Signal Processing 110 (2015) 164–177

Contents lists available at ScienceDirect

Signal Processing

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

3D Human motion tracking by exemplar-based conditional particle filter

Jigang Liu [a,*], Dongquan Liu [b], Justin Dauwels [c], Hock Soon Seah [a]

[a] School of Computer Engineering, Nanyang Technological University, Singapore
[b] Singapore Polytechnic, Singapore
[c] School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore

ARTICLE INFO

Article history:
Received 4 May 2014
Received in revised form
13 August 2014
Accepted 16 August 2014
Available online 27 August 2014

Keywords:
3D Human motion tracking
Particle filter
Monocular camera tracking
Motion exemplar

ABSTRACT

3D human motion tracking has received increasing attention in recent years due to its broad applications. Among various 3D human motion tracking methods, the particle filter is regarded as one of the most effective algorithms. However, there are still several limitations of current particle filter approaches such as low prediction accuracy and sensitivity to discontinuous motion caused by low frame rate or sudden change of human motion velocity. Targeting such problems, this paper presents a full-body human motion tracking system by proposing exemplar-based conditional particle filter (EC-PF) for monocular camera. By introducing a conditional term with respect to exemplars and image data, dynamic model is approximated and used to predict current states of particles in prediction phase. In update phase, weights of particles are refined by matching images with projected human model using a set of features.

This method retains advantages of classic particle filters while increases prediction accuracy by replacing the smooth motion model with exemplars-based dynamic model which constrains evolved particles within an area closer to true state. Therefore, tracking robustness to discontinuous motion is improved such as under conditions of sudden change in motion velocity or using low-frame rate cameras. To verify the effectiveness and efficiency of the proposed algorithm, a variety of datasets are selected for testing and the results are also compared with the state-of-the-art methods in this domain.

1. Introduction

3D human motion tracking is a process in which the configuration of body parts are estimated from one or several sensor inputs [1]. It has received great attention in the last two decades due to its wide applicability in many areas, such as surveillance, virtual reality, medical analysis, computer animation and human–computer interaction [2,3]. Although a large amount of research on human tracking has been carried out, the task of 3D human motion tracking from monocular image is still challenging due to following reasons. First, human pose has high degree of freedom leading to heavy computations of searching in high-dimensional state space. Second, there are a number of issues significantly affecting tracking results need to be addressed, such as imaging conditions, features extracted from images, occlusions, velocity changes, etc. Currently, most of commonly-used particle filter-based approaches concentrate on developing motion prior that allows efficient prediction in high-dimensional pose state. However, the effectiveness of such motion priors, which are derived

* Corresponding author. Tel.: +65 92342149.
E-mail addresses: liujg@ntu.edu.sg (J. Liu), liu_dongquan@sp.edu.sg (D. Liu), jdauwels@ntu.edu.sg (J. Dauwels), ASHSSEAH@ntu.edu.sg (H.S. Seah).

http://dx.doi.org/10.1016/j.sigpro.2014.08.028
0165-1684/

from learning statistical models of captured human motion data or from simple dynamics models, remains an open problem. As a result, many current existing particle filter-based methods suffer from inaccurate human pose prediction and sensitivity to discontinuous motions, especially under conditions of using low-frame rate cameras or existing of sudden changes in human motion. This issue is even more serious when a simple dynamic model which assumes continuous and smooth human motion is applied in prediction. When discontinuous or abrupt motion occurs, the prediction accuracy of methods based on such models will significantly drop which results in either low tracking accuracy or large searching computation as a compensation.

In this paper, a new method named exemplar-based conditional particle filter (EC-PF) is proposed for 3D human motion tracking. For EC-PF, system state is constructed to be conditional to image data and exemplars in order to improve prediction accuracy. Current frame can be thought of as a snapshot of the current human pose when human motion is captured. It contains accurate current human pose information comparing with temporal coherence [4] and learnt dynamics model [5]. In this proposed approach, a small amount of exemplars with prior knowledge, such as joint locations and relative depth information between two joints, are created and stored beforehand. In prediction step, 3D human pose estimation [6,7] is conducted by using shape context matching [8,9] with exemplars. A dynamic model from previous frame to current frame is then built based on the human pose information from current and previous frame in order to predict particles' current states. In the update step, a set of features extracted from current frame are used as observations to update particles' weights. Since smooth or pre-learnt motion model [10,11] used by classic particle filters is replaced by an exemplar-based dynamic model here, particles can evolve within an area closer to true state even in the case of using low-frame rate cameras or existing of sudden velocity changes in human motion. Experimental results in Section 5 demonstrate the robustness of the proposed method to discontinuous motions.

To sum up, the contributions of this paper include

1. With theoretical deduction, EC-PF is proposed by conditioning system state with images and exemplars in order to improve prediction accuracy.
2. An exemplars-based dynamic model construction via shape context matching is introduced to effectively estimate three dimensional pose with a monocular camera setup.
3. A full-body human motion tracking system for monocular camera is realized based on proposed EC-PF. By comparison, its advantage in robustness to abrupt motion velocity change and low frame rates is verified.

This paper is organized as follows. In Section 2, we briefly review related works. In Section 3, classic particle filters and EC-PF are described, and the differences between these two filters are discussed. Section 4 presents the whole human motion tracking process by applying proposed EC-PF. In detail, 3D human model is first introduced, followed by description on pose estimation method and likelihood measurement. Experimental results are provided in Section 5. Section 6 concludes this paper.

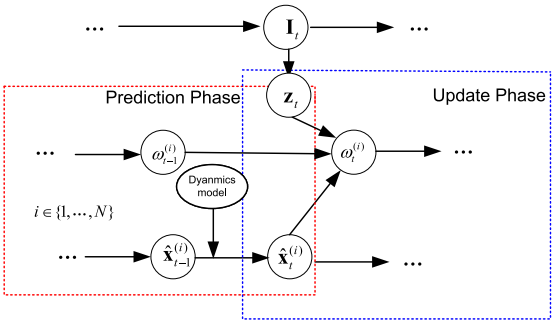

Fig. 1. The process of classic particle filters.

2. Related works

There has been a large amount of work on human pose tracking in the last two decades. There are mainly two categories in human motion tracking: machine learning methods and object tracking methods [12]. In the first approach, researchers proposed the use of machine learning methods that exploit prior knowledge in gaining more stable estimates of 3D human body pose [13–15]. However, these algorithms require a large amount of samples which limit their applications. Object tracking methods [16,17] commonly follow two sequential steps: human pose features are extracted and tracked in each frame, then human

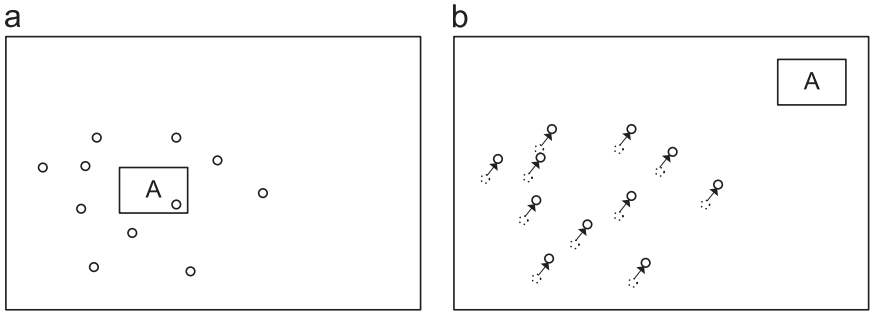

Fig. 2. Challenges in classic particle filters. From time t to time $t+1$, object A experiences abrupt motion: (a) Object A with particles denoted by circles with solid lines at time t. (b). Object A and particles at time $t+1$. Particles at time t are denoted by circles with dotted lines. Particles at time $t+1$ are denoted by circles with solid lines.

Table 1
The pseudo-code of EC-PF.

For time steps $t=0,1,2,\ldots$

1. Initialization at time 0: for $i=1,\ldots,N_p$, sample $\mathbf{x}_0^{(i)} \sim p(\mathbf{x}_0)$, $\omega_0^{(i)} = \frac{1}{N_p}$
2. Estimation of prediction function $f_{exemplars,\mathbf{I}_{1:t}}$.
   Compare current frame $\mathbf{I}_t$ with exemplars to estimate current system state $\widehat{\mathbf{x}}_{matched,t}$, calculate transition function $f_{exemplars,\mathbf{I}_{1:k}}$ according to $\widehat{\mathbf{x}}_{matched,t} = f_{exemplars,\mathbf{I}_{1:t}}(\widehat{\mathbf{x}}_{t-1})$
3. Importance sampling. For $i=1,\ldots,N_p$, draw samples $\widetilde{\mathbf{x}}_t^{(i)} \sim p(\mathbf{x}_t|\mathbf{x}_{t-1}^{(i)}, exemplars, \mathbf{I}_{1:t})$.
4. Weight update. For $i=1,\ldots,N_p$, update the importance weights $\omega_t^{(i)} = \omega_{t-1}^{(i)} p(\mathbf{z}_t|\widetilde{\mathbf{x}}_t^{(i)}, exemplars, \mathbf{I}_{1:t})$.
5. Normalize the importance weights: $\tilde{\omega}_t^{(i)} = \frac{\omega_t^{(i)}}{(1/N_p)\sum_{j=1}^{N_p}\omega_t^{(j)}}$.
6. Compute the optimal results: $\widehat{\mathbf{x}}_t = \sum_{i=1}^{N_p} \widetilde{\mathbf{x}}_t^{(i)} \tilde{\omega}_t^{(i)}$
7. Resampling if necessary.
8. Repeat Steps 2–7.

