

文章编号 1004-924X(2014)10-2870-09

## 采用改进粒子群优化粒子滤波的三维人手跟踪

李东年<sup>1,2</sup>, 周以齐<sup>1,2\*</sup>

- (1. 山东大学 机械工程学院 虚拟工程研究中心, 山东 济南 250061;
2. 山东大学 高效洁净机械制造教育部重点实验室, 山东 济南 250061)

**摘要:**针对高维人手状态空间中的采样稀疏问题,提出了一种基于改进粒子群优化粒子滤波的关节人手跟踪方法,用于从 Kinect 获取的深度图像序列中恢复三维人手运动。首先,利用简单几何基元建立三维人手模型,并为其添加自由度节点,用于在跟踪过程中生成可与观测特征进行比较的人手姿势假设。然后,在粒子滤波框架下,使用深度图像作为观测输入,融合深度特征与区域特征建立了系统观测模型。最后,将粒子群优化粒子滤波应用于关节人手运动跟踪。为避免在高维空间中的早熟收敛问题,利用模拟退火思想和局部随机化方法对算法进行改进,增强了算法的全局搜索能力。通过合成序列和真实序列上的跟踪实验对该方法进行了评价,结果表明该方法的关节角度跟踪误差均值约为  $2.3^\circ$ ,标准差约为  $1.7^\circ$ ,优于标准粒子滤波和标准粒子群优化跟踪方法,可以准确、鲁棒地从深度图像跟踪三维人手运动。

**关键词:**三维人手跟踪;深度图像;粒子滤波;粒子群优化;模拟退火;局部随机化

**中图分类号:** TP391.4 **文献标识码:** A **doi:** 10.3788/OPE.20142210.2870

### Three dimensional hand tracking by improved particle swarm optimized particle filter

LI Dong-nian<sup>1,2</sup>, ZHOU Yi-qi<sup>1,2\*</sup>

- (1. Virtual Engineering Research Center, School of Mechanical Engineering, Shandong University, Jinan 250061, China;
  2. Key Laboratory of High Efficiency and Clean Mechanical Manufacture of the Ministry of Education, Shandong University, Jinan 250061, China)
- \* Corresponding author, E-mail: yqzhou@sdu.edu.cn

**Abstract:** To overcome the difficulty of dense sampling in a high-dimensional hand state space, an improved Particle Swarm Optimized Particle Filter (PSOPF) algorithm was proposed to track articulated hand motion from single depth images obtained by a Kinect sensor. Firstly, a 3D hand model was built with basic geometric primitives, and the nodes with degrees of freedom (DOFs) were added into the model to generate the hand pose hypotheses to compare with the observation feature in the tracking process. Then, the single depth images were used as the only input, and the system observation model was established by combining depth features and silhouette features in the framework of particle filter. Finally, the PSOPF was applied to articulated hand tracking. To avoid the premature conver-

收稿日期:2014-03-24;修订日期:2014-05-29.

基金项目:山东省科技攻关计划资助项目(No. 2010G0020233)

gence in the high-dimensional space, the global search ability of the algorithm was improved by applying simulated annealing and partial randomization on the particles. Experiments were conducted both on synthetic data and real sequences for evaluation of the proposed method. It shows that the average of the joint angle errors for proposed method is about  $2.3^\circ$  and the standard deviation is about  $1.7^\circ$ , which are better than those of the standard Particle Filter(PF) and the standard Particle Swarm Optimization (PSO) method. The results show that the proposed method may track 3D articulated hand motion accurately and robustly from single depth images.

