Generative Estimation of 3D Human Pose Using Shape Contexts Matching

Xu Zhao and Yuncai Liu

Institute of Image Processing & Pattern Recognition, Shanghai Jiao Tong University
200240, Shanghai, China
{zhaoxu,whomliu}@sjtu.edu.cn

Abstract. We present a method for 3D pose estimation of human motion in generative framework. For the generalization of application scenario, the observation information we utilized comes from monocular silhouettes. We distill prior information of human motion by performing conventional PCA on single motion capture data sequence. In doing so, the aims for both reducing dimensionality and extracting the prior knowledge of human motion are achieved simultaneously. We adopt the shape contexts descriptor to construct the matching function, by which the validity and the robustness of the matching between image features and synthesized model features can be ensured. To explore the solution space efficiently, we design the Annealed Genetic Algorithm (AGA) and Hierarchical Annealed Genetic Algorithm (HAGA) that searches the optimal solutions effectively by utilizing the characteristics of state space. Results of pose estimation on different motion sequences demonstrate that the novel generative method can achieves viewpoint invariant 3D pose estimation.

1 Introduction

Capturing 3D human motion from visual cues has received increasing attention in recent years, due to the drive from a wide spectrum of potential applications such as behavior understanding, content-based image retrieval and visual surveillance. Although having been attacked by many researchers, this challenging problem is still long standing because of the difficulties conduced mainly by complicated nature of 3D human motion and incomplete information of 2D images for 3D human motion analysis.

In the context of graphical models, the state-of-art approaches of 3D human motion estimation can be classified as generative and discriminative [1]. *Generative methods* [2,3,4,5,6,7] follow the bottom-up Bayes' rule and model the state posterior density using observation likelihood or cost function. Given an image observation and prior state distribution, the posterior likelihood is usually evaluated using Bayes' rule. This approach has a sound framework of probabilistic support and can achieve significant success for recovering complex unknown motions by utilizing well-defined state constrains. However, it is generally computationally expensive because one had to perform complex search over the state space in order to locate the peaks of the observation likelihood. Moreover, prediction model and initialization are also the bottlenecks of the approach especially in tracking situation.

In this paper, we propose a novel generative approach in the framework of evolutionary computation, by which we try to widen the bottlenecks mentioned above with effective search strategy embedded in the extracted state subspace. Considering the generalization of application scenario, the observation information we utilized comes from an uncalibrated monocular camera. This makes the state estimation get into severe ill-conditioned problem. And, we have to confront the curse of dimensionality because there are more than forty degrees of freedom (DOFs) of full body joints in our 3D human model. Therefore, the process for searching optimal solutions should be performed in some compact state space by the search algorithms which suit for the characteristics of this space. In doing so, infeasible solutions, namely, the absurd poses can be avoided naturally. To this end, we consider to reduce the dimensionality of state space by principal component analysis (PCA) of motion capture data. Actually, the motion capture data embody the prior knowledge about human motion. By PCA, the aims at both reducing dimensionality and extracting the prior knowledge of human motion are achieved simultaneously. From the theoretical view, PCA is optimal in the sense of reconstruction because it allows the minimal information loss in the course of state transformation from the subspace to original state space. Different from the previous works [8,9], we perform the lengthways PCA, by which the subspace can be extracted from only single sequence of motion capture data.

To explore the solution space efficiently, we design the Annealed Genetic Algorithm (AGA) combining the ideas of simulated annealing and genetic algorithm [10]. As the promoted version of AGA, Hierarchical Annealed Genetic Algorithm (HAGA) searches the optimal solutions more effectively than AGA by utilizing the characteristics of state space. According to the theory of PCA, in our problem, the first principle component captures the most important part of human motion and the rest of principle components capture the detailed parts of this motion. In monocular uncalibrated camera situation, the fitness function (observation likelihood function) is very sensitive to the change of global motion. The HAGA performs hierarchical search automatically in the extracted state subspace by localizing priorly the state variables such as the global motions and the coordinate of the first principle component which dominate the topology of state space. We adopt the shape contexts descriptor [11] to construct the fitness function, by which the validity and the robust matching between image features and synthesized model features can be achieved.