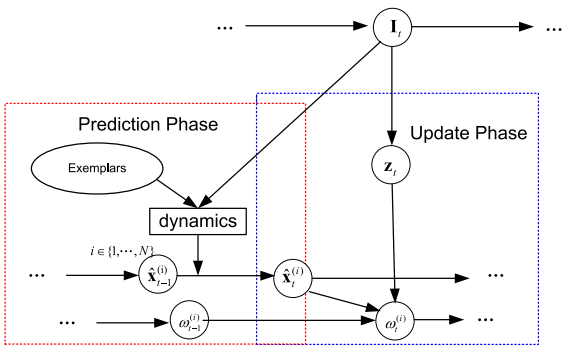

Fig. 3. The process of EC-PF.

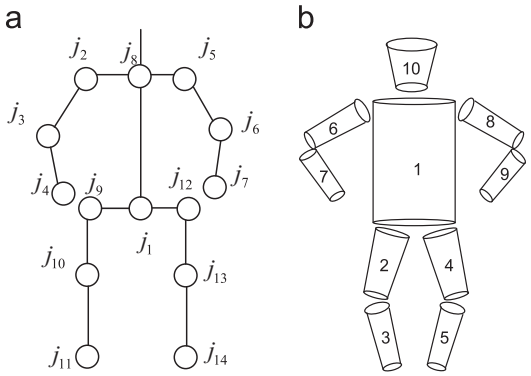

Fig. 4. 3D human model (a) kinematic chain and (b) 3D human body model.

pose is reconstructed from the obtained features. Many researchers have conducted studies on the first step where people usually use the configuration in current frame and a dynamic model to predict the next configuration [18]. However, 3D human motion tracking from monocular video sequences is still a challenging task due to self-occlusion and depth ambiguities. Some restrictions or prior knowledge may be required beforehand such as in [19,20]. Also, to recover the parameters of a human pose correctly from video sequences is a difficult task due to the high degree of freedom in human body configurations.

Particle filter [21], which is able to track non-linear motion, can solve problems related to complex human motion. Generally, a number of particles are sampled using a dynamic model, including a noise component. Each particle is assigned with an associated weight which is updated according to a likelihood measurement function. The pose estimation is obtained by the weighted sum of all particles. Aiming to overcome the high dimensionality of system states and sample impoverishment [22], many methods have been proposed to make the human pose more tractable. The first group of methods is to use priors on human movement. This kind of methods can be recognized as learning motion models to guide particle prediction effectively in prediction step. Raskin [23] proposed the Gaussian Process Annealed Particle Filter (GPAPF) to track 3D human motion. GPAPF combines the annealed particle filter (APF) [24] with the Gaussian process dynamic model (GPDM) [25] to reduce the dimensionality of state vector. This method improves the tracker's performance and increases its stability and ability to recover from losing the target. In [26], walking dynamics was learnt from a training set to predict the human pose. These kinds of learning motion models need a large number of training datasets and a complex training process; otherwise the derived dynamics is impossible to correctly describe human motion in the real world due to the complexity of human motion.

Covariance scaled sampling (CSS) is introduced to guide particles by inflating the posterior covariance of previous frame in [27]. This method focuses on particles in regions where there is uncertainty, such as depth ambiguities in monocular tracking. Deutscher et al. [24,28,29] proposed APF to track 3D human motion with three calibrated cameras. They use simulated annealing to make particles locate on the global maxima of the posterior at the expense of multiple iterations per frame. Particles are initially sampled widely, and then their range of movement is decreased gradually over time. These methods spread particles in a large area at initialization or enlarge posterior uncertainty to track the human motion more accurately. Therefore, the computation burden becomes heavy, which results in slow processing.

In [30], a dynamical simulation prior was proposed based on the truth of physical ground–person interaction. This prior is able to constraint human motion in the range of

Table 2
Lengths of the segments of human model.

Segment	Length (m)
Lower arm	0.25
Upper arm	0.25
Thigh	0.43
Calf	0.40
Torso	0.46
Head	0.30

physical plausibility. Chang and Lin [31] proposed a progressive particle filter to decrease computational cost by employing hierarchical searching. This hierarchical searching approach decomposes high-dimensional space into several lower-dimensional spaces. In this method, there are some restrictions on initial human pose.

Smooth or pre-learnt dynamic models are used in the above methods. Once abrupt motion of human object occurs between previous frame and current frame, this kind of dynamic models cannot describe human motion correctly; these methods may fail to track human motion.

3. Classic particle filters and exemplar-based conditional particle filter

In this section, the concept of classic particle filters is first described, as well as its limitations. Targeting such limitations, EC-PF is introduced. The theoretical deduction and working process of EC-PF are followed.

3.1. Classic particle filters

A particle filter is a recursive process that estimates the posterior probability from a set of weighted particles. Let $\mathbf{x}_t$ and $\mathbf{z}_t$ denote the state vector and the observation at time step t respectively. The history of observations from time step 1 to t is expressed as $\mathbf{Z}_t=\{\mathbf{z}_1,...,\mathbf{z}_t\}$. $\hat{\mathbf{x}}$ is the state estimation. N_p weighted particles at time t are expressed as $\{\mathbf{x}_t^{(i)},\omega_t^{(i)},\ i=1,...,N_p\}$. $\mathbf{x}_t^{(i)}$ represents the state of ith particle at time step t, its corresponding weight $\omega_t^{(i)}$. Classic particle filters contain two major steps, namely, prediction and update. The details of each step are described as follows.

3.1.1. State prediction

Firstly, sample a new set of particles by choosing the particles with the highest posterior probabilities $p(\mathbf{x}_{t-1}^{(i)}|\mathbf{z}_{t-1})$ among the previous particle set at time step $t-1$. Then, assume that the pdf $p(\mathbf{x}_{t-1}|\mathbf{Z}_{t-1})$ is available at time step $t-1$, system state is predicted according to

$$p(\mathbf{x}_t|\mathbf{Z}_{t-1})=\int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{Z}_{t-1})d\mathbf{x}_{t-1}$$

3.1.2. State update

Compute the posterior $p(\mathbf{x}_t|\mathbf{Z}_t)$, weight current particles by the predicted prior probability using the above equation, and apply observation $\mathbf{z}_t$ to estimate the likelihood probability $p(\mathbf{z}_t|\mathbf{x}_t)$. The posterior $p(\mathbf{x}_t|\mathbf{Z}_t)$ can be expressed in Bayesian form as

$$p(\mathbf{x}_t|\mathbf{Z}_t)=\frac{p(\mathbf{z}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{Z}_{t-1})}{p(\mathbf{z}_t|\mathbf{Z}_{t-1})} \tag{1}$$

where $p(\mathbf{z}_t|\mathbf{Z}_{t-1})$ denotes normalizing constant in the denominator as

$$p(\mathbf{z}_t|\mathbf{Z}_{t-1})=\int p(\mathbf{z}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{Z}_{t-1})d\mathbf{x}_t \tag{2}$$

The normalized weight $\omega_t^{(i)}$ is regarded as the posterior probability

$$\omega_t^{(i)}\propto p(\mathbf{z}_t|\hat{\mathbf{x}}_t^{(i)}),\quad \sum_{i=1}^{N_p}\omega_t^{(i)}=1 \tag{3}$$

Finally, the mean state at time t can be estimated as

$$\hat{\mathbf{x}}_t=\sum_{i=1}^{N_p}\omega_t^{(i)}\hat{\mathbf{x}}_t^{(i)}$$

Classic particle filter is a popular non-linear filter. Its process is shown in Fig. 1, there are two main phases, prediction phase and update phase. In prediction phase, each particle $\hat{\mathbf{x}}_t^{(i)}$ is diffused based on predefined dynamics model. In update phase, observation $\mathbf{z}_t$ is extracted from current frame $\mathbf{I}_t$ and used to refine particles' weights.

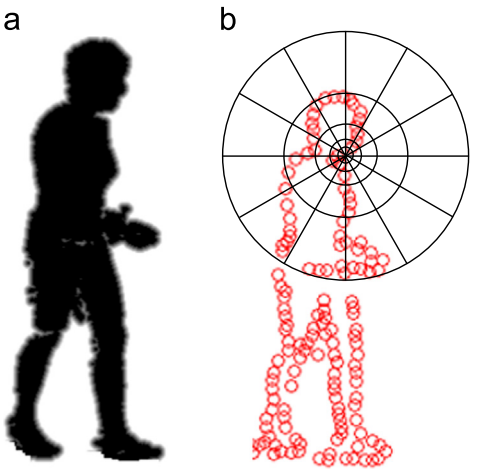

Fig. 5. Examples of shape contexts: (a) input image and (b) sampled edge point and example log-polar histogram bins.

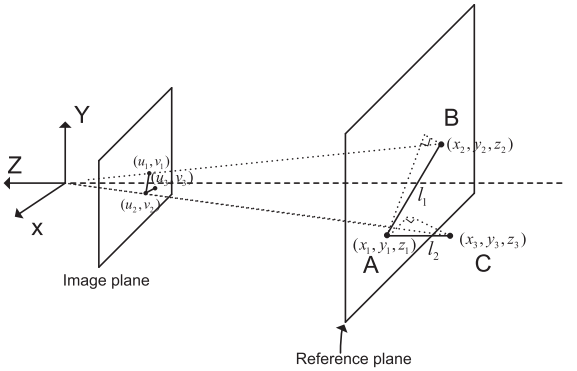

Fig. 6. The projections from world coordinate system onto image plane.

