their first element, i.e. pairs having identical first element are assigned to the same class. Then for any shoulder joint angle  $\mathbf{q}_s$ , provided that class  $(\mathbf{q}_s, \bullet)$  includes  $K$  pairs  $(\mathbf{q}_s, \mathbf{q}_e^k)$ ,  $k = 1 \dots K$ , the conditional distribution  $p(\mathbf{q}_e | \mathbf{q}_s)$  is represented by Gaussian distribution  $G(\boldsymbol{\mu}, \boldsymbol{\Sigma}^2)$  where  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  are the mean and standard deviation of  $\mathbf{q}_e^k$ ,  $k = 1 \dots K$ . Figure 1(d) gives the motion constraint for the elbow joint. Similarly, the motion constraints for the knee and ankle joint are learnt in the same way. We also derive intervals of valid value for each motion parameter from training data by specifying its maximal and minimal value. All of the samples are constrained in their associated intervals by setting the exceeding samples to its minimum or maximum.

### 3 Tracking

The main task here is to relate the image data to the posture vector  $P = \{x, y, \mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_{10}\}$  defined in Section 2.1. We decompose the task into several sub-tasks, viz. locating the global position and tracking each limb separately. Such decomposition is strongly supported by the experiments.

In our approach, we firstly predict the global position from the centroid of the detected moving human and then refine it by searching entirely from the neighborhood of the predicted position. Each limb is tracked under the CONDENSATION framework [17]. As a popular method in visual tracking, the CONDENSATION algorithm uses learnt dynamical models, together with visual observations, to propagate the random sample set over time. The rule of state density propagation over time is:

$$p(x_t | Z_t) = k_t p(z_t | x_t) \int_{x_{t-1}} p(x_t | x_{t-1}) p(x_{t-1} | Z_{t-1}) dx_{t-1} \quad (4)$$

where  $x_t$  are the motion parameters at time  $t$ ,  $Z_t = (z_1, z_2, \dots, z_t)$  is the image sequence up to time  $t$ , and  $k_t$  is a normalization constant independent on  $x_t$ . According to this rule, the posterior distribution of  $p(x_t | Z_t)$  can be derived from the posterior  $p(x_{t-1} | Z_{t-1})$  at the previous time step and three other premises: the prior distribution  $p(x_0)$  at time 0, i.e. the initialization process; the dynamical model  $p(x_t | x_{t-1})$  to predict the motion parameters  $x_t$  by drifting and diffusing  $x_{t-1}$ ; observation density  $p(z_t | x_t)$  computed by the PEF. They are detailed in the following subsections.

### 3.1 Initialization

Unlike previous work on initialization that attempted to roughly estimate the posture from a single frame, we accomplish the initialization using spatio-temporal information of the first  $N$  frames. Thus our approach is more robust, and most importantly, it also achieves real-time speed by avoiding evaluating the cost function. In what follows we consider the initialization process that includes a learning process and an estimation process.

In the learning process, the moving human in each frame in the training data ( $m = 9$  examples from 5 different subjects) is detected by subtracting the background image and extracted edges using Sobel operator, and then the moving area is clipped and normalized to the same size. Similar to the preprocessing of learning motion model, the normalized examples are adjusted to the same phase and their periods are rescaled to the same length. Also  $n$  normalized walking cycles  $V_j$  with  $j = 1 \dots n$ , are segmented from the  $m$  preprocessed examples. We use the average walking cycle

$$V = \frac{1}{n} \sum_{j=1}^n V_j \quad (5)$$

as the reference cycle. In the estimation process, the first  $N$  frames  $v$ , after detection, edge extraction and normalization as that in the learning process, is located in the reference cycle by searching the major peak in the correlation function  $\text{corr}(V, v)$ . Referring the location (assumed to be  $t$ ) to the motion model, the posture of the last frame in  $v$  is roughly estimated as the 10-dimensional vector  $(u_{3,t}, u_{4,t}, \dots, u_{12,t})$ . Accordingly, the prior distribution for tracking is the Gaussian distribution  $G((u_{3,t}, u_{4,t}, \dots, u_{12,t}), \boldsymbol{\Sigma}_t, I_{10})$  where  $I_{10}$  is an  $10 \times 10$  identity matrix, and  $\boldsymbol{\Sigma}_t = (\boldsymbol{\Sigma}_{3,t}, \boldsymbol{\Sigma}_{4,t}, \dots, \boldsymbol{\Sigma}_{10,t})$ .

