Yi Li · Zhengxing Sun

Generative Tracking of Human Motion by Sequential Clonal Selection Algorithm

Abstract High dimensional pose state space is the main challenge in articulated human motion tracking. In this paper, we propose a novel generative approach in the framework of artificial immune model, by which we try to widen the bottleneck with effective search strategy embedded in the extracted state subspace. Firstly, we learn the latent space of pose state and propose a manifold reconstruction method to establish the inverse mapping. Pose analysis in this latent space is more effective and accurate. Secondly, we apply Clone Selection Algorithm (CSA) for human pose estimation. In order to make CSA suitable for pose tracking, we propose a sequential CSA (S-CSA) framework by incorporating the temporal continuity information into the traditional CSA. Experimental results show that our method achieves better results than state-of-art methods.

Keywords Human Motion analysis · Manifold learning · Clonal selection algorithm

1 Introduction

Recovering 3D human poses from marker-free images is a highly active research area due to the number of potential applications and its inherent complexity. However, although having been attacked by many researchers, this challenging problem is still long standing. In general, approaches to vision-based human pose analysis can be broadly divided into two categories: generative and discriminative [1]. Discriminative methods attempt to learn a direct mapping from image features to 3D pose from training data. The mapping is often approximated using nearest neighbor [2] or regression models [3]. While effective and fast, they are inherently limited by the amount and quality of the training data. Moreover, the relationship between image feature and the human pose in often multimodal which makes it difficult to build the mapping accurately. In contrast, generative methods exploit the fact that although the mapping from visual features to poses is complex and multimodal, the reverse mapping is often well-posed. Therefore, pose recovery is tackled by optimizing an object function that encodes the pose-feature correspondence [4], or by sampling posterior pose probabilities [5]. Compared with discriminative methods, generative methods are usually more accurate, but they suffer a large computational load, especially in high-dimensional state space. Moreover, optimization method and initialization are also the bottlenecks of the approach especially in tracking situation.

In the previous work, several possible strategies have been proposed for reducing the dimensionality of the configuration space, including using motion models [6], hierarchical search [7] and dimensionality reduction [4, 8]. Of which, dimensionality reduction method gains more interest, as it can reduce the dimensionality and extracting the prior knowledge of human motion simultaneously. However, linear methods as PCA [4] is often inadequate, since the mapping between the original pose space and the latent space is in general non-linear. Although non-linear methods as manifold learning [8] can learn this non-linear mapping, they are not invertible. How to build the smooth bidirectional mappings is still a well-solved problem. Optimization method is another key research problem of generative pose tracking methods, which is typically tackled using either deterministic methods or stochastic methods. Deterministic methods, as gradient descent search [9] are usually computationally efficient but they easily become trapped in local minima. In contrast, stochastic methods, as particle filter [5] are usually more robust, but they suffer a large computational load, especially in high-dimensional state space. In recent years, evolutional computing methods, such as genetic algorithm [4] and particle swarm optimization [7] have received increasing attention. Although considerable work has already been done above, a more effective optimization method is still intensively needed for robust visual tracking.

In this paper, we propose a novel generative approach. The main contributions of this paper are as follows. Firstly, we perform human pose analysis in latent space of pose state learned by ISOMAP, where the inverse mapping is established by a manifold reconstruction method. Secondly, we design Clone Selection Algorithm (CSA) and Sequential CSA for pose estimation and tracking. To the best of our knowledge, the proposed algorithm is new in human motion tracking literature. Experimental results show that our method achieves better results than state-of-art methods.
2 State space analysis

We use ISOMAP [10] to learn the low-dimensional latent space. Then a manifold reconstruction method is proposed to establish the inverse mapping.

ISOMAP-based manifold learning: we use motion capture data from CMU [11] for latent space learning. The subspace learned by ISOMAP is shown as Fig.1. We can see that similar low-dim subspace can be extracted from the training sequences that belong to the same type of motions but performed by different subjects. And the training sequences corresponding to different type of motions produce different subspaces.