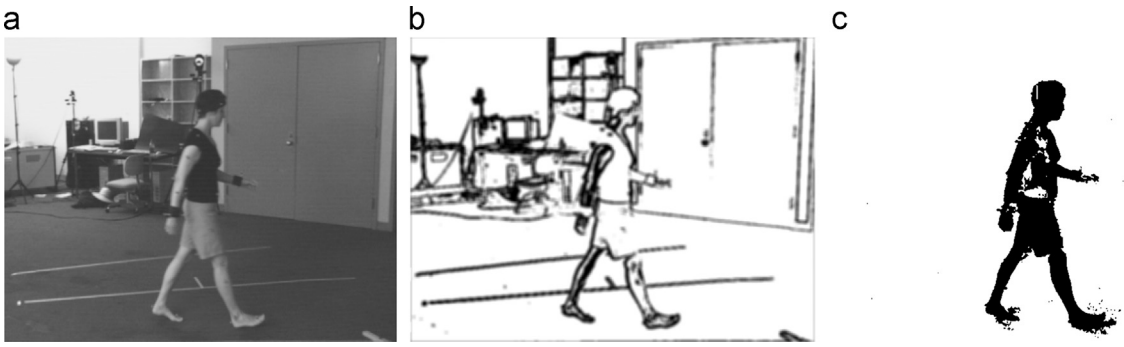

Fig. 7. Example features: (a) original image, (b) edge feature and (c) silhouette feature.

Table 3
The pseudo-code of proposed 3D human motion tracking method.

Initialization: $t=0$

Generate N_p particles with corresponding weights $\{\mathbf{x}_0^{(i)}, \omega_0^{(i)}, i=1,...,N_p\}$ sample

$\mathbf{x}_0^{(i)} \sim p(\mathbf{x}_0)$, $\omega_0^{(i)} = \frac{1}{N_p}$

for $t=1$–k

Prediction

Shape contexts feature is first extracted from current frame and to be compared with exemplars. Calculate current 3D human pose $\hat{\mathbf{x}}_{matched,t}$ according to the method in Section 4.2.

Compute transition function $f_{exemplars,\mathbf{I}_{1:t}}$ based on $\hat{\mathbf{x}}_{matched,t}$ and previous state $\hat{\mathbf{x}}_{t-1}$

Predict each particle state according to $f_{exemplars,\mathbf{I}_{1:t}}$

$\hat{\mathbf{x}}_t^{(i)} = f_{exemplars,\mathbf{I}_{1:t}}(\hat{\mathbf{x}}_{t-1}^{(i)})$

Update

Edge and silhouette information of human body are extracted from current frame and denoted as $\mathbf{Z}_{e,t}$ and $\mathbf{Z}_{s,t}$.

Calculate weights of each particle according to likelihood measurement function.

$\omega_t^{(i)} = \omega_{t-1}^{(i)} p(\mathbf{z}_t|\widehat{\mathbf{x}}_t^{(i)}, exemplars, \mathbf{I}_{1:t})$

Normalize the weights $\omega_t^{(i)} = \frac{\tilde{\omega}_t^{(i)}}{(1/N_p)\sum_{j=1}^{N_p}\tilde{\omega}_t^{(j)}}$

Output

Estimate the optimal states at time t as

$\hat{\mathbf{x}}_t = \sum_{i=1}^{N_p} \hat{\mathbf{x}}_t^{(i)} \omega_t^{(i)}$

Resample N_p particles if necessary.

end

As an example, one challenging case for classic particle filters is shown in Fig. 2. A classic particle filter is applied to track object A by using 11 particles denoted by circles. If object A experiences abrupt motion from time step t to time step $t+1$, particles' states predicted by smooth dynamic model would not be around true state of object A (see Fig. 2(b)). Finally, the classic particle filter fails to track object A.

3.2. EC-PF

Classic stochastic filters such as Kalman filter, extended Kalman filter and particle filters are constructed for dynamical systems discretized in the time domain. A smooth dynamic model is always chosen as a prior used for state prediction in these filters. Estimated states by these filters may diverge from true states if smooth dynamic model does not describe object's motion correctly. However such smooth dynamic model cannot describe complex motions in the real world. In order to solve this problem, in this paper EC-PF is proposed by conditioning system state with image data and exemplars.

The working mechanism of EC-PF is based on the following two equations:

State equation $\mathbf{x}_t = f_{exemplars,\mathbf{I}_{1:t}}(\mathbf{x}_{t-1}) + \mathbf{w}_t$

Measurement equation $\mathbf{z}_t = h(\mathbf{x}_t) + \mathbf{v}_t$

where $f_{exemplars,\mathbf{I}_{1:t}}$ is the function for predicting current state from previous state based on exemplars and image data. h computes observation from system state. $\mathbf{w}_t$ and $\mathbf{v}_t$ represent process noise and observation noise respectively.

The details are given as follows; the posterior distribution is empirically represented by a weighted sum of N_p samples drawn from the posterior distribution

$$p(\mathbf{x}_t|\mathbf{Z}_t, exemplars, \mathbf{I}_{1:t}) \approx \frac{1}{N_p}\sum_{i=1}^{N_p} \delta(\mathbf{x}_t - \mathbf{x}_t^{(i)}) \equiv \hat{p}(\mathbf{x}_t|\mathbf{Z}_t, exemplars, \mathbf{I}_{1:t}) \tag{4}$$

where $\delta(x) = \begin{cases} 0, & x \neq 0 \\ +\infty, & x = 0 \end{cases}$ and $\int_{-\infty}^{+\infty} \delta(x)dx = 1$, $\mathbf{x}_t^{(i)}$ are assumed to be identically and independently distributed (i.i.d.) particles drawn from $p(\mathbf{x}_t|\mathbf{Z}_t, exemplars, \mathbf{I}_{1:t})$. N_p is the number of particles. When N_p is sufficiently large,

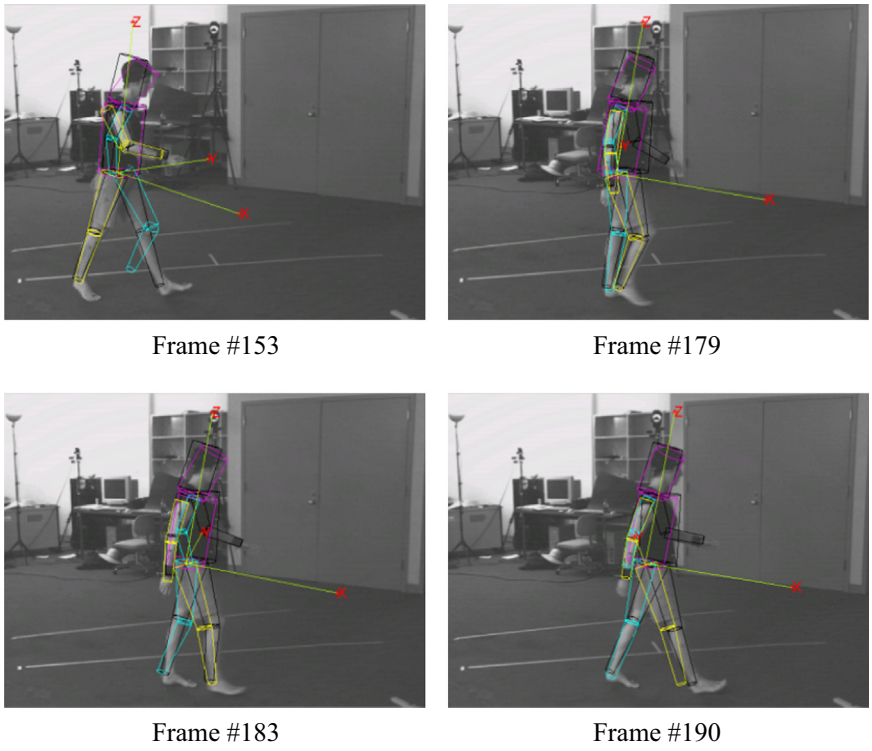

Fig. 8. Tracking results by EC-PF at Frame #153, #179, #183 and #190 under the condition of sudden change in human velocity.

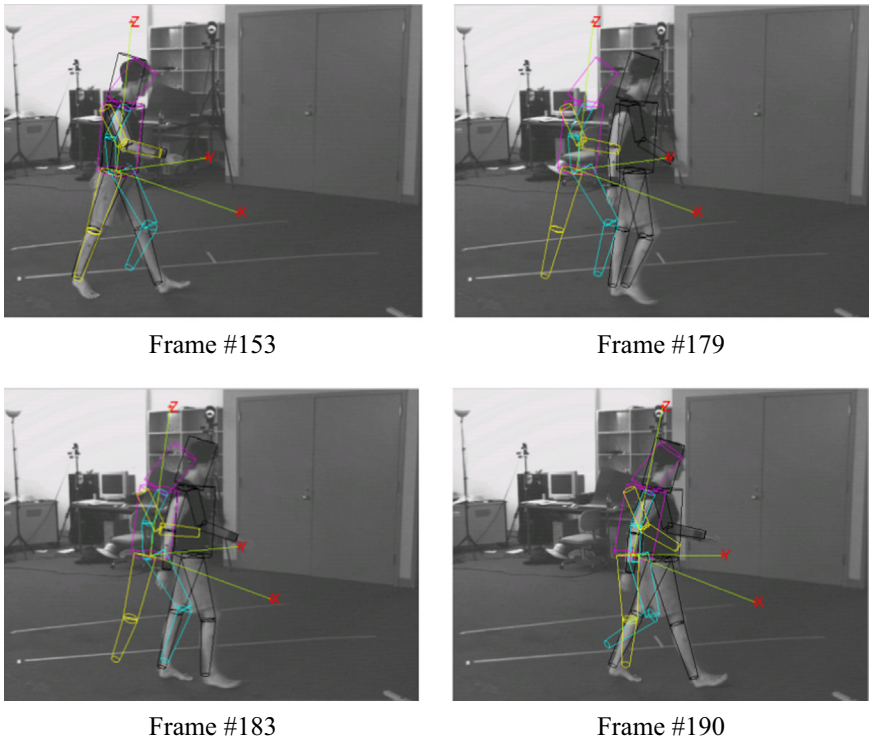

Fig. 9. Tracking results by annealed particle filter at Frame #153, #179, #183 and #190 under the condition of sudden change in human velocity.