The initialization process can also be used to recover from severe tracking failures probably caused by occlusion, accumulated error, or image noise. When a failure occurs, the tracker will stop for  $N$  frames and reinitialize using the spatio-temporal information derived from such  $N$  frames to estimate the current posture. However, it is worthwhile to mention that the real-time speed and robustness of the initialization and bootstrap are achieved at the expense of the first  $N-1$  frames where tracking is stopped.

### 3.2 Dynamic Model

The dynamic model is often exquisitely designed to improve the efficiency of factored sampling. The idea is to concentrate the samples in the areas of state-space containing most information about the posterior. The desired effect is to avoid as far as possible generating samples that have low weights, since they contribute little to the posterior. In this paper, the learnt motion model

served as prior is integrated into the dynamic model to achieve efficiency of sampling. In detail, at any time-step  $t$  the  $i$ th motion parameter  $\mathbf{q}_{i,t}$  satisfies the dynamic model

$$p(\mathbf{q}_{i,t} | \mathbf{q}_{i,t-1}) = G(\mathbf{a}u_{i,t} + \mathbf{b}u_{i,t-1} + \mathbf{g}\mathbf{q}_{i,t-1}, \mathbf{I}(\mathbf{s}_{i,t} + \mathbf{s}_{i,t-1})) \quad (6)$$

where  $G$  is Gaussian distribution, and  $\mathbf{a} + \mathbf{b} + \mathbf{g} = 1$  makes the drifting of  $\mathbf{q}_{i,t}$  not only from the history  $\mathbf{q}_{i,t-1}$  but also from the motion model, and  $\mathbf{I}$  is a scalar that is often set to 1. But when the walking of the tracked person is very normal, a smaller  $\mathbf{I}$  is expected to restrict the sampling more effectively to portions of the parameter space that are most likely to correspond to human motion.

This dynamic model is generally sufficient for all motion parameters, but motion constraints can further concentrate the samples for motion parameters: elbow, knee and ankle joint. For instance, after the shoulder joint  $\mathbf{q}_{s,t}$  is sampled, sample positions generated from the conditional distribution  $p(\mathbf{q}_{e,t} | \mathbf{q}_{s,t})$  (see Section 2.2) for the elbow joint  $\mathbf{q}_{e,t}$  also contain much information. So a mixed-state CONDENSATION [18] can be included in the factored sampling scheme by choosing with probability  $q$  to generate samples from the dynamic model and with probability  $1-q$  to generate samples from the conditional distribution  $p(\mathbf{q}_{e,t} | \mathbf{q}_{s,t})$ , i.e.  $\mathbf{q}_{e,t}$  satisfies the dynamic model

$$p(\mathbf{q}_{e,t} | \mathbf{q}_{e,t-1}, \mathbf{q}_{s,t}) = q G(\mathbf{a}u_{e,t} + \mathbf{b}u_{e,t-1} + \mathbf{g}\mathbf{q}_{e,t-1}, \mathbf{I}(\mathbf{s}_{e,t} + \mathbf{s}_{e,t-1})) + (1-q)p(\mathbf{q}_{e,t} | \mathbf{q}_{s,t}) \quad (7)$$

where  $\mathbf{a}, \mathbf{b}, \mathbf{g}, \mathbf{I}$  are defined as above. Equations similar to (7) can also be provided for knee and ankle joints.

### 3.3 Pose Evaluation Function

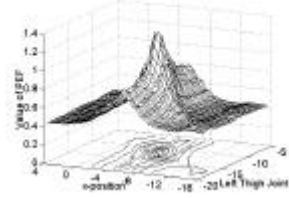
The PEF reveals the observation density  $p(z_t | x_t)$  of the image  $z_t$  given that the human model has posture  $x_t$  at time  $t$ . To match the image  $z_t$  with the generative model, the model must be projected into the image plane. Furthermore detection based on background subtraction and Sobel operator are applied successively to the image  $z_t$  to acquire both region and boundary information. Then the information is employed to compute both the boundary match error  $E_b$  and the region match error  $E_r$  (see [19] for details).