Fig.1 ISOMAP based dimensionality reduction results. (a) (b) are manifolds of two sequences of walking and running in 3D subspace respectively.

Manifold reconstruction method: Let the pose state space be $X \subset \mathbb{R}^n$ and the low-dim state space be $Y \subset \mathbb{R}^d$. Denote the mapping as: $f : \mathbb{R}^n \rightarrow \mathbb{R}^d$ and $g : \mathbb{R}^d \rightarrow \mathbb{R}^n$, let the set of input instances be $\{ \mathbf{x}_i | x_i \in X, i = 1,...,T \}$ and their corresponding points in the embedding space be $\{ \mathbf{y}_i | y_i \in Y, i = 1,...,T \}$. The ISOMAP is computed using $L_2$ distance and ε-Isomap neighbor. Assume the neighbors of point $\mathbf{x}_i$ are $\{ \mathbf{x}_j | j = 1,...,l \}$, where $l_i$ is the number of neighbors. And their corresponding points in the embedding space are $\{ y_j | j = 1,...,l \}$. Then the ISOMAP-based manifold reconstruction method can be described as follows:

Algorithm 1: ISOMAP-based manifold reconstruction

**Input:** The training data set $\{ \mathbf{x}_i | x_i \in X, i = 1,...,T \}$ and the parameter $\varepsilon$.

**Output:** Given $\mathbf{x}_o \in X$, output the corresponding low-dim point $\mathbf{y}_o \in \mathbb{R}^d$. Given $\mathbf{y}_o \in \mathbb{R}^d$, output the corresponding high-dim point $\mathbf{x}_o \in X$.

**Step 1:** (preparing)

(1) use the ISOMAP method to compute the low-dim vector $\{ y_i | y_i \in Y, i = 1,...,T \}$ for the original input vector $\{ x_i | x_i \in X, i = 1,...,T \}$;

(2) construct the following matrixes: $X_i = (x_i - x_1,...,x_i - x_T) \in \mathbb{R}^{m \times d}$, $Y_i = (y_i - y_1,...,y_i - y_T) \in \mathbb{R}^{d \times d}$, where $\{ x_i | j = 1,...,l_i \}$ is the $\varepsilon$-neighbor of $x_i$;

(3) compute $Q = X_i Y_i^T (Y_i Y_i^T)^{-1} \in \mathbb{R}^{m \times d}$, where $(Y_i Y_i^T)^{-1}$ is the generalized inverse matrix of $Y_i Y_i^T$;

**Step 2:** (manifold mapping)

(1) given a high-dim pose vector $\mathbf{x}_o$, the corresponding low-dim vector $\mathbf{y}_o$ can be computed as:

(1.1) find the nearest neighbor of $\mathbf{x}_o$ in $\{ \mathbf{x}_i | i = 1,...,l \}$, set it to be $\mathbf{x}_s$;

(1.2) compute $\mathbf{y}_o = \mathbf{g}(\mathbf{x}_o) = \mathbf{y}_s + Q \mathbf{r} (\mathbf{x}_o - \mathbf{x}_s)$;

(1.3) output $\mathbf{y}_o$.

(2) given a high-dim pose vector $\mathbf{y}_o$, the correspondence low-dim vector $\mathbf{x}_o$ can be computed as:

(2.1) find the nearest neighbor of $\mathbf{y}_o$ in $\{ y_i | i = 1,...,l \}$, set it to be $\mathbf{y}_s$;

(2.2) compute $\mathbf{x}_o = \mathbf{f}(\mathbf{y}_o) = \mathbf{x}_s + Q (\mathbf{y}_o - \mathbf{y}_s)$;

(2.3) output $\mathbf{x}_o$.

3 Clonal selection algorithm for pose estimation

Clonal selection algorithm (CSA) [12], as a novel evolution method, has been another hotspot succeeding genetic algorithms and particle swarm optimization for its success in solving pattern recognition and multimodal optimization problems. In this paper, we apply CSA for pose estimation.