**Key words:** 3D hand tracking; depth image; particle filter; particle swarm optimized; simulated annealing; partial randomization

## 1 引 言

基于计算机视觉的三维人手跟踪是一个具有挑战性的研究课题,可以应用于许多领域,如运动捕捉、人机交互、手势识别、机器人示教学习等。由于人手具有多个自由度,该问题本质上是一个高维空间中的运动跟踪问题,其开发受困于多个复杂因素,如高维空间问题、人手运动过程中的自遮挡问题、跟踪失败后的自动恢复问题等。目前,基于视觉的三维人手跟踪方法通常分为两类:基于表观的方法和基于模型的方法。

基于表观的方法<sup>[1-5]</sup>是通过学习来建立一个映射,将图像特征空间映射到人手状态空间,从而直接从图像特征估计出人手状态。该类方法不需要初始化,跟踪速度快,但其精确度受训练样本的影响;基于模型的方法<sup>[6-11]</sup>使用预先建立的三维人手模型生成人手姿势假设,将从模型中提取的特征与从视觉观测提取的特征进行比较,评价二者的相似度,通过某种优化方法在模型状态空间中搜索出一组具有最佳相似度的人手状态。该类方法可以利用较多的先验信息(如人手形状、关节约束等),但其跟踪过程需要初始化,并且面临在高维空间中搜索的难题。

近些年发展起来的粒子滤波(Particle Filter, PF)算法已成为解决非线性、非高斯动态系统参数估计与状态滤波的主流方法<sup>[12-14]</sup>,许多研究人员用 PF 框架实现基于模型的三维人手<sup>[6-7]</sup>或人体<sup>[8]</sup>运动跟踪。标准 PF 需要大量样本来逼近系统的真实后验概率密度分布,尤其对于人手运动跟踪这样的高维问题,由于高维空间中粒子采样的极度稀疏性,难以使用有限个数粒子来有效表达人手状态的真实后验分布,很容易导致跟踪

失败。针对 PF 高维跟踪困难的问题,Bray 等<sup>[6]</sup>通过将随机梯度下降方法与 PF 结合起来跟踪三维人手运动,在 PF 框架基础上,以动态模型提供的预测粒子为初始值,采用随机梯度下降对样本进行优化来指导重要性分布函数的生成。吕治国等<sup>[7]</sup>采用层次遗传算法来提高粒子采样的优化效率,并使用最优解附近采样的粒子和预测模型中采样的粒子一起表达当前时刻的真实后验分布,在隐状态空间中跟踪三维人手运动。

粒子群优化(Particle Swarm Optimization, PSO)算法<sup>[5]</sup>是一种基于群体智能理论的新兴演化计算方法,文献[15]将 PSO 算法引入 PF,提出了粒子群优化粒子滤波(Particle Swarm Optimized Particle Filter, PSOPF)算法,并用其解决移动机器人的跟踪与定位问题。该算法采用一种改进的 PSO 算法——高斯 PSO 算法<sup>[16]</sup>来优化 PF 的采样过程,驱动所有粒子向高似然区域运动,从而克服了粒子贫乏问题,可以使用少量的粒子实现鲁棒跟踪。本文将 PSOPF 算法用于三维人手运动跟踪,并采用模拟退火思想和局部随机化方法<sup>[17]</sup>对其改进,以克服早熟收敛问题。

在观测方面,近年来,以 Kinect 为代表的一系列深度相机的出现使研究人员能够更容易地获取准确、鲁棒的深度图像,而深度图像的使用有利于避免二维图像的歧义性问题和观测尺度问题,许多研究人员开始采用深度图像来研究人手的运动估计问题<sup>[1-3,10-11]</sup>。Keskin 等<sup>[1-2]</sup>建立了一个概率决策树,即通过每像素分类估计人手部位的近似位置,其方法依赖于预先进行的繁重的训练。Oikonomidis 等<sup>[10]</sup>采用基于模型的单假设跟踪方法,结合颜色信息与深度信息跟踪三维人手运动。根据本文应用需要,本文研究了单一深度信息下的人手运动跟踪,只采用单一深度信息作为观测

输入,并没有使用颜色信息。由于 Kinect 获取的深度信息比较粗糙,这样虽然增加了对人手运动进行准确跟踪的难度,但却增强了系统对环境的适应性,避免了系统对光照条件和背景颜色的限制。