By introducing a factor  $\mathbf{a}$  to adjust their weights, both boundary and region match errors are combined into the PEF to achieve both accuracy and robustness. The PEF is also modeled with a robust radial term  $r_i(s, \mathbf{s}) = v e^{-s^2}$  [11]:

$$PEF(E) = v e^{-(\mathbf{a} \times E_b + (1-\mathbf{a}) \times E_r) / s^2} \quad (8)$$

Apart from its robustness, the radial term can also improve the efficiency of factored sampling because it

assigns heavier weights to important samples and severely reduces the weights of insignificant ones. It is worthwhile to mention that we often select a bigger  $\mathbf{a}$  for the upper limbs to minish the influence of region match error. The reason is that the upper limbs and the torso often have clothes with the same texture and they also frequently occlude each other, and therefore the region information is a little unimportant.



**Figure 2.** The curve of the PEF with the global position  $x$  and the joint angle of the left thigh changing smoothly and other parameters remaining constant. Also shown is the contour of the function. The curve is shifted up by 0.4 to make the contour clear.

Figure 2 shows the effectiveness of our PEF (8). Its curve is basically smooth and has no local maxima at the neighborhood of the global maximum. Furthermore, according to the contour of the PEF and other experiments, we can conclude that the global position  $(x, y)$  is much more significant than other joint angles with respect to the PEF. This is one of the reasons why the global position can be firstly determined with other parameters unchanged.

## 4 Experiments and Discussion

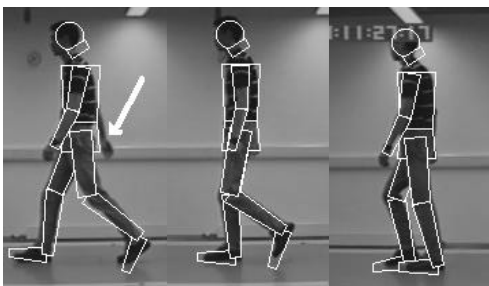
We conduct experiments of tracking on several persons having various shapes and walking characteristics in gray image sequences with low quality and significant self-occlusion. These sequences, selected from the SOTON gait database, were captured by a stationary camera at a rate of 25 frames per second, and the original resolution of each image is  $384 \times 288$  pixels. Our approach is implemented with the MATLAB on a personal desktop workstation using the first 15 frames of each sequence to automatically initialize the prior distribution. The tracked sequences are selected both from the training data and from new instances. For each sequence from the new instances the implementation takes approximately 7.5 minutes/frame with 300 state samples for the upper limbs and 100 state samples for the lower limbs. Comparatively, due to the accuracy of the motion model, each sequence from the training data only requires 100 and 50 state samples for the upper and lower limbs respectively, consuming a little more than 3 minutes/frame.

### 4.1 Tracking Results

Started automatically by the initialization process

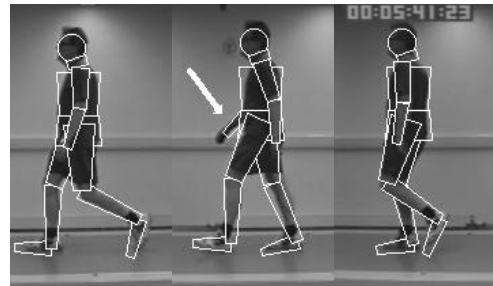
described in Section 3.1, the program mostly tracks successfully in the entire image sequences with people walking parallel to the camera plane, although sometimes stopped and reinitialized for severe failures perhaps caused by occlusion, accumulated errors, or drastic image noise. Here a sequence from the training data (see Figure 3) and a new instance (see Figure 4) are showed as tracking results. Due to the space constraint, only the human areas clipped from the original image sequences are shown. The most difficult part of the data, which verifies the effectiveness of our approach, is that the sequences include the configuration in which the two legs and thighs occlude each other severely (e.g. frame 53 in Figure 4), causing most part of one leg or thigh is unseen. Other challenges include: shadow under the feet; the arm and the torso have the same color; various colors and styles of clothes; different shapes of tracked people; and low quality of the image sequences. It is worthwhile to mention that sometimes the arms far from the camera in these sequences were lost for the severe occlusion by the torso (see frame 19 in Figure 3). However, their motion parameters can usually be properly estimated using the motion model (see frame 27 in Figure 4) or using the symmetric value of the other arms.