3.1 Apply CSA for pose optimization

Our modified CSA for pose optimization can be summarized as follows: **Initialization:** create an initial random set of antibodies population $A_0$; **Immunity process:** generates more antibodies after iteration of selection, clone and mutation operations; **Update:** the antibodies that meet the affinity requirements will be added to the antibody population. Some details are discussed below.

**Encoding and initialization:** We represent the antibody as $g = (g_0, g_1, g_2, ..., g_n)$, where $n$ is the dimensionality of state vector. For our problem, each gene $g_i$ of the antibody corresponds to a component of the pose vector. Here, we use real encodings and represent the population as $A_0 = \{ g_0, g_1, g_2, ..., g_N \}$, where $N$ is the total number of antibodies. Moreover, we restrict every component of the pose vector to be in the scope $[\min(g_i), \max(g_i)]$, where the bound $\min(g_i)$ and $\max(g_i)$ are learned from the motion training data.

**Affinity measure:** For each antibody, an affinity measure needs to be computed to estimate how well a given antibody (pose) matches the observed images. Here we use the bi-directional likelihood proposed by Ref. [13].

**Immunity operators:** The optimization process can be formulated as:

$A_k \xrightarrow{r} C_k \xrightarrow{t} Y_k \xrightarrow{u} Z_k \cup A_k \xrightarrow{v} A_{k+1}$

Where $T_1$, $T_2$, $T_3$, $T_4$ are the select, clone, mutate and update operators respectively. We set numbers of antibodies $N = 60$, selected number of individuals $h = 40$, newly generated number of individuals $r = 5$ and maximum number of generations $K = 100$ respectively. Detailed introduction about CSA can be found in Ref. [12].

Based on the design above, the CSA based pose optimization algorithm can be described as follows.

**Algorithm 2:** CSA based pose optimization algorithm
(1) **Initialization:** generate the initial population with \( N \) antibodies, represented as \( A_0 = \{g_1^{1,0}, g_2^{2,0}, ..., g_N^{N,0}\} \).

(2) **Repeat:**

- While \((k < K)\) do
  - (2.1) Immunity process: perform select, clone, mutate operations on current population \( A_k \) according to the defined operators \( T_s, T_c, \) and \( T_m \).
  - (2.2) Update: update \( A_k \) with matured antibodies and randomly generated individuals according to defined operator \( T_u \). Thus, the next population \( A_{k+1} \) generate;
  - (2.3) \( k = k + 1 \).

End while

(3) **Output:** output converged populations \( A_K \).

### 3.2 Clonal selection algorithm for pose estimation

Pose estimation is the process to estimate articulated human pose from single image which can be formulated as an optimization process. We apply CSA for pose estimation. First, with the aim of both cutting the search space and determining the motion direction roughly, we incorporated the global motion process step [4] into the framework of CSA. Based on the proposed CSA pose optimization algorithm, the antibody with highest affinity in population \( A_k = \{y_1^{1,k}, y_2^{2,k}, ..., y_N^{N,k}\} \) will be selected to be the estimated pose. Fig. 2 is the process of pose estimation.

![Fig. 2](image)

**Fig. 2** The process of human pose estimation, where (a) is one video frame, (b) is the initialized poses, (c) (d) (e) are results with different times of iteration respectively.

### 4 Sequential clonal selection algorithm for pose tracking

![Flowchart](image)

**Fig. 3** Overview of the sequential CSA algorithm

In tracking applications, the data is typically a time sequence, and hence the task is essentially a dynamic optimization problem which distinguishes it from traditional optimization problems. We proposed a sequential CSA (S-CSA) based framework for human motion tracking. The flowchart of the S-CSA framework is shown in Fig.3. There are three major stages: automatically initialization, next frame propagation, CSA-based optimization.

#### Initialization of tracking:

We achieve the automatic initialization by determining the pose of the first frame using the CSA based human pose estimation algorithm.