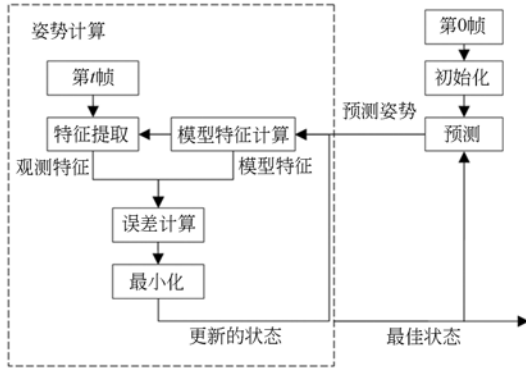


图 1 基于模型的人手运动跟踪系统框图

Fig. 1 Block diagram of model-based hand motion tracking

## 2 跟踪系统整体框架及人手模型

### 2.1 跟踪系统整体框架

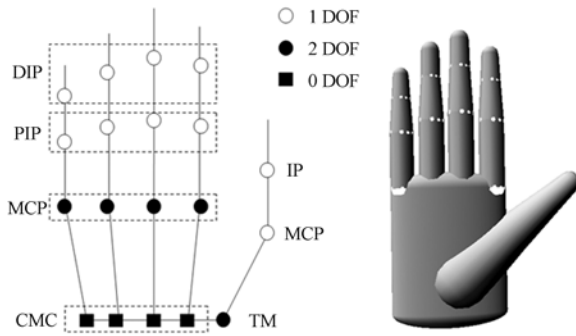
本文采用基于模型的方法进行人手运动跟踪。该方法的通用系统框图<sup>[18]</sup>如图 1 所示。该方法需要预先构造一个状态可配置的人手几何模型,然后在跟踪过程中利用此模型生成人手姿势假设,再计算模型特征组和输入图像特征组之间的匹配误差,最后,在人手状态空间中搜索使匹配误差最小化的状态参数,即为该帧对应的最优解。

### 2.2 人手模型的建立

在参数化三维建模软件 PTC Pro/Engineer 中以简单几何单元建立三维人手的几何模型,通过中间格式 .obj 将其导入可视化建模软件 Multigen-Paradigm Creator,在 Creator 中为人手模型建立树状层次组织结构并为其添加 DOF 自由度节点。

图 2(a)和图 2(b)分别为人手运动学模型和几何模型。如图 2(a)所示,本文采用的三维人手模型共 26 个自由度,包括 6 个全局自由度和 20 个局部自由度。各手指的 CMC 关节固定,手掌作为一个刚体建模,其运动对应人手的 6 个全局自由度(3 个平移和 3 个旋转),由掌心的一个固定点表示。5 个手指的运动对应 20 个局部自由

度,是本文研究的重点。除拇指外,各手指的 MCP 关节均包含 2 个自由度(1 个屈伸和 1 个外展内收)。而拇指的 MCP 关节只包含 1 个屈伸自由度。拇指的 IP 关节,其余 4 个手指的 PIP 和 DIP 关节均各含 1 个屈伸自由度。另外,拇指的 TM 关节包含 2 个自由度。各手指关节角度取值被限定在一定范围内。



(a) 运动学模型

(b) 几何模型

(a) Kinematical model

(b) Geometric model

图 2 人手运动学模型和几何模型

Fig. 2 Kinematical and geometric models for hand

## 3 观测模型

采用深度特征信息与区域特征信息融合来构建观测似然函数。以 Kinect 深度相机获取的深度图像作为观测输入,通过简单的深度阈值分割将人手对应区域提取出来,生成深度图  $z_d$ 。对于每一个人手姿势假设  $\mathbf{x}$ ,在给定深度相机校准信息的情况下,使用图形渲染手段生成对应的深度图  $r_d$ 。然后,由观测得到的深度图像  $z_d$  生成人手区域图  $z_s$ ,由渲染得的深度图像  $r_d$  生成人手区域图  $r_s$ 。 $z_s$  和  $r_s$  均为二值图,在前景人手区域处取值为 1,在背景处取值为 0。