Also, further experiments are carried out to deeply analyze our approach. It is mentioned that tracking sequences from the training data requires much less state samples than tracking new examples. The chief reason is that our motion model is learnt from limited training data and cannot accurately represent the variations of all sequences, especially of abnormal ones. When encountered a novel instance, the deficiency of the motion model will reduce the accuracy of the prediction of the dynamic model. Fortunately, increasing the state samples will compensate it. This is demonstrated in an experiment. In Figure 5(a), when the sequence is included in the training data, the prediction is extremely close to the refined results. This closeness, which also proves the effectiveness of the dynamic model, requires a small set of samples. Contrastively, in Figure 5(b) when the same sequence is intentionally removed from the training data, the prediction is not very good but the larger sample set offsets the inaccuracy.

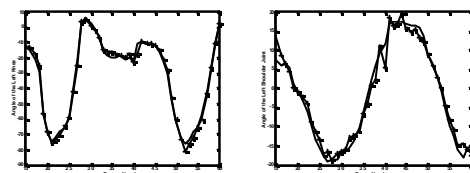


**Figure 3.** Tracking persons in training data. Frame 19, 23 and 43 in sequence 2 of subject 3 (vin2).

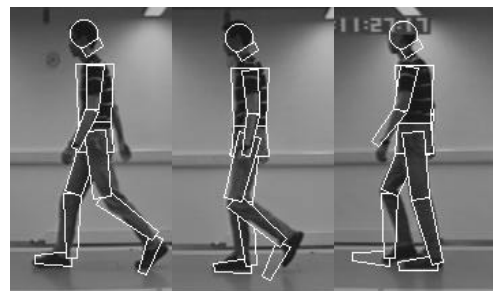
Then a question arises: what a role does the motion model play in the tracking? We track a sequence without sampling, so that the motion parameters are entirely estimated from the motion model. The tracking results illustrated in Figure 6 reveal that, although roughly true, the results often deviate in detail in contrast to that in Figure 3. Therefore the motion model does not unduly affect the tracking.



**Figure 4.** Tracking persons not in training data. Frame 19, 27 and 53 in sequence 1 of subject 1 (dh1).



**Figure 5.** Estimation results: Predicted results are in bold lines and refined results by factored sampling are in thin lines with markers. (a) The angle of the left knee of dh1 that is included in the training data. (b) The angle of the left shoulder joint of the same sequence that is removed from the training data.



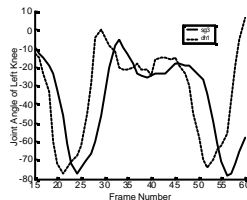
**Figure 6.** Analyzing the significance of the motion model. Tracking results of frame 19, 23 and 43 in sequence 2 of subject 3 (vin2) without sampling.

An important improvement of our experiments is the small size of the sample set (300 for each upper limb and 100 for each lower limb) comparing to 500 in [13] and 512 in [5], but with good tracking results. The effect is due partly to the effectiveness of our dynamic model, and partly to the accurate and robust PEF modeled with a radial term. Of course, further experiments are needed to determine the minimal required number of the sample set in order to make the computational cost as low as possible.

However, run time analysis reveals that most of the time is spent on evaluating the PEF. So a PEF with little computational cost will contribute more to the high speed of a fast tracker.

## 4.2 Human Motion Data and Synthesis

Our purpose of the motion model based tracking is to acquire such high-level information of walking as joint angles and velocity for virtual reality. Figure 7 gives some human motion data obtained from the tracking results. With these data, we can synthesize the walking process with parameterized 3D articulated-models. The synthetic process, which can be inspected from any viewing angle, gives us a more vivid impression of walking. Figure 8 provides a synthetic sequence of walking.