1.1   Related Work

There has been considerable previous work on capturing human motion from image information. The earlier work on this research topic had been reviewed comprehensively by the survey papers [12,13,14]. Generally speaking, to recover 3D human pose configuration, more information are required than image can provide especially in the monocular situation. Therefore, much work focus on using prior knowledge and experiential data in order to alleviate the ill-condition of this problem. Explicit body model embodies the most important prior knowledge about pose configuration and thus be widely used in human motion analysis. Another class of important prior knowledge comes from the experiential data such as motion capture data acquired by commercial

motion capture system and other hand-labeled data. The combination of the both prior information can produces favorable techniques for solving this problem.

Agarwal et al. [5] distill prior information (the motion model) of human motion from hand-labeled training sequences using PCA and clustering on the base of a simple 2D human body model. This method presents a good autoregressive based tracking scheme but has no description about pose initialization. In the framework of generative approach, the prior information is usually employed to constrain or reduce the search space. Urtasun et al. [15,9] construct a differentiable objective function based on the PCA of motion capture data and then find the poses of all frames simultaneous by optimizing a function in low-dim space. Sidenbladh et al. [3,8] present similar methods in the framework of stochastic optimization. For a specific activity, such methods need many example sequences of motion capture to perform PCA and all of these sequences must keep same length and same phase by interpolating and aligning. Ning et al. [6] learn a motion model from semi-automatically acquired training examples which are aligned with correlation function, and then, some motion constrains are introduced to cut the search space. Unlike these methods, we extract the state subspace from only one example sequence of a specific activity using the lengthways PCA and thus have no use for interpolating or aligning. In addition, useful motion constraints are included naturally in the low-dim subspace.

In recent years, particle filter [16] (also known as condensation algorithm) based optimization methods are used widely for recovering human pose in generative framework [2,3,4,5,6,7]. However, as a stochastic search algorithm, we think that particle filter is essentially similar with evolutionary algorithm (EA) if having no explicit temporal dynamic model. The EA can provide more flexible evolutionary mechanism such as crossover operator. This is the important motivation for us to solve this problem in the framework of EA. A noticeable example showing the relationship between particle filter and EA is the work of Deutscher et al. [17]. By introducing the crossover operator, the annealed particle filter proposed in theirs earlier work [2] get remarkable improvement.

Comparing with previous generative methods, extracting the common characteristic of a special types of motion from prior information and represent them with some compact forms are of particular interests to us. At the same time, we ensure the motion individuality of the input sequences with effective evolutionary search strategy suiting for the characteristic of state subspace.

2 State Space Analysis

The potential special interests motivate us to analyze the characteristics and structure of the state space. Such interests involve mainly modeling the human activities effectively in the extracted state subspace and eliminating the curse of dimension.

2.1 Pose Representation

We use a explicit model that represent the articulated structure of the human body. Our fundamental 3D skeleton model (see Figure 1.a) is composed of 34 articulated rigid sticks. The pose is described by a 44 dimensional vector $\mathbf{x} = \{\mathbf{x}_g, \mathbf{x}_j\}$, where 3D vector
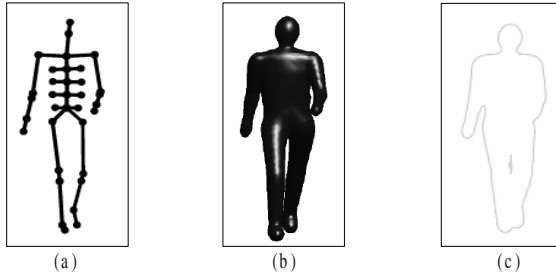
Fig. 1. (a) The 3D human skeleton model. (b) The 3D human convolution surface model. (c) The 2D convolution curves.

$\mathbf{x}_g$ represents the global rotations of human motion and 41D vector $\mathbf{x}_j$ represents the joint angles.

Figure 1.b shows the 3D convolution surface [18] human model which actually is an isosurface in a scalar field defined by convolving the 3D body skeleton with a kernel function. Similarly, the 2D convolution curves of human body as shown in Figure 1.c are the isocurves generated by convolving the 2D projection skeleton. As the synthetical model features, the curves are used to match with the edges of image silhouettes for constructing the likelihood function.