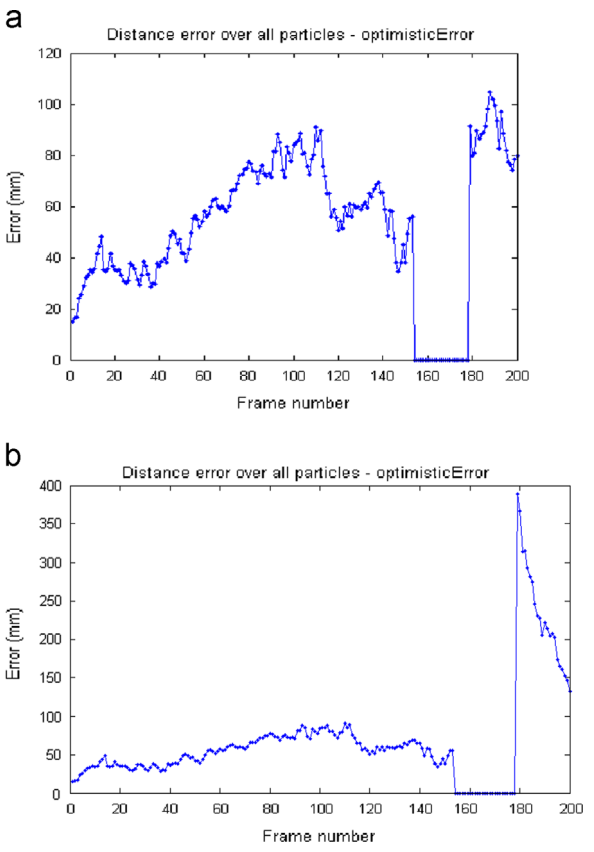

Fig. 10. Tracking error under the condition of sudden change in human velocity: (a) tracking error by EC-PF and (b) tracking error by annealed particle filter.

$\hat{p}(\mathbf{x}_t|\mathbf{Z}_t, exemplars, \mathbf{I}_{1:t})$ approximates the true posterior $p(\mathbf{x}_t|\mathbf{Z}_t, exemplars, \mathbf{I}_{1:t})$.

Since it is usually impossible to sample from the true position, it is common to sample from an easy-to-implement distribution, the so-called proposal distribution denoted by $q(\mathbf{x}_t|\mathbf{Z}_t, exemplars, \mathbf{I}_{1:t})$. Hence, we can estimate the mean of system state

$$\begin{aligned} E[\mathbf{x}_t] &= \int \mathbf{x}_t \frac{p(\mathbf{x}_t|\mathbf{Z}_t, exemplars, \mathbf{I}_{1:t})}{q(\mathbf{x}_t|\mathbf{Z}_t, exemplars, \mathbf{I}_{1:t})} q(\mathbf{x}_t|\mathbf{Z}_t, exemplars, \mathbf{I}_{1:t}) d\mathbf{x}_t \\ &= \int \mathbf{x}_t \frac{\omega_t(\mathbf{x}_t)}{p(\mathbf{Z}_t|exemplars, \mathbf{I}_{1:t})} q(\mathbf{x}_t|\mathbf{Z}_t, exemplars, \mathbf{I}_{1:t}) d\mathbf{x}_t \\ &= \frac{1}{p(\mathbf{Z}_t|exemplars, \mathbf{I}_{1:t})} \int \mathbf{x}_t \omega_t(\mathbf{x}_t) q(\mathbf{x}_t|\mathbf{Z}_t, exemplars, \mathbf{I}_{1:t}) d\mathbf{x}_t \end{aligned} \tag{5}$$

where

$$\omega_t(\mathbf{x}_t) = \frac{p(\mathbf{Z}_t|\mathbf{x}_t, exemplars, \mathbf{I}_{1:t}) p(\mathbf{x}_t|exemplars, \mathbf{I}_{1:t})}{q(\mathbf{x}_t|\mathbf{Z}_t, exemplars, \mathbf{I}_{1:t})}$$

Eq. (5) can be rewritten as

$$\begin{aligned} E[\mathbf{x}_t] &= \frac{\int \mathbf{x}_t \omega_t(\mathbf{x}_t) q(\mathbf{x}_t|\mathbf{Z}_t, exemplars, \mathbf{I}_{1:t}) d\mathbf{x}_t}{\int p(\mathbf{Z}_t|\mathbf{x}_t, exemplars, \mathbf{I}_{1:t}) p(\mathbf{x}_t|exemplars, \mathbf{I}_{1:t}) d\mathbf{x}_t} \\ &= \frac{\int \mathbf{x}_t \omega_t(\mathbf{x}_t) q(\mathbf{x}_t|\mathbf{Z}_t, exemplars, \mathbf{I}_{1:t}) d\mathbf{x}_t}{\int \omega_t(\mathbf{x}_t) q(\mathbf{x}_t|\mathbf{Z}_t, exemplars, \mathbf{I}_{1:t}) d\mathbf{x}_t} \\ &= \frac{E_{q(\mathbf{x}_t|\mathbf{Z}_t, exemplars, \mathbf{I}_{1:t})}[\mathbf{x}_t \omega_t(\mathbf{x}_t)]}{E_{q(\mathbf{x}_t|\mathbf{Z}_t, exemplars, \mathbf{I}_{1:t})}[\omega_t(\mathbf{x}_t)]} \end{aligned} \tag{6}$$

By drawing the i.i.d. particles $\{\mathbf{x}_t^{(i)}\}$ from $q(\mathbf{x}_t|\mathbf{Z}_t, exemplars, \mathbf{I}_{1:t})$, we can approximate Eq. (6) by

$$\begin{aligned} E[\mathbf{x}_t] &\approx \frac{(1/N_P)\sum_{i=1}^{N_p} \mathbf{x}_t^{(i)} \omega_t(\mathbf{x}_t^{(i)})}{(1/N_P)\sum_{i=1}^{N_p} \omega_t(\mathbf{x}_t^{(i)})} \\ &= \sum_{i=1}^{N_p} \widetilde{\omega}_t(\mathbf{x}_t^{(i)}) \mathbf{x}_t^{(i)} \end{aligned} \tag{7}$$

To construct a recursive expression of proposed particle filter, the proposal distribution $q(\mathbf{x}_t|\mathbf{Z}_t, exemplars, \mathbf{I}_{1:t})$ is assumed to have the following form:

$$q(\mathbf{x}_{0:t}|\mathbf{Z}_t, exemplars, \mathbf{I}_{1:t}) = q(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{Z}_t, exemplars, \mathbf{I}_{1:t}) q(\mathbf{x}_{0:t-1}|\mathbf{Z}_{t-1}, exemplars, \mathbf{I}_{1:t-1})$$

Thus the importance weight $\omega_t(\mathbf{x}_t^{(i)})$ can be updated recursively

$$\omega_t(\mathbf{x}_t^{(i)}) = \omega_{t-1}(\mathbf{x}_{t-1}^{(i)}) \frac{p(\mathbf{z}_t|\mathbf{x}_t^{(i)}, exemplars, \mathbf{I}_{1:t}) p(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(i)}, exemplars, \mathbf{I}_{1:t})}{q(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(i)}, \mathbf{Z}_t, exemplars, \mathbf{I}_{1:t})} \tag{8}$$

The proposal distribution here is defined as

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{Z}_t, exemplars, \mathbf{I}_{1:t}) = p(\mathbf{x}_t|\mathbf{x}_{t-1}, exemplars, \mathbf{I}_{1:t})$$

Thus Eq. (8) is transformed to

$$\omega_t(\mathbf{x}_t^{(i)}) = \omega_{t-1}(\mathbf{x}_{t-1}^{(i)}) p(\mathbf{z}_t|\mathbf{x}_t^{(i)}, exemplars, \mathbf{I}_{1:t}) \tag{9}$$

The problem of this filter is that the distribution of importance weights becomes more and more skewed as time increases. This phenomenon is called weight degeneracy or sample impoverishment. To monitor how bad the weight degeneration is, a measure for degeneracy called effective sample size N_{eff} is introduced,

$$N_{eff} = \frac{1}{\sum_{i=1}^{N_p} (\omega_k(\mathbf{x}_k^{(i)}))^2}$$

If N_{eff} is less than a predefined threshold N_T (usually $N_p/2$ or $N_p/3$), resampling operation should be performed. The whole process of EC-PF is described in Table 1.

Compared with classic particle filters, the dynamic model in EC-PF is not predefined or assumed to be only suitable
for smooth motion. EC-PF performs more accurately in prediction, especially under conditions such as abrupt velocity changes, due to the introduction of conditioning with respect to exemplars and image data. The process of EC-PF is shown in Fig. 3.

4. 3D human motion tracking

This section describes, in detail, the proposed 3D human motion tracking method based on EC-PF discussed in Section 3.2. Firstly, 3D human model and relative motion parameters applied in this method are presented and explained. Then, techniques in 3D human motion tracking by using EC-PF are described, such as shape context-based method for estimating dynamic model in prediction phase and likelihood measurement for evaluating similarity between the human pose and human profile in image.