**Figure 7.** Human motion data. Temporal curve of joint angle of the left knee.



**Figure 8.** Synthetic walking corresponding to the sequence1 of subject1 (dh1).

## Acknowledgements

The authors would like to thank Dr. M. Nixon and Dr. C. Yam from University of Southampton, U.K, for their help with the SOTON gait database. This work is supported by NSFC (Grant No. 69825105 and 60105002) and Institute of Automation (Grant No. IM01J02), Chinese Academy of Sciences.

## References

- [1] D. Gavrilu, the Visual Analysis of Human Movement: a Survey, *Computer Vision and Image Understanding* 73 (1), pp. 82-98, 1999.
- [2] J. Aggarwal and Q. Cai, Human Motion Analysis: a Review, *Computer Vision and Image Understanding*, 73 (3), pp. 428-440, 1999.
- [3] K Toyama and A Blake, Probabilistic Tracking in a Metric Space, in *Proc. of International Conference on Computer*

- Vision*, pp. 50-57, 2001.
- [4] A Baumberg and D. Hogg, An Efficient Method for Contour Tracking Using Active Shape Models, in *Proc. of IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, pp. 194-199, 1994.
- [5] T. Zhao, T.S. Wang and H.Y. Shum, Learning a Highly Structured Motion Model for 3D Human Tracking, in *Proc. of 5<sup>th</sup> Asian Conference on Computer Vision*, Melbourne, Australia, 2002.
- [6] H.J. Lee and Z. Chen, Determination of 3D Human Body Posture from a Single View, *Computer Vision, Graphics, Image Processing*, 30, pp. 148-168, 1985.
- [7] D. Hogg, Model-based Vision: A Program to See a Walking Person, *Image and Vision Computing*, 1(1), pp. 5-20, 1983.
- [8] S. Wachter and H. H. Nagel, Tracking Persons in Monocular Image Sequences, *Computer Vision and Image Understanding*, 74(3), pp. 174-192, 1999.
- [9] Q. Delamarre and O. Faugeras, 3D Articulated Models and Multi-View Tracking with Physical Forces, *Computer Vision and Image Understanding*, 81, pp. 328-357, 2001.
- [10] Q. Delamarre and O. Faugeras, 3D Articulated Models and Multi-View Tracking with Silhouettes, in *Proc. of 7<sup>th</sup> International Conference on Computer Vision*, Kerkyra, Greece, 1999.
- [11] C. Sminchisescu and B. Triggs, Covariance Scaled Sampling for Monocular 3D Body Tracking, in *Proc. of International Conference on Computer Vision and Pattern Recognition*, Kauai, HI, 2001.
- [12] R. Plankers and P. Fua, Articulated Soft Objects for Video-based Body Modeling, in *Proc. of 9<sup>th</sup> International Conference on Computer Vision*, Vancouver, Canada, 2001.
- [13] H. Sidenbladh, M. Black and David Fleet, Stochastic Tracking of 3D Human Figures Using 2D Image Motion, in *Proc. of European Conference on Computer Vision*, 2000.
- [14] J.C. Cheng, and J.M.F. Moura, Capture and Representation of Human Walking in Live Video Sequence, *IEEE Transactions on Multimedia*, 1(2), pp. 144-156, 1999.
- [15] C. Bregler, Learning and Recognizing Human Dynamics in Video Sequences, In *Proc. IEEE Conference Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, 1997.
- [16] M. Murray, Gait as a Total Pattern of Movement, *American Journal of Physical Medicine*, 46, pp. 290-333, 1967.
- [17] M. Isard and A. Blake, CONDENSATION – Conditional Density Propagation for Visual Tracking, *International Journal of Computer Vision*, 29(1), pp. 5-28, 1998.
- [18] M. Isard and A. Blake, A Mixed-state Condensation Tracker with Automatic Model Switching, in *Proc. of International Conference on Computer Vision*, pp. 107-112, 1998.
- [19] H. Ning, L. Wang, W. Hu and T. Tan, Model-based Tracking of Human Walking in Monocular Image Sequences, in *IEEE Region 10 Technical Conference on Computers, Communication, Control and Power Engineering*, 2002.