2.2 Subspace Extraction

All of the 3D poses distribute in the state space $\mathbb{X}$. The pose set which belongs to a special activity, such as walking, running, handshaking, etc., generally crowd in a subspace of $\mathbb{X}$. We extract the subspace $\mathbb{X}_s$ from motion capture data obtained from the CMU database (http://mocap.cs.cmu.edu/).

Assuming $\{\mathbf{x}_t \mid \mathbf{x}_t \in \mathbb{X}\}$ is a given data sequence of motion capture corresponding to one motion type, where t is the time tag, the subspace $\mathbb{X}_s$ is extracted by PCA as follows:

1. Centering the state vectors and assembling them into a matrix (by rows):
   $\mathbf{X} = [(\mathbf{x}_1 - c); (\mathbf{x}_2 - c); \ldots ; (\mathbf{x}_T - c)]$, where c is the mean vector.
2. Performing a singular value decomposition of the matrix to project out the dominant directions: $\mathbf{X} = \mathbf{U}\,\mathbf{D}\,\mathbf{V}^T$.
3. Projecting the state vectors into the dominant subspace: each state vector is represented as a reduced vector $\mathbf{x}_s = (\mathbf{x} - c)\,\mathbf{U}_m$, where $\mathbf{U}_m$ is the matrix consisting of first m columns of $\mathbf{U}$, by which the m-D subspace $\mathbb{X}_s$ is spanned.

Therefore, the original state vector $\mathbf{x}$ can be reconstructed by:

$$\mathbf{x} = c + \mathbf{x}_s\,\mathbf{U}_m^T \qquad (1)$$

The dimensionality m of subspace $\mathbb{X}_s$ is determined according to the cumulative sum ε of principal component variance percentage. With our experiences, the value of ε is set to be not smaller than 0.95; accordingly, the value of m is not greater than 6 generally.

3 Fitness Function

In generative framework, pose capturing can be formulated as Bayesian posterior distribution inference:

$$p(\mathbf{x}_s \mid \mathbf{y}) \propto p(\mathbf{x}_s)p(\mathbf{y} \mid \mathbf{x}_s) \tag{2}$$

The function $p(\mathbf{y} \mid \mathbf{x}_s)$ represents the likelihood observing in image $\mathbf{y}$, conditioned on a pose candidate $\mathbf{x}_s$. It is used to evaluate every pose candidate generated from $p(\mathbf{x}_s)$. In the context of evolutionary algorithm, the likelihood function is just the fitness function. We propose a fitness function on the basis of shape contexts matching [11].

   We choose the image silhouette of subject as the observed image feature, which is extracted using statistical background subtraction. The shape context descriptor is used to describe the shape of image silhouette and convolution curves generated by the pose candidate (see Figure 1). Figure 2 illustrate the shape contexts [11] (histograms of local edge pixels into log-polar bins) of human shape. Our shape contexts contain 12 angular $\times$ five radial bins, giving rise to 60-dimensional histograms as shown in Figure 2.b. In the matching process, the regularly spaced points on the edge of the silhouette are sampled as the query shape. The point set sampled from the convolution curves is viewed as the candidate shape. Before matching, the image shape and the candidate shape are normalized to same scale. We represent the query shape and the candidate shape as $S_{query}(\mathbf{y})$ and $S_m(\mathbf{x}_s)$ respectively. To this end, the matching cost function is formulated as:

$$F(S_{query}(\mathbf{y}), S_m(\mathbf{x}_s)) = \sum_{j=1}^{r} \chi^2(SC_{query}^j(\mathbf{y}), SC_m(\mathbf{x}_s)^*) \tag{3}$$

where SC is the shape context, r is the number of sample point on the edge of image silhouette, and $SC_m(\mathbf{x}_s)^* = argmin_u\chi^2(SC_{query}^j(\mathbf{y}), SC_m^u(\mathbf{x}_s))$. Here, we use the χ^2 distance as the similarity measurement. In AGA, the optimization mechanism are designed for searching the maximal value of object function. Therefore, according to Eq. (3), the fitness function can be reformulated as:

$$\mathcal{F}(S_{query}(\mathbf{y}), S_m(\mathbf{x}_s)) = C \cdot \exp(-F(S_{query}(\mathbf{y}), S_m(\mathbf{x}_s))) \tag{4}$$

where C is a constant for adjusting the value range of fitness function.