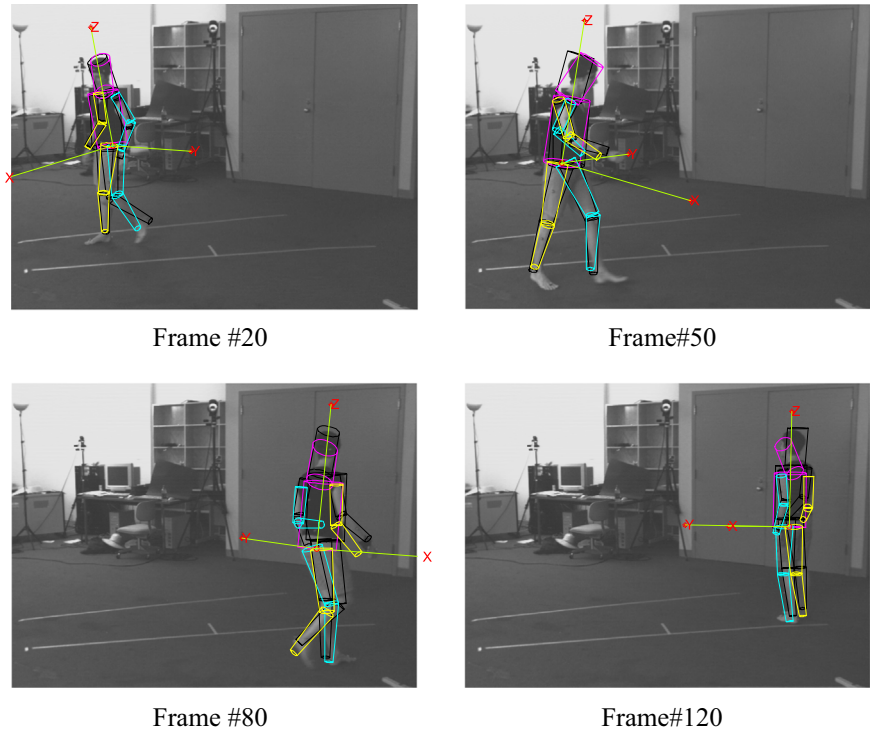

Fig. 11. Tracking results by EC-PF at Frame #20, #50, #80 and #120 in the case of using low-frame rate camera.

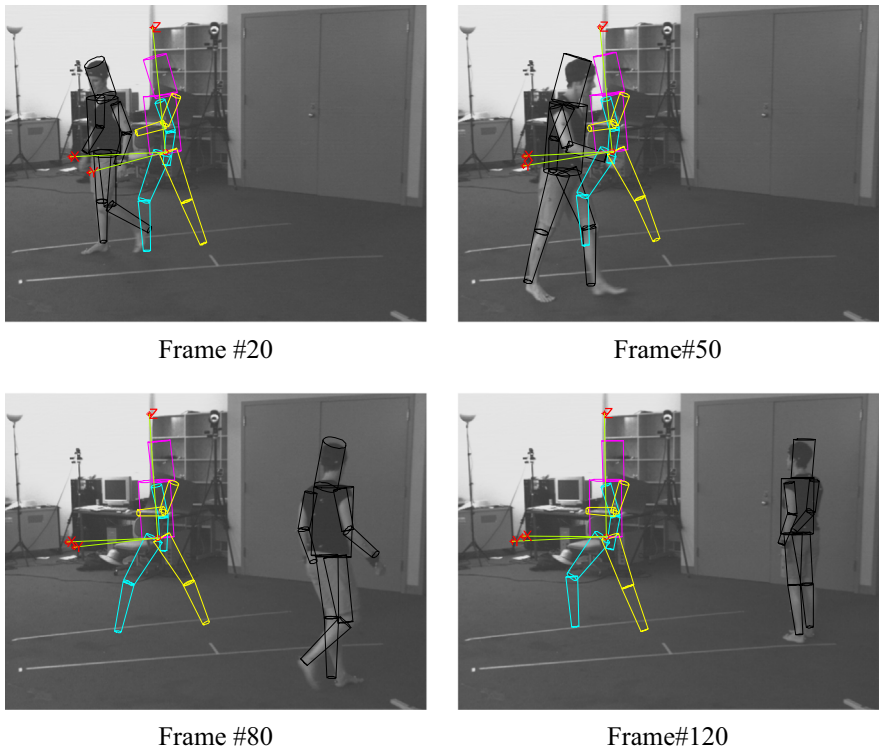

Fig. 12. Tracking results by annealed particle filter at Frame #20, #50, #80 and #120 in the case of using low-frame rate camera.

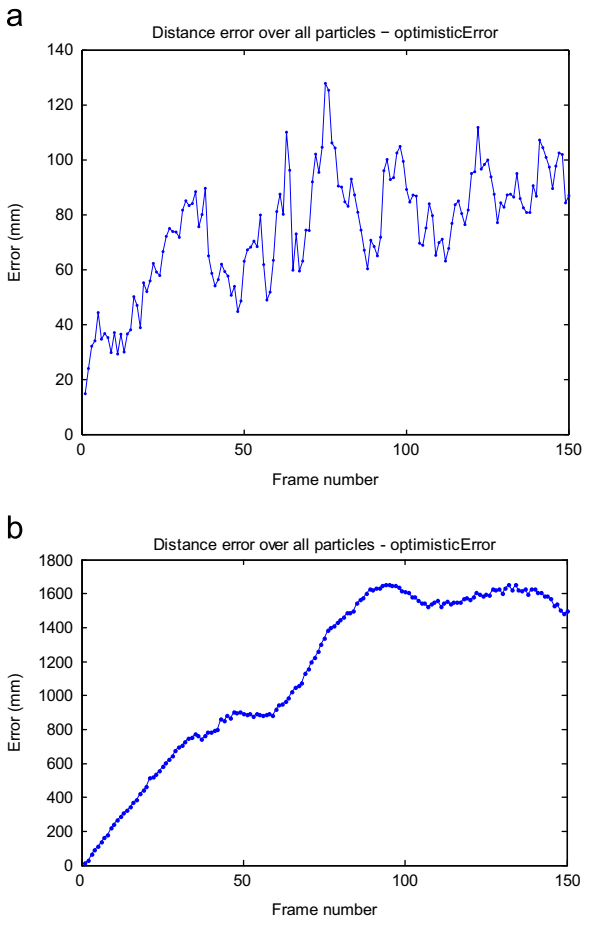

Fig. 13. Tracking error in the case of using low-frame rate camera: (a) tracking error by EC-PF and (b) tracking error by annealed particle filter.

4.1. 3D human model

A 3D human body model used in this paper is built to simulate human pose based on the framework of a kinematic chain (shown in Fig. 4(a)). The 3D human body model consists of 10 rigid segments connected by joints, including a head, a torso, two upper arms, two lower arms, two thighs and two legs, and each segment is depicted by a cylinder. Every two adjacent segments are connected via a joint (see Fig. 4(b)).

Six degrees of freedom are given to translation and rotation of the human global parameters. Each limb has a local coordinate system with Z-axis directed along the limb. Rigid transformations are used to specify relative positions and orientations of each limb. 3D rotation is given to thighs, upper arms and head. 1D rotation is assigned for calves and lower arms. This results in a model with 25 degrees of freedom and the corresponding human state vector is denoted as $\mathbf{x}=(x_1 \quad \cdots \quad x_{25})^{\mathrm{T}}$.

The relative lengths of segments in the human model are predefined based on measurements of an average human body. The value of each body segment used in this paper is indicated in Table 2.

4.2. Prediction step

In prediction phase, the human motion function $f_{exemplars,\mathbf{I}_{1:t}}$ is first estimated from previous human pose, image data and exemplars. The key step is to construct correspondences from image data and exemplars. In recent years, a number of research was done on object correspondence construction. Yu et al. [32,33] proposed a semisupervised patch alignment framework and a semi-supervised multiview subspace learning algorithm for object correspondence construction. In this paper, the 3D human body configuration method using shape contexts [6] is applied to estimate current human pose.

The process consists of several steps which are described as follows:

1. A shortlist of 2D exemplars with prior knowledge is built.
2. Human body edges are extracted from current frame, points are sampled along edges and shape context features are constructed.
3. Current frame is compared with exemplars using shape context matching method to derive joints' locations.
4. 3D human pose is reconstructed based on derived joints' locations and predefined kinetic information.

4.2.1. Shape context

Shape context is a rich descriptor in shape matching and used to find the correspondences between exemplars and current frame. Shapes of 2D human are represented by a discrete set of n points $P=\{\mathbf{p}_1,\ldots,\mathbf{p}_n\}$, $\mathbf{p}_i \in \Re^2$, which are sampled from contours of the shape. The descriptor for a point $\mathbf{p}_i$ is the histogram $\hat{\mathbf{h}}_i=(\hat{\mathbf{h}}_i^1,\ldots,\hat{\mathbf{h}}_i^d)$ for d log-polar histogram bins [34]:

$$\hat{\mathbf{h}}_i^k=\sum_{\mathbf{q}_j\in Q}\mathbf{t}_j, \quad \text{where } Q=\{\mathbf{q}_j\neq\mathbf{p}_i,(\mathbf{q}_j-\mathbf{p}_i)\in bin(k)\} \tag{10}$$

$\mathbf{t}_j$ is a unit length tangent vector and it is along the direction of the edge at $\mathbf{q}_j$. In each histogram bin $\hat{\mathbf{h}}_i^k$, the descriptor is calculated by summing the tangent vectors for all points falling in the bin. An example of shape contexts is shown in Fig. 5.

Each bin holds a single vector in the direction of the dominant orientation of edges in the bin. We compare two histograms at point $\mathbf{p}_i$ and point $\mathbf{q}_i$ using a distance as

$$d(\mathbf{p}_i,\mathbf{q}_i)=\frac{1}{2}\sum_{bins(k)}\frac{||\hat{\mathbf{h}}_i^k-\hat{\mathbf{h}}_j^k||^2}{||\hat{\mathbf{h}}_i^k||+||\hat{\mathbf{h}}_j^k||} \tag{11}$$

Bipartite graph matching method is used to compare all pairs of points $\mathbf{p}_i$ sampled from current frame and $\mathbf{q}_i$ from exemplars. The target is to minimize the total cost of matching

$$H(\pi)=\sum_i d(\mathbf{p}_i,\mathbf{q}_{\pi(i)}) \tag{12}$$

where $\mathbf{p}_i$ is a point on the shape extracted from current frame, $\mathbf{q}_i$ is on the shape derived from an exemplar, π is a permutation. This problem can be solved by using Hungarian method [35].