通过匹配误差函数来表达观测  $\mathbf{z} = \{z_d, z_s\}$  和人手姿势假设  $\mathbf{x}$  之间的匹配度,小的匹配误差意味着高的匹配度。本文采用的匹配误差函数定义如下:

$$E(\mathbf{z}, \mathbf{x}) = \lambda_d E_d(\mathbf{z}, \mathbf{x}) + \lambda_s E_s(\mathbf{z}, \mathbf{x}) + \lambda_m E_m(\mathbf{x}), \quad (1)$$

式中:  $E(\mathbf{z}, \mathbf{x})$  由深度特征项  $E_d$ 、区域特征项  $E_s$  和平滑项  $E_m$  三部分组成,  $\lambda_d$ 、 $\lambda_s$  和  $\lambda_m$  为各部分的常数权重因子,在本文实验中,分别取为 1.0、2.22 和 0.005。

$E_d$  度量观测所得深度图  $z_d$  和人手姿势假设

$\mathbf{x}$  所对应合成深度图  $r_d$  之间的深度偏差,其定义如下:

$$E_d(z, x) = \frac{\sum \min(|z_d - r_d|, T_d)}{\sum (z_s \vee r_s)}.$$

深度偏差(以 mm 为度量单位)在整个特征图上逐像素计算,并进行累加,将累加和除以人手像素区域的总面积来进行规范化。某些大深度偏差会引起函数值发生较大的变化,从而影响优化搜索方法的性能。因此,引入最大深度偏差常量  $T_d$  将各像素上深度偏差的范围限定为  $[0, T_d]$ 。在本文实验中,  $T_d$  取值为 40 mm。

$E_s$  通过计算观测区域图  $z_s$  与渲染区域图  $r_s$  之间不重叠区域的大小来描述区域特征匹配度,其定义如下:

$$E_s(z, x) = \frac{\sum z_s(1 - r_s)}{\sum z_s} + \frac{\sum r_s(1 - z_s)}{\sum r_s}.$$

区域特征项  $E_s$  对目标函数起到了平滑作用,减少了全局最小值周围的局部极小值,从而使优化过程更好地收敛到实际全局最小值,增强了优化过程的鲁棒性。

为避免相邻两帧人手姿势的突变,为匹配误差函数  $E(z, x)$  增加了一个平滑项  $E_m$  来平滑恢复后的人手姿势在时域上的运动,其定义如下:

$$E_m(\mathbf{x}_t) = \|\mathbf{x}_t - \hat{\mathbf{x}}_{t-1}\|,$$

其中:  $\mathbf{x}_t$  为当前帧的人手姿势假设;  $\hat{\mathbf{x}}_{t-1}$  为上一帧所恢复得的人手姿势。

观测似然函数与匹配误差函数  $E(z, x)$  之间呈单调递减关系,将观测似然函数定义如下:

$$p(z|\mathbf{x}) \propto \exp(-\lambda_e \cdot E(z, x)), \quad (2)$$

其中:  $\lambda_e$  为常数规范化因子,其取值由观测噪声决定。

## 4 跟踪算法

### 4.1 PF 算法概述

PF 算法是一种具有强鲁棒性的运动跟踪框架,因其通过多个样本在时间上的传播,具有表达多峰分布的特性,其基本思路为:根据  $t-1$  时刻系统状态后验概率分布  $p(\mathbf{x}_{t-1} | \mathbf{z}_{1:t-1})$  的加权样本值  $\{\mathbf{x}_{t-1}^i, \mathbf{w}_{t-1}^i\}_{i=1, \dots, N}$ , 利用系统状态预测模型  $p(\mathbf{x}_t | \mathbf{x}_{t-1})$  和观测似然  $p(z_t | \mathbf{x}_t)$ , 寻找一组逼近  $t$  时刻系统状态后验概率分布  $p(\mathbf{x}_t | \mathbf{z}_{1:t})$  的加权样

本值  $\{\mathbf{x}_t^i, \mathbf{w}_t^i\}_{i=1, \dots, N}$ 。其中,上标  $i$  为粒子序号,  $\mathbf{x}_t$  为  $t$  时刻系统状态矢量,在本文中表示  $t$  时刻人手姿势参数;  $\mathbf{w}_t$  为  $\mathbf{x}_t$  对应权值;  $\mathbf{z}_{1:t}$  为系统从 1 时刻累计到  $t$  时刻的观测值。