#### Next-frame propagation:

Given the converged antibodies at frame \( t \), the antibodies in the next frame are initialized by sampling a Gaussian distribution[14] centered in the current best antibodies. We don’t incorporate any motion model here, which is motivated by two considerations: generality and the effectiveness of our CSA pose optimization.

#### CSA-based optimization:

Estimate pose of current frame based on the initialized antibodies.

### 5 Experimental results

**Experimental data:** The data for latent space training is from CMU database [11]. We quantitatively evaluate our method on synthesized image sequences as in Ref. [3]. We also give experimental results on real image sequences from Ref. [15], CMU database [11] and HumanEva [13].

**Evaluation measures:** We use the evaluation measures proposed in Ref.[13]. The average error over all joint angles (in degrees) and the standard deviation of the performance are computed.

**CSA based pose estimation:** We test our pose estimation method on 100 frames of images of three types of motions. The mean errors of joint angle are reported, as shown in Fig. 4. The mean errors of most joints for three sequences are less than 5 degrees. Our results are competitive with others reported in the related literatures.

![Mean Errors](image)

**Fig.4** The mean errors of individual joint angle for different sequences.

We can see that: on most of the frames, the occlusion and left–right confusion problems are tackled by searching the
optimal pose in the extracted subspace. And the pose estimator is view invariant, mainly because of the special step for searching the global motion. In addition, the experiment results on walking and running sequence demonstrate that our algorithm is efficient for different types of motions.

**S-CSA based pose tracking results:** We compare S-CSA quantitatively with other tracking methods, include PF [5], PSO [7] and GA [4]. The mean errors of different methods over all joint angles of the test sequences are shown in Fig.5. Table 1 is the statistics of the average errors and the standard deviations. We can see that our method achieves better results. The average errors and the standard deviations over all frames are near 3° and 1°, respectively, in general. These indicate that our method can achieve stable tracking of 3D human pose.

![Fig. 5 Comparison of different tracking methods](image)

**Table 1 Results of different tracking methods**

<table>
<thead>
<tr>
<th></th>
<th>Walking</th>
<th></th>
<th>Running</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean error</td>
<td>Standard deviations</td>
<td>Mean error</td>
<td>Standard deviations</td>
</tr>
<tr>
<td>PF</td>
<td>4.6713</td>
<td>2.5157</td>
<td>4.4669</td>
<td>2.0188</td>
</tr>
<tr>
<td>PSO</td>
<td>4.4369</td>
<td>1.5181</td>
<td>4.3949</td>
<td>0.9821</td>
</tr>
<tr>
<td>PCA+GA</td>
<td>3.5705</td>
<td>1.5651</td>
<td>4.1494</td>
<td>1.4779</td>
</tr>
<tr>
<td>S-CSA</td>
<td>2.5217</td>
<td>1.2377</td>
<td>2.8313</td>
<td>1.2941</td>
</tr>
</tbody>
</table>

![Fig. 6 Human tracking results on real image sequences](image)

**6 Conclusions**

We propose a novel generative approach for human pose analysis form monocular image sequences. The proposed manifold reconstruction method can establish smooth inverse mapping, which enable us to perform pose analysis in the latent space. In the search strategy, we apply Clone Selection Algorithm for human pose estimation. And a sequential CSA framework is proposed for tracking by incorporating the temporal continuity information into the traditional CSA. Experimental results show that our CSA based pose estimation method can achieve viewpoint invariant 3D pose reconstruction and the S-CSA based tracking method can achieve accurate and stable tracking of 3D human motion.

**Acknowledgements** This paper is supported by the grants from “the National Natural Science Foundation of China” [Grants No.61021062 and 61100110], “the Natural Science Foundation of Jiangsu Province” [Grant No.BK2009230 and BK2010375], “Key Technology R&D Program of Jiangsu Province” [Grant No.BE2010072 and BE2011058].

**References**

11. CMU database. http://mocap.cs.cmu.edu/