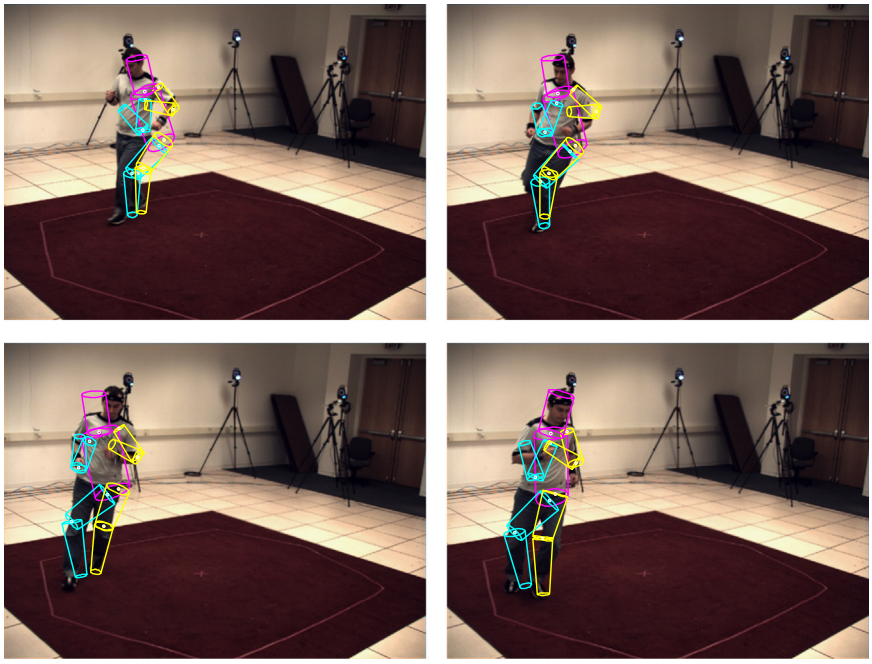

Fig. 14. Some tracking results by the baseline algorithm at Frame #485, #500, #515 and #530 in the case of sudden change in human velocity.

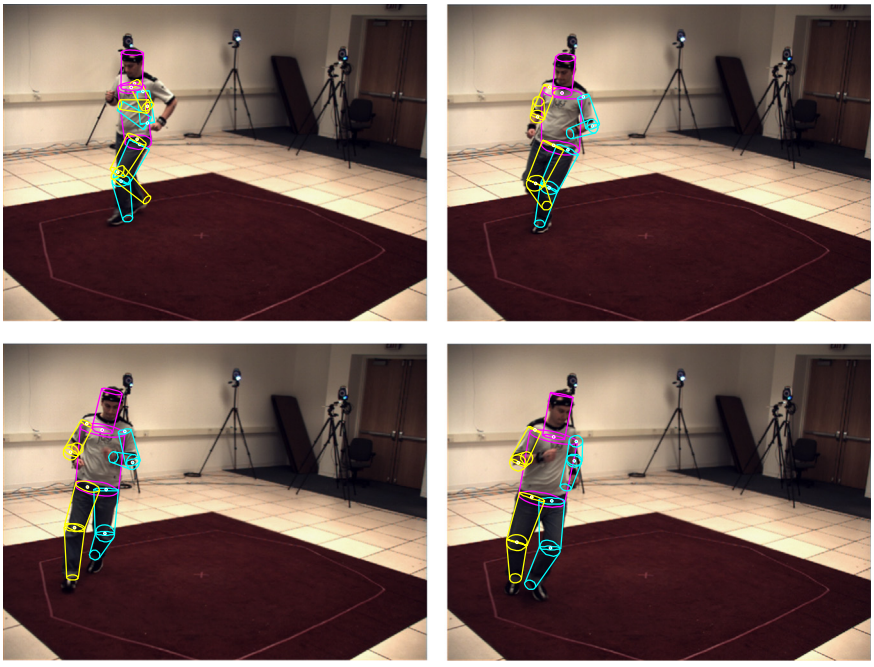

Fig. 15. Tracking results by EC-PF at Frame #485, #500, #515 and #530 in the case of sudden change in human velocity.

4.2.2. 3D human pose recovery

Once the best exemplar is chosen by comparing with current frame, the deformation model proposed in [6] is used to do deformation. Based on the image coordinates of joint points, Taylor's method [36] is then applied to estimate 3D human pose. Fig.6 shows the projection of two perpendicular line segments, whose lengths are l_1 and l_2, onto image plane by projective projection.

In this case, for line segment AB with known length l_1, the projection of two end points,(x_1, y_1, z_1) and (x_2, y_2, z_2) onto image plane are represented by (u_1, v_1) and (u_2, v_2) respectively. For line segment AC, the projections of two end points

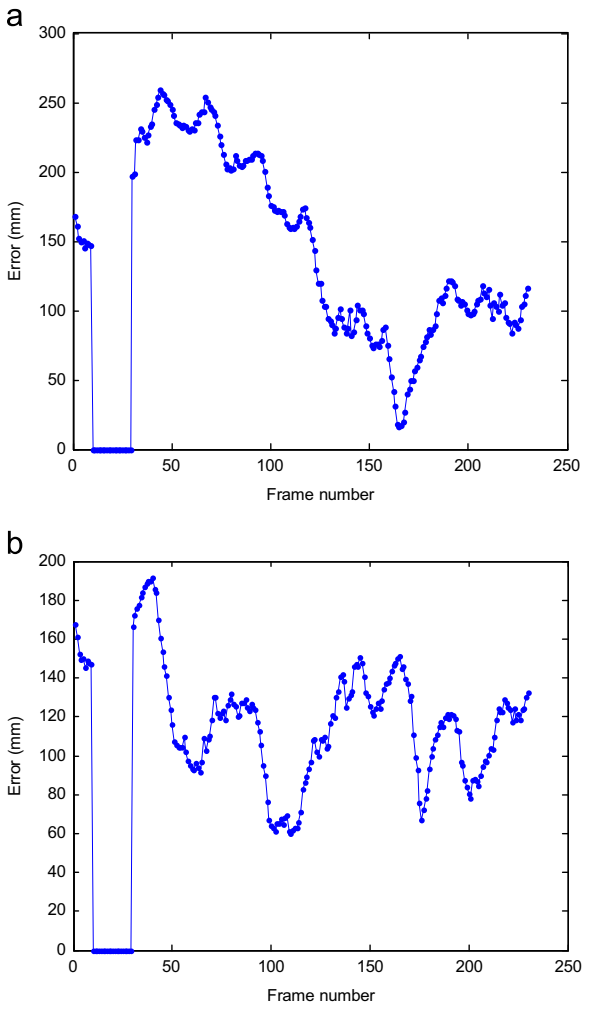

Fig. 16. Tracking error in the case of sudden change in human velocity: (a) tracking error by the baseline algorithm and (b) tracking error by EC-PF.

(x_1, y_1, z_1) and (x_3, y_3, z_3) are (u_1, v_1) and (u_3, v_3) respectively. Assume the scale factors of the three end points are known as s_1,s_2 and s_3, it would be simple to compute the relative depth of the two end points. We have

$$l_1^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2$$
$$l_2^2 = (x_1 - x_3)^2 + (y_1 - y_3)^2 + (z_1 - z_3)^2$$
$$(x_1, y_1, z_1) = \left(\frac{u_1}{s_1}, \frac{v_1}{s_1}, \frac{f}{s_1}\right)$$
$$(x_2, y_2, z_2) = \left(\frac{u_2}{s_2}, \frac{v_2}{s_2}, \frac{f}{s_2}\right)$$
$$(x_3, y_3, z_3) = \left(\frac{u_3}{s_3}, \frac{v_3}{s_3}, \frac{f}{s_3}\right)$$

where f is the camera focal length which is known. Moreover, the two line segments are perpendicular. So

$$(x_1 - x_2)(x_1 - x_3) + (y_1 - y_2)(y_1 - y_3) + (z_1 - z_2)(z_1 - z_3) = 0 \quad (13)$$

The relative depth information between two end points is derived from the matched exemplar, for example,

$$z_1 < z_2 \text{ and } z_1 < z_3$$

According to the above equations and inequalities, s_1,s_2 and s_3 can be derived. The 3D coordinates of the three end points can be computed. In the human model (see Fig. 4 (a)), segment $j_1 j_8$ is always perpendicular to segment $j_9 j_{12}$. Base on the projected image coordinates of the joints, the unknown scale used in [6] can be computed. Therefore other joints' locations can be derived accordingly.

Then Euler angles of relevant joints are calculated by using Inverse Kinemics (IK) based on the obtained 3D joint coordinates to estimate human body states.

Assume that previous state is $\hat{\mathbf{x}}_{t-1}$, human pose derived by above method is $\hat{\mathbf{x}}_{matched,t}$, the motion transition function $f_{exemplars,\mathbf{I}_{1:t}}$ can be computed by solving the following equation:

$$\hat{\mathbf{x}}_{matched,t} = f_{exemplars,\mathbf{I}_{1:t}}(\hat{\mathbf{x}}_{t-1})$$

The motion transition function $f_{exemplars,\mathbf{I}_{1:t}}$ is then used to direct importance sampling. Because $f_{exemplars,\mathbf{I}_{1:t}}$ describes how the human moves from previous pose to current pose, prediction accuracy by applying $f_{exemplars,\mathbf{I}_{1:t}}$ will be improved especially in the case of sudden change in motion velocity. For each particle, it will be predicted as

$$\hat{\mathbf{x}}_t^{(i)} = f_{exemplars,\mathbf{I}_{1:t}}(\hat{\mathbf{x}}_{t-1}^{(i)})$$

4.3. Likelihood measurement function

For each particle, a likelihood measure needs to be computed to estimate how well the projection of a given human body pose fits the observation. Two image features, edges and silhouette (see Fig. 7), are chosen to construct the likelihood measurement function.