标准 PF 算法的一个主要问题是采用先验模型  $p(\mathbf{x}_t | \mathbf{x}_{t-1})$  作为重要密度函数,而  $p(\mathbf{x}_t | \mathbf{x}_{t-1})$  没有考虑最新观测值  $z_t$ 。因而粒子的重要性采样过程是次优的。对于三维人手运动跟踪这样的高维问题,在可计算的前提下,由于高维空间中粒子采样的极度稀疏性,难以使用有限个粒子来有效表达人手状态的真实后验分布,很容易导致跟踪失败。

### 4.2 PSOPF 算法

文献[15]将高斯 PSO 算法融入 PF 来优化 PF 的采样过程,提出了 PSOPF 算法。高斯 PSO 算法<sup>[16]</sup>是一种改进的 PSO 算法,它采用高斯分布来生成速度矢量,收敛性较经典的 PSO 算法<sup>[5]</sup>好,其粒子速度、位置更新方程如下:

$$\mathbf{v}_{k+1}^i = |\text{randn}|(\mathbf{p}_k^i - \mathbf{x}_k^i) + |\text{randn}|(\mathbf{g}_k - \mathbf{x}_k^i), \quad (3)$$

$$\mathbf{x}_{k+1}^i = \mathbf{x}_k^i + \mathbf{v}_{k+1}^i, \quad (4)$$

其中:  $|\text{randn}|$  为正的高斯分布随机数,可由标准高斯分布  $N(0, 1)$  的绝对值产生;向量  $\mathbf{x}_k^i$  和向量  $\mathbf{v}_k^i$  分别为第  $i$  个粒子在第  $k$  代的位置和速度;向量  $\mathbf{p}_k^i$  为个体历史最优解,其为当前迭代次数  $k$  内第  $i$  个粒子所经历过的最优位置;向量  $\mathbf{g}_k$  为全局最优解,用于储存整个群体到目前为止找到的最优解。

文献[15]用 PSOPF 算法解决移动机器人的跟踪与定位问题。通过将最新观测值引入待优化目标函数,采用高斯 PSO 算法对由预测模型采样得到的粒子进行演化,在权值更新前使其更加趋向于高似然概率区域,如图 3 所示。

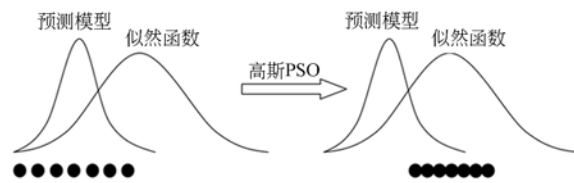


图 3 PSOPF 算法的采样优化过程

Fig. 3 Sampling optimization process of PSOPF

### 4.3 基于改进 PSOPF 的人手运动跟踪算法

本文将 PSOPF 算法应用于高维状态空间的

人手运动跟踪,采用高斯 PSO 算法来优化第 3 节中式(1)所示的匹配误差函数  $E(\mathbf{z}, \mathbf{x})$ 。在帧间状态转移过程中,采用  $t-1$  时刻收敛后个体的最优位置来初始化  $t$  时刻的粒子:

$$\mathbf{x}_{t,0}^i = \mathbf{p}_{t-1}^i + \mathbf{r}, \quad (5)$$

其中:  $\mathbf{r} \sim N(0, \Sigma)$  为 0 均值多维高斯噪声,  $\Sigma$  为协方差矩阵,其对角线元素根据最大帧间角度或位移变化来确定。

然而,高斯 PSO 算法同标准 PSO 算法一样在搜索过程的中后期容易陷入局部最优,产生早熟收敛现象。对于人手运动跟踪这样的高维多峰问题,此问题尤为突出。本文同时采用两种改进方法来克服算法的早熟收敛问题,提高粒子多样性,以增强算法的全局搜索能力。