4 Pose Estimation Using HAGA

In this section, we describe the key algorithms of the generative framework, namely, the AGA and HAGA, and theirs adaption for pose estimation from monocular silhouettes.

4.1 Hierarchical Annealed Genetic Algorithm

Combining simulated annealing (SA) and genetic algorithm (GA), we design the annealed genetic algorithm, which actually is a hybrid $(1 + 1)$ evolutionary strategy. In our algorithm, the local optimal solutions are avoided by introducing several genetic evolutionary principles. We represent chromosome by state vector as $\mathbf{x} = [x_1, x_2, \ldots, x_n]$,
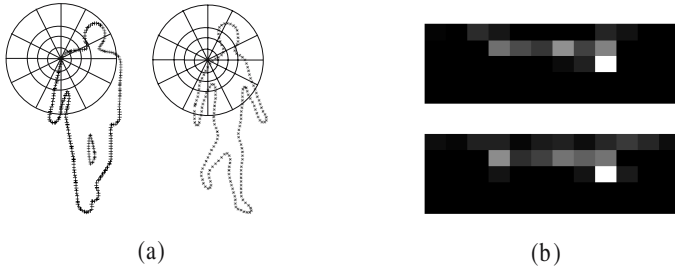
Fig. 2. (a) The shape contexts computed on edge points of image silhouette (right) and sampled points of convolution curves (left). (b) The example shape contexts for reference samples showed in (a) of image silhouette (bottom) and convolution curves (top).

where the genes $\{x_i \mid i = 1, 2, \ldots, n\}$ are random numbers uniformly distributed in the interval $(0, 1)$. We use real encodings. The algorithm searching for optimal solutions with the AGA is described as follows:

> **Parameter initialization** set values for evolution control parameters:
> S_t – stop criteria;
> N_t – termination condition;
> E_t – times for searching a equation state;
> **for** $s_t = 1$ **to** S_t **do**:
> $NonImproveNum \leftarrow 0$;
>   Generate the genes of **x** uniformly at random in the interval $(0, 1)$;
>   Evaluate the fitness function $\mathcal{F}(\mathbf{x})$ by mapping **x** onto the problem domain;
>   **while** ($NonImproveNum < N_t$) **do**
>   **for** $e_t = 1$ **to** E_t **do**:
>   Evolution of **x** driven by the genetic operators; (see Table 1.)
>   Evaluate $\mathcal{F}(\mathbf{x})$;
>   **end for**
>   If the value of fitness function is improved, $NonImproveNum \leftarrow 0$, else
> $NonImproveNum \leftarrow NonImproveNum + 1$;
>   **end while**
>   Record the optimal **x**;
> **end for**

We design five genetic operators, which are executed orderly in AGA. The operators are introduced by evolving a example chromosome $\mathbf{x} = [x_1, x_2, x_3, x_4, x_5, x_6]$. The new chromosome generated by the operators is denoted as $\mathbf{x}'$. Assuming the positions generated randomly are number 2 and number 6 or number 3 (for point mutation operator), for example, the five operators are illustrated in Table 1. (The new genes are represented as x'). On the basis of AGA, we develop a HAGA by utilizing the characteristics of state space $\mathbb{X}$. In HAGA, the state space is decomposed automatically by computing the variances of state components which are generated in each annealing run. According to the variances of state components, the state space is partitioned by localizing down the important components to a small area of theirs range. It is explainable in theory because