Edges produced by a human subject in an image usually provide a good outline of visible arms and legs, and they are mostly invariant to color, texture and lighting. In most situations, edges provide a good measurement for the likelihood function. A gradient-based edge detection mask is used to detect edges. The result is thresholded to eliminate spurious edges, smoothed with a Gaussian mask. The example edge image by this method is shown in Fig. 7(b). A sum-squared difference (SSD) function $SSD(\mathbf{X}, \mathbf{Z})$ is computed as

$$SSD^e(\mathbf{X}, \mathbf{Z}_e) = \frac{1}{N_1} \sum_{i=1}^{N_1} (1 - p_i^e(\mathbf{X}, \mathbf{Z}_e))^2 \quad (14)$$

where $\mathbf{X}$ is the projection of the human model onto the image plane and $\mathbf{Z}_e$ is the image edge image. $p_i^e(\mathbf{X}, \mathbf{Z}_e)$ is the value of edge map at the sampling points taken along the model's edge at point i. N_1 is the number of foreground points.

Another feature extracted from current frame is *silhouette*, which has been generated by learning a Gaussian mixture model for each pixel over background images and comparing the background pixel probability with that of a uniform foreground model. The value of foreground pixels is set to 1 and background to 0. This time, another SSD

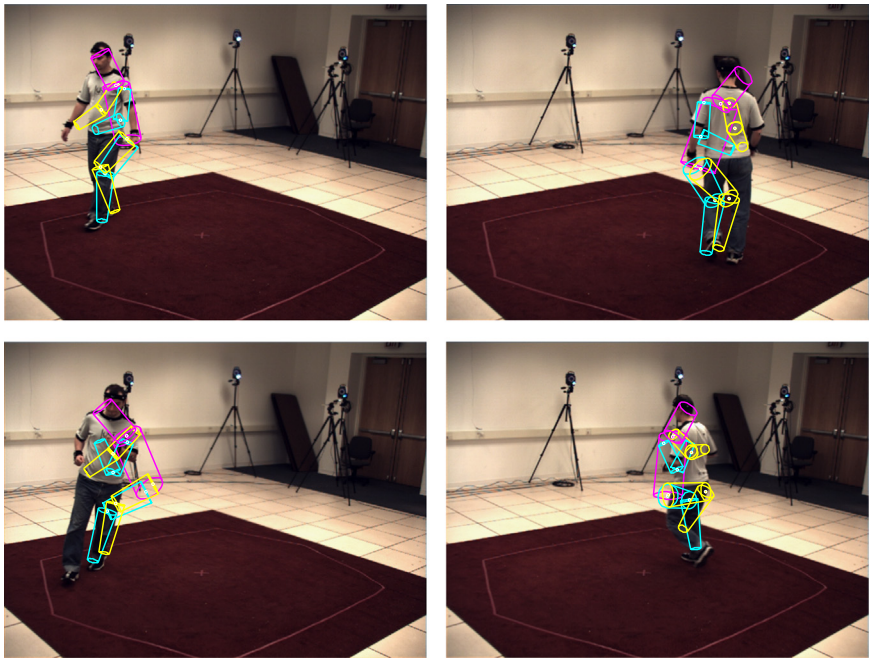

Fig. 17. Some tracking results by the baseline algorithm at Frame #485, #500, #515 and #530 in the case of in the case of using low-frame rate camera.

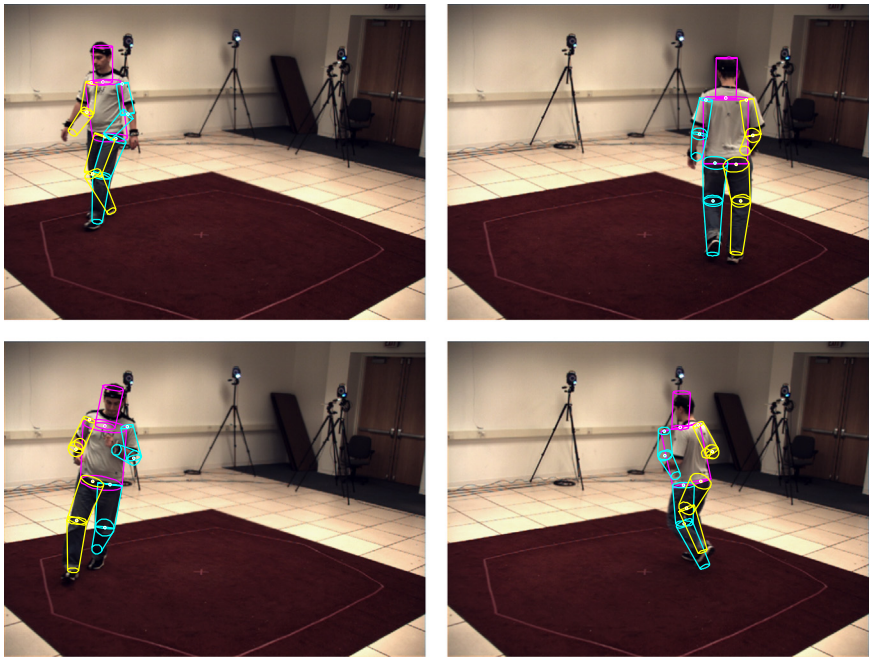

Fig. 18. Tracking results by EC-PF at Frame #485, #500, #515 and #530 in the case of using low-frame rate camera.

function is constructed as

$$SSD^s(\mathbf{X},\mathbf{Z}_s)=\frac{1}{N_2}\sum_{i=1}^{N_2}(1-p_i^s(\mathbf{X},\mathbf{Z}_s))^2 \tag{15}$$

where $p_i^s(\mathbf{X},\mathbf{Z}_s)$ is the value of the sampling point i. To combine these two features, the weight function is proposed as

$$\omega(\mathbf{X},\mathbf{Z})=\exp(-((SSD^e(\mathbf{X},\mathbf{Z}_e)+SSD^s(\mathbf{X},\mathbf{Z}_s))) \tag{16}$$

The entire process of the proposed 3D human tracking algorithm is demonstrated in Table 3.

5. Experiments and discussion

To demonstrate the robustness of EC-PF to sudden change in human velocity and low frame rate, Brown dataset [37] and HumanEva dataset [38] were used for testing.

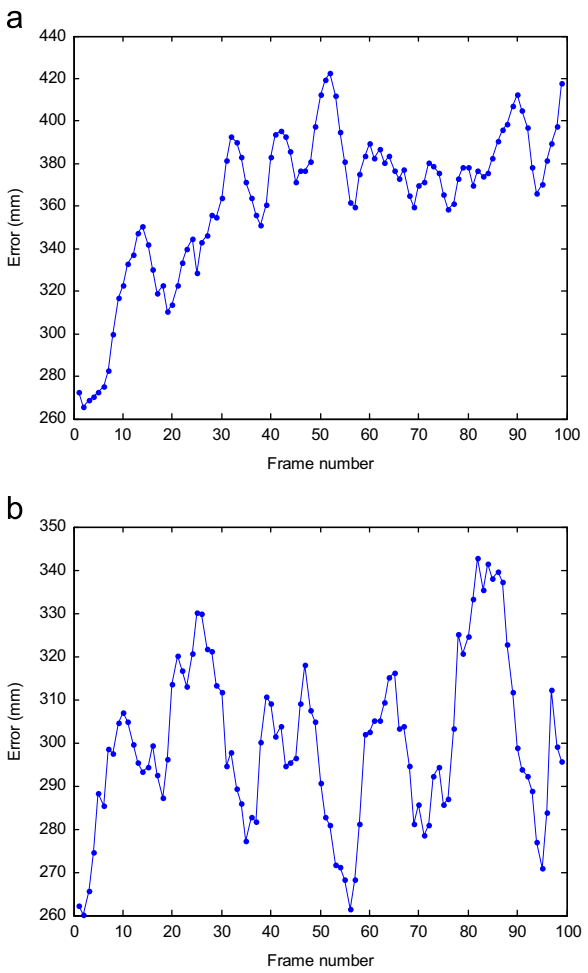

Fig. 19. Tracking error in the case of using low-frame rate camera: (a) tracking error by the baseline algorithm and (b) tracking error by EC-PF.

When experiments were conducted on Brown dataset, performances of EC-PF and annealed particle filter were compared. When experiments were conducted on HumanEva dataset, performance of EC-PF was compared with the baseline algorithm in [38].

5.1. Brown dataset

The images from Brown dataset were taken with four synchronized cameras at 60 Hz, and the provided information includes camera calibration parameters, binary maps for foreground silhouettes, and motion capture data. Motion capture data was collected using a six-camera Vicon system at 120 Hz. Image data and motion capture data were then synchronized in software. Motion capture data was processed separately to obtain body model parameters and ground truth values for joint angles, position and orientation. In this experiment, only images captured from the first camera was used for testing the performance of EC-PF.

A portion of a circular walking sequence was selected from Brown dataset. From this sequence, a set of 400 exemplars was pruned to a shortlist of representative shape contexts with manually added prior information such as joint relative locations. The program was written in Matlab and run on a Dell desktop with Intel Core i7 3.07 GHz CPU and 4 GB memory.

5.1.1. Sudden change in human motion velocity

To test the performance of EC-PF and APF under the condition of sudden change in human velocity, some intermediate frames from chosen image sequence were omitted to simulate such situation. In this experiment, the first 200 frames in Brown dataset were selected by deleting Frame nos. 154–178. The tracking results by EC-PF and original annealed particle filter are given in Figs. 8 and 9 respectively. For annealed particle filter, four cameras were used while one camera was used for EC-PF.