首先,将模拟退火思想引入高斯 PSO,采用模拟退火算法对粒子个体极值的更新条件进行改进,使其既接受优化解,也可以以一定概率接受恶化解,从而扩大全局搜索范围。如下式所示:

$$\mathbf{p}_{k+1}^i = \begin{cases} \mathbf{x}_{k+1}^i & \text{if } r^i < \min(1, \exp(-\Delta D/T_{k+1})) \\ \mathbf{p}_k^i & \text{otherwise} \end{cases}, \quad (6)$$

其中:  $T_{k+1}$  为第  $k+1$  次迭代的退火温度,  $T_{k+1} = \alpha T_k$ ,  $\alpha$  为降温系数且  $\alpha \in (0, 1)$ , 在本文的所有实验中,  $\alpha$  均取 0.8;  $\Delta D$  为粒子失真变化,  $\Delta D = 1/f(\mathbf{x}_{k+1}^i) - 1/f(\mathbf{p}_k^i)$ ,  $f(\mathbf{x}_{k+1}^i)$  和  $f(\mathbf{p}_k^i)$  分别为新粒子适应值和当前个体最优适应值,在本文中  $f(\mathbf{x}) = p(\mathbf{z}|\mathbf{x})$ , 即粒子的适应值  $f(\mathbf{x})$  采用 3.2 节中式(2)定义的观测似然  $p(\mathbf{z}|\mathbf{x})$  计算。  $r^i$  为在区间  $[0, 1]$  上均匀分布的随机数。

其次,从高斯 PSO 第 3 次迭代开始,采用文献[17]中的方法在手指关节角度相应的维度上执行局部随机化处理。更具体地说,每次迭代过程中,在进行完粒子位置更新后,以一个小概率对粒子位置进行重新取值,所取值为在对应维度取值范围内的均匀采样。如下式所示

$$x_k^{ij} = \begin{cases} U(x^{j,\min}, x^{j,\max}) & \text{if } \text{rand}^{ij} < P_r \\ x_k^{ij} & \text{otherwise} \end{cases}, \quad (7)$$

其中:  $x^{j,\min}$  和  $x^{j,\max}$  分别为搜索空间第  $j$  维所限定的最小取值、最大取值;  $U(x^{j,\min}, x^{j,\max})$  则为在区间  $[x^{j,\min}, x^{j,\max}]$  均匀分布的随机数;  $\text{rand}^{ij}$  等于  $U(0, 1)$ , 为区间  $[0, 1]$  上均匀分布的随机数;  $P_r$  为对粒子进行随机化处理的概率阈值常量,在本文实验中取为 0.01。

改进 PSOPF 的算法步骤具体可描述如下:

Step1: 初始化。从先验分布  $p(\mathbf{x}_0)$  采样得到  $N$  个粒子, 权值均为  $1/N$ , 表示为  $\{\mathbf{x}_0^i, 1/N\}_{i=1, \dots, N}$ 。

Step2: 粒子状态转移, 即重要性采样过程。通过式(5)获取粒子初始位置, 将最新观测值引入待优化目标函数中, 采用式(3)和式(4)对粒子集进行迭代演化, 驱动粒子向高似然概率区域运动, 同时采用式(6)和式(7)提高粒子的多样性, 避免早熟收敛。

Step3: 权值更新。利用观测似然更新粒子权值  $w_t^i \propto w_{t-1}^i p(\mathbf{z}_t | \mathbf{x}_t^i)$ , 并对权值进行归一化。以最大后验准则输出系统状态估计值。

Step4: 重采样。为避免粒子权值退化, 根据权值大小对样本集  $\{\mathbf{x}_t^i, w_t^i\}_{i=1, \dots, N}$  进行重采样, 得到新的等权样本集  $\{\mathbf{x}_t^i, 1/N\}_{i=1, \dots, N}$ 。