Table 1. The genetic operators in AGA

Operators	Example
Exchange	$\mathbf{x} = [x_1, x_2, x_3, x_4, x_5, x_6] \longrightarrow \mathbf{x}' = [x_1, x_6, x_3, x_4, x_5, x_2]$
Segment reversion	$\mathbf{x} = [x_1, x_2, x_3, x_4, x_5, x_6] \longrightarrow \mathbf{x}' = [x_1, x_6, x_5, x_4, x_3, x_2]$
Segment shift	$\mathbf{x} = [x_1, x_2, x_3, x_4, x_5, x_6] \longrightarrow \mathbf{x}' = [x_1, x_6, x_2, x_3, x_4, x_5]$
Point mutation	$\mathbf{x} = [x_1, x_2, x_3, x_4, x_5, x_6] \longrightarrow \mathbf{x}' = [x_1, x_2, x_3', x_4, x_5, x_6]$
Segment mutation	$\mathbf{x} = [x_1, x_2, x_3, x_4, x_5, x_6] \longrightarrow \mathbf{x}' = [x_1, x_2', x_3', x_4', x_5', x_6']$

the important state components dominate the topology of the state space and the little changes of theirs value can produce great effect whereas the values of other state components had little influence on whether they were selected or not. This theory is illustrated in Figure 3. Focusing only on one annealing run of sate evolution ($s_t \rightarrow s_t + 1$), we describe the detailed HAGA as follows.

1. Generate initial chromosome $\mathbf{x} = [x_1, x_2, \ldots, x_n]$ at random, where $\{x_i \,|\, i = 1, 2, \ldots, n\}$ are random numbers uniformly distributed in the interval $(0, 1)$. Mapping it linearly into the variance domain:

$$\mathbf{x} \mapsto \mathbf{x}_t \in (\min \mathbf{x}_t, \max \mathbf{x}_t) \tag{5}$$

   In the first round of state evolution, $(\min \mathbf{x}_1, \max \mathbf{x}_1) = (\mathbf{0}, \mathbf{1})$. Evaluating the fitness function $\mathcal{F}(\mathbf{x})$.
2. Evolve the chromosome according to the state evolutionary mechanism of AGA. Before evaluating the fitness function, every new chromosome needs to be mapped onto the variance domain as formulated in Eq.5.
3. Store N best states (chromosomes) and computing the covariance matrix:

$$\mathbf{V}_{t+1} = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{x}_{t+1}^i - \mathbf{x}_{t+1}^c)^T \cdot (\mathbf{x}_{t+1}^i - \mathbf{x}_{t+1}^c) \tag{6}$$

   where $\mathbf{x}_{t+1}^c$ is the mean vector, and the covariance matrix $\mathbf{V}_{t+1}$ is a diagonal matrix on the assumption that the state components are independent each other. To this end, the variance domain can be formulated as:

$$\begin{cases} \min \mathbf{x}_{t+1} &= \mathbf{x}_{t+1}^c - \mathbf{c}_{t+1}\mathbf{V}_{t+1} \\ \max \mathbf{x}_{t+1} &= \mathbf{x}_{t+1}^c + \mathbf{c}_{t+1}\mathbf{V}_{t+1} \end{cases} \tag{7}$$

   where $\mathbf{c}_{t+1} = [c_{t+1}, c_{t+1}, \ldots, c_{t+1}]$ is used to adjust the variance domain and c_{t+1} is a positive constant.
4. The variance domain $(\min \mathbf{x}_{t+1}, \max \mathbf{x}_{t+1})$ is used to cut down the state space in next state evolution.

4.2 Experiments

We demonstrate our method by extracting subspaces for different classes of human motion and using them to estimate 3D body pose in unseen video sequences.
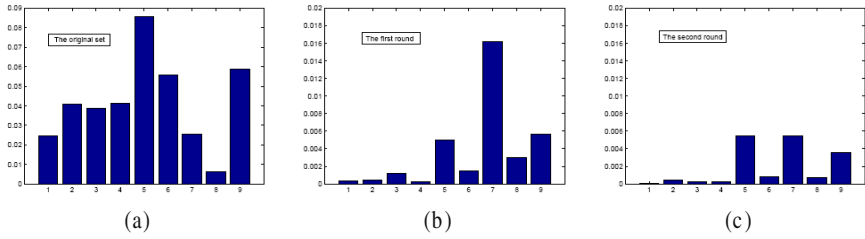
(a)   (b)   (c)

Fig. 3. Variance reduction contrast between principal state components and other state components. Graph (a) shows the variances of state set in which the chromosomes have not been evolved, displaying almost equal variances for each components. Graph (b) shows the variances of state set which have come through one round of state evolution, noticing that the variances of first four principal state components have been greatly reduced whereas the variances of other components have been reduced with a slighter extent. In graph (c), the variances of the principal components have been reduced to very small scopes indicating advanced localization after coming through two rounds of state evolution.