Fig. 8 displays some of the output poses estimated by EC-PF. Experiments shows that EC-PF could effectively track 3D human motion under the condition of sudden human velocity change. In contrast, annealed particle filter failed to track the human motion (see Fig. 9).

Fig. 10 shows the tracking errors by annealed particle filter and EC-PF. Before Frame no. 154, both algorithms worked fine with error less than 100 mm. However, after Frame no. 178 when a sudden change in human velocity occurred, the resulting error of annealed particle filter is almost 400 mm, while that of EC-PF is still satisfactory.

5.1.2. Low frame rate

To test the performance of EC-PF and APF when low-frame rate camera is used, testing frames were re-sampled every 4 frames from the original image sequence. Thus, the new image sequence is at 15 Htz. The tracking results by EC-PF and original annealed particle filter are shown in Figs. 11 and 12 respectively. Annealed particle filter failed to track the subject while EC-PF still tracked the human correctly in such condition.

Comparison of tracking errors between EC-PF and annealed particle filter is given in Fig. 13. The resulting error by annealed particle filter increases to around 1600 mm, while error by EC-PF is less than 140 mm.

5.2. HumanEva dataset

HumanEva datasets [38] contain multiple subjects performing a set of predefined actions with a number of repetitions. Multi-view video sequences were collected at 60 Hz and synchronized with 3D body poses obtained from a motion capture system. In this section, experiments were conducted on HumanEva-II dataset by using EC-PF and the baseline algorithm [38] separately.

5.2.1. Sudden change in human motion velocity

To compare performances of EC-PF with the baseline algorithm under the condition of sudden change in human velocity, Frame nos. 450–680 in HumanEva-II were chosen for experiments and Frame nos. 460–480 were omitted to simulate such condition. Tracking results by EC-PF and the baseline algorithm are given in Figs. 14 and 15 respectively. From the obtained results, the estimated human body by EC-PF is close to ground truth, while the baseline algorithm gives a wrong estimation of human pose.

Fig. 16 shows the comparison of tracking error by EC-PF and the baseline algorithm. From the results, EC-PF outperforms the baseline algorithm by giving less error.

5.2.2. Low frame rate

The same frame re-sampling method as described in Section 5.1.2 was used here. Some tracking results by the baseline algorithm and EC-PF are given in Figs. 17 and 18 respectively. We note that human poses generated by the baseline algorithm are quite different from ground truth while proposed EC-PF still tracked correctly.

Fig. 19 presents errors of human pose derived by the baseline algorithm and EC-PF when a low-frame-rate image sequence was used. The results show that EC-PF is more robust to low frame rate than the baseline algorithm.

6. Conclusion

Particle filter is an effective framework for 3D human motion tracking. Researchers have improved particle filters, such as annealed particle filter and progressive particle filter, to reduce computational burden and improve tracking accuracy. However, the problem of needs for sampling particles in a large enough uncertainty area due to low prediction accuracy remains, especially in cases where sudden human velocity change exists or low frame rate camera is used. Targeting this issue, EC-PF is proposed by introducing a conditional system states with respect to image data and exemplars. In order to obtain 3D poses, an exemplar-based dynamic model is constructed to guide human motion prediction so that particles are able to evolve within an area close to true state. Therefore, this approach is robust to usage of low frame rate camera and sudden motion change of subject. Moreover, by adopting shape context-based exemplar matching, proposed 3D motion tracking approach EC-PF can be effectively achieved with a monocular camera setup, which suggests a better potential for future applications in real world.

References

[1] R. Poppe, Vision-based human motion analysis: an overview, Comput. Vis. Image Underst. 108 (2007) 4–18.
[2] Y. Motai, S.K. Jha, D. Kruse, Human tracking from a mobile agent: optical flow and Kalman filter arbitration, Signal Process. 27 (2012) 83–95.
[3] Y. Yi, Y.K. Lin, Human action recognition with salient trajectories, Signal Process. 93 (2013) 2932–2941.
[4] J. Deutscher, I.D. Reid, Articulated body motion capture by stochastic search, Int. J. Comput. Vis. 62 (2005) 185–205.
[5] R. Urtasun, D.J. Fleet, A. Hertzmann, P. Fua, Priors for people tracking from small training sets, in: IEEE International Conference on Computer Vision, 2005, pp. 403–410.
[6] G. Mori, J. Malik, Recovering 3D human body configurations using shape contexts, IEEE Trans. Pattern Anal. Mach. Intell. 28 (2006) 1052–1062.
[7] G. Mori, J. Malik, Estimating human body configurations using shape context matching, in: Proceedings of the European Conference on Computer Vision, 2002, pp. 150–180.
[8] G. Mori, S. Belongie, J. Mali, Efficient shape matching using shape contexts, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 1832–1837.
[9] G. Mori, S. Belongie, J. Malik, Shape contexts enable efficient retrieval of similar shapes, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2001, pp. 723–730.
[10] J.Y. Zuo, Y. Liang, Y.Z. Zhang, Q. Pan, Particle filter with multimode sampling strategy, Signal Process. 93 (2013) 3192–3201.
[11] Y.H. Yu, Combining $H\infty$ filter and cost-reference particle filter for conditionally linear dynamic systems in unknown non-Gaussian noises, Signal Process. 93 (2013) 1871–1878.
[12] B.J. Zhou, S. Chen, C. Shi, U.M. Providence, Automatic reconstruction of 3D human motion pose from uncalibrated monocular video sequences based on markerless human motion tracking, Pattern Recognit. 42 (2009) 1559–1571.
[13] C. Hong, J. Yu, X. Chen, Image-based 3D human pose recovery with locality sensitive sparse retrieval, in: Proceedings of IEEE International Conference on System, Man, and Cybernetics, 2013, pp. 2103–2108.
[14] Y. Song, X. Feng, P. Perona, Towards detection of human motion, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2000, pp. 810–817.
[15] K. Grauman, G. Shakhnarovich, T. Darrell, Inferring 3D structure with a statistical image-based shape model, in: Proceedings of IEEE Conference on Computer Vision and Patter Recognition, 2003, pp. 641–647.
[16] C. Sminchisescu, B. Triggs, Kinematic jump processes for monocular 3d human tracking, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2003, pp. 69–77.
[17] K. Toyama, A. Blake, Probabilistic tracking with exemplars in a metric space, Int. J. Comput. Vis. 48 (2002) 9–19.
[18] C. Bregler, J. Malik, Tracking people with twists and exponential maps, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1998, pp. 8–15.
[19] S. Niyogi, E. Adelson, Analysing and recognising walking figures in xyt, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1994, pp. 469–474.
[20] R. Liu, G. Shakhnarovich, J.K. Hodgins, H. Pfister, P.A. Viola, Learning silhouette features for control of human motion, ACM Trans. Comput. Graph. 24 (2005) 1303–1331.
[21] M. Isard, A. Blake, Condensation – conditional density propagation for visual tracking, Int. J. Comput. Vis. 29 (1998) 5–28.
[22] O.D. King, D.A. Forsyth, How does condensation behave with a finite number of samples? in: Proceedings of the European Conference on computer Vision, 2000, pp. 695–709.
[23] L. Raskin, E. Rilvlin, M. Rudzsky, Dimensionality reduction for articulated body tracking, in: Proceedings of 3DTV Conference, 2007, pp. 1–4.
[24] J. Deutscher, A. Blake, I. Reid, Articulated body motion capture by annealed particle filtering, in: Proceedings of IEEE Conference on Computer Visiona and Pattern Recognition, 2000, pp. 126–133.
[25] N.D. Lawrence, Gaussian process latent variable models for visualisation of high dimensional data, Adv. Neural Inf. Process. Syst. 16 (2003) 329–336.
[26] A.O. Balan, L. Sigal, M.J. Black, A quantitative evaluation of video-based 3D person tracking, in: Proceedings of the 2nd Joint IEEE International Workshop on Visual Surveilance and Performance Evaluation of Tracking and Surveillance, 2005, pp. 349–356.
[27] C. Sminchisescu, B. Triggs, Estimating articulated human motion with covariance scaled sampling, Int. J. Robot. Res. 22 (2003) 371–392.
[28] J. Deutscher, I.D. Reid, Articulated body motion capture by stochastic search, Int. J. Comput. Vis. 61 (2005) 185–205.
[29] J. Deutscher, A. Davison, I. Reid, Automatic partitioning of high dimensional search spaces associated with articulated body motion capture, CVPR (2001) 669–676.
[30] M. Vondrak, L. Sigal, O.C. Jenhins, Dynamical simulation priors for human motion tracking, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2013) 52–65.
[31] I.C. Chang, S.Y. Lin, 3D human motion tracking based on a progressive particle filter, Pattern Recognit. 43 (2010) 3621–3635.
[32] J. Yu, D.Q. Liu, D.C. Tao, H.S. Seah, Complex object correspondence construction in two-dimensional animation, IEEE Trans. Image Process. 20 (2011) 3257–3269.
[33] J. Yu, D.Q. Liu, D.C. Tao, H.S. Seah, On combining multiple features for cartoon character retrieval and clip synthesis, IEEE Trans. Syst. Man Cybern. Part B 42 (2012) 1413–1427.
[34] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 509–522.
[35] C. Papadimitriou, K. Stieglitz, Combinatorial Optimization: Algorithms and Complexity, Prentice-Hall, New Jersey, 1982.
[36] C.J. Taylor, Reconstruction of articulated objects from point correspondences in a single uncalibrated image, Comput. Vis. Image Underst. 880 (2000) 677–684.
[37] S. Bhatia, S. Roth, M.J. Black, M. Isard, Tracking loose-limbed people, in: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2004 pp. I421–I428.
[38] L. Sigal, A.O. Balan, M.J. Black, HumanEva: synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion, Int. J. Comput. Vis. 87 (2010) 4–27.