Walking motion: straight walk and turning walk. To extract the motion subspace of walking, a data set consisting of motion capture data of a single subject was used. The total frame number is 316. It was found that the different subject and different frame numbers can produce generally identical subspace. To keep the ratio of information loss lower than 0.05, the dimensionality of the subspace was choose to be 5. For the sequence of one subject walking in a straight line, the parameters of HAGA are set as $S_t = 2, N_t = 2, E_t = 5$. The results are showed in Figure.4. It can be seen that the estimator is successful in determining the correct global motion as well as the 3D pose of the subject. The occlusion problem are tackled by searching the optimal pose in the extracted subspace because the prior knowledge about walking motion is contained in
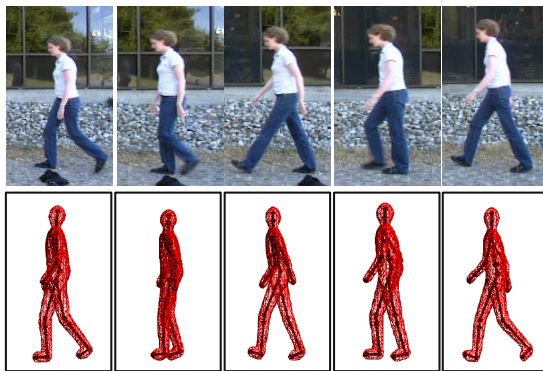

Fig. 4. Results of recovering the poses of a subject walking straight. (the images are part of a sequence from www.nada.kth.se/ hedvig/data.html). The second pose demonstrated the left-right confusion in the silhouette.
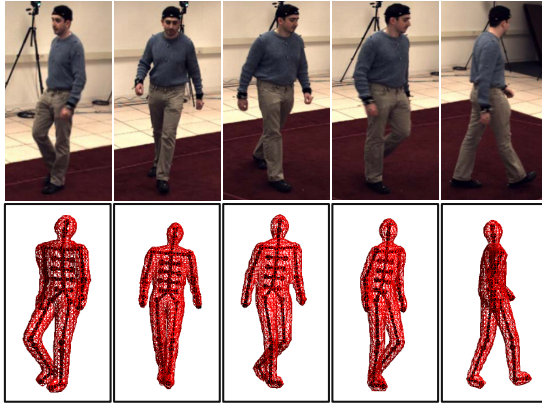
Fig. 5. Results of recovering the poses of a subject performing a turning walking motion

this space. The left-right confusion is mostly disambiguated, however, in few frames, the left-right confusion conduced by silhouette ambiguity still exist. This can be seen from Figure. 4. We test the generalization capability of our method in a turning walk sequence. In this sequence [19], a subject is performing continuing turning walking motion around a circle therefore the global motion is changed in a wide range. The parameters of HAGA are set as $S_t = 2, N_t = 2, E_t = 5$. The results can be seen in Figure.5.

Running motion. The subspace of running motion is extracted from motion capture data that consisted of 130 frames. This subspace is more compact than that of walking motion. Figure.6 shows the estimation results of 3D poses.
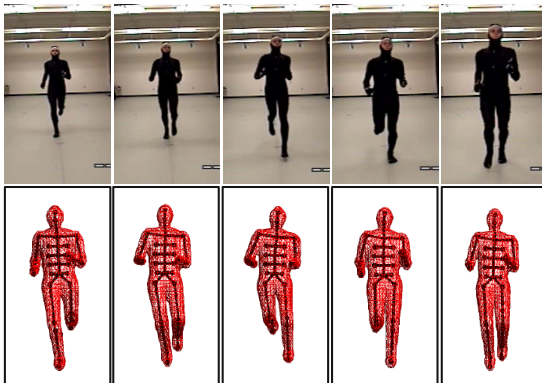

Fig. 6. Results of recovering the poses of a subject performing a running motion. The images are extracted from the video taken from the web site http://mocap.cs.cmu.edu/.

5   Conclusion

We have discussed a novel generative approach to estimating 3D human pose from a single camera. Our approach is a step towards describing motion characteristic of high-dimensional data spaces by extracting its subspace. From motion capture data, we not only distilled the prior knowledge about human motion, but also reduced the dimensionality of problem. In the compact subspace, we perform effective search for finding the optimal poses. To explore the solution space efficiently, we designed the AGA and HAGA, by which the optimal solutions can be searched effectively by utilizing the characteristics of state subspace. The robust shape contexts descriptor allows us using the silhouettes as image features. The approach was tested on different human motion sequences with good results, and allows the estimation of complex unseen motions in the presence of image ambiguities. In terms of future work, the more interior edge information need be added to disambiguate some challenging sequences. Including a wider range of motion capture data would allow the estimator to cover more types of human motions.

Acknowledgements

References

1. Sminchisescu, C., Kanaujia, A., Li, Z., Metaxas, D.: Discriminative density propagation for 3d human motion estimation. In: Proc. Conf. Computer Vision and Pattern Recognition, pp. 217–323 (2005)
2. Deutscher, J., Blake, A., Reid, I.: Articulated body motion capture by annealed particle filtering. In: Proceedings of the 2000 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 126–133 (2000)
3. Sidenbladh, H., Black, M., Fleet, D.: Stochastic tracking of 3D human figures using 2D image motion. In: European Conference on Computer Vision, vol. 2, pp. 702–718 (2000)
4. Sminchisescu, C., Triggs, B.: Covariance scaled sampling for monocular 3D body tracking. In: IEEE International Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 447–454 (2001)
5. Agarwal, A., Triggs, B.: Tracking articulated motion using a mixture of autoregressive models. In: Pajdla, T., Matas, J. (eds.) ECCV 2004. LNCS, vol. 3023, pp. 54–65. Springer, Heidelberg (2004)
6. Ning, H., Tan, T., Wang, L., Hu, W.: People tracking based on motion model and motion constraints withautomatic initialization. Pattern Recognition 37(7), 1423–1440 (2004)
7. Mori, G., Malik, J.: Recovering 3 D Human Body Configurations Using Shape Contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(7), 1052–1062 (2006)
8. Sidenbladh, H., Black, M., Sigal, L.: Implicit Probabilistic Models of Human Motion for Synthesis and Tracking. In: European Conference on Computer Vision, vol. 1, pp. 784–800 (2002)

9. Urtasun, R., Fleet, D., Fua, P.: Monocular 3-D Tracking of the Golf Swing. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, IEEE Computer Society Press, Los Alamitos (2005)

10. Michalewicz, Z.: Genetic algorithms+ data structures= evolution programs. Springer, Heidelberg (1996)

11. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(4), 509–522 (2002)

12. Aggarwal, J., Cai, Q.: Human motion analysis: A review. Computer Vision and Image Understanding 73(3), 428–440 (1999)

13. Gavrila, D.: Visual analysis of human movement: A survey. Computer Vision and Image Understanding 73(1), 82–98 (1999)

14. Moeslund, T., Granum, E.: A survey of computer vision-based human motion capture. Computer Vision and Image Understanding 81(3), 231–268 (2001)

15. Urtasun, R., Fua, P.: 3D Human Body Tracking using Deterministic Temporal Motion Models. In: European Conference on Computer Vision, vol. 3, pp. 92–106 (2004)

16. Arulampalam, M., Maskell, S., Gordon, N., Clapp, T., Sci, D., Organ, T., Adelaide, S.: A tutorial on particle filters for online nonlinear/non-GaussianBayesian tracking. IEEE Transactions on Signal Processing, IEEE Transactions on [see also Acoustics, Speech, and Signal Processing 50(2), 174–188 (2002)

17. Deutscher, J., Davison, A., Reid, I.: Automatic partitioning of high dimensional search spaces associated with articulated body motion capture. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society Press, Los Alamitos (2001)

18. Jin, X., Tai, C.: Convolution surfaces for arcs and quadratic curves with a varying kernel. The Visual Computer 18(8), 530–546 (2002)

19. Sigal, L., Black, M.J.: Humaneva: Synchronized video and motion capture dataset for evaluation of articulated human motion. Technical Report CS-06-08, Brown University (2006)