Step5: 判断是否结束。若是则退出本算法, 否则转到 Step2。

## 5 实验结果

本文实验系统是在 Visual Studio 2010 平台上采用开源跨平台三维图形引擎 OpenScene-Graph(OSG) 3.0.0 开发的。将带有 DOF 自由度节点的三维人手模型加载到 OSG 中, 通过 `osgSim::DOFTransform` 类来控制各自由度的运动。另外, 每次迭代时, 在 OSG 中使用帧缓存对象(FBO)渲染生成人手模型的深度图像, 用于算法中各粒子匹配误差和观测似然值的计算。

本文分别在合成序列和真实序列上进行实验, 以验证本文基于改进 PSOPF 的人手运动跟踪方法的有效性, 同时将本文方法与标准 PF 方法、采用标准 PSO<sup>[18]</sup> 的单假设跟踪方法进行了比较。为公平起见, 在所有实验中, 本文方法与标准 PSO 方法均采用 60 个粒子, 且对于每帧图像输入, 均迭代优化 30 次, 而标准 PF 方法采用 1 800 个粒子。对于这 3 种方法, 每跟踪一帧图像均需要进行 1 800 次的匹配误差计算。由于匹配误差计算是跟踪过程中最主要的耗时操作, 故 3 种方法的跟踪速度基本相同。实验在双核 Core 2 2.0 GHz CPU, 2.0 GB 内存和 Nvidia GeForce 8400M GS GPU 的普通 PC 机上运行, 在关闭垂直同步的情况下, 跟踪一帧图像平均耗时 5 s。

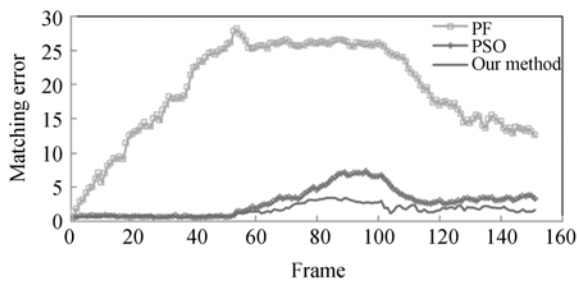
### 5.1 合成序列实验

从相机拍摄的图像序列直接获取真实数据是非常困难的,故本文采用附带真实数据的合成序列来对算法进行量化评价。首先,定义几个关键姿态,图 4 为其直观表达;然后,在关键姿态间以一定的步长进行线性插值,生成中间各时刻的姿态,并通过三维模型渲染的方式合成各时刻姿态对应的深度图像。本文所采用的合成图像序列共 150 frame,包含连续的 3 段动作,每段动作为 50 frame。



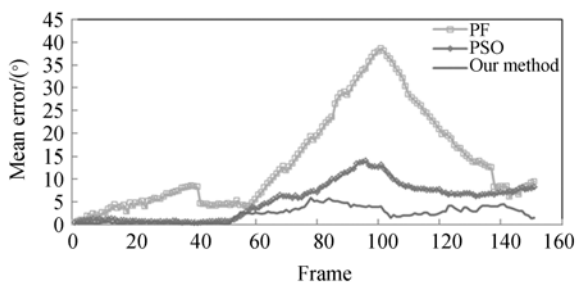
图 4 合成序列的关键姿态

Fig. 4 Key poses of synthetic sequence



(a) 匹配误差比较

(a) Comparison of matching errors



(b) 关节角度平均误差比较

(b) Comparison of average errors for joint angles

图 5 合成序列跟踪误差比较

Fig. 5 Comparison of tracking errors on synthetic sequence

将此合成序列作为跟踪系统的输入,进行三维人手跟踪实验。图 5 给出了 3 种方法的跟踪误差比较结果。其中,匹配误差采用第 3 节中式(1)

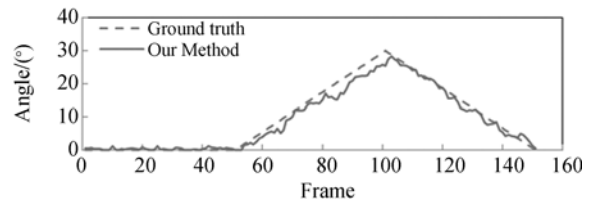
求取,而关节角度平均误差则为各关节角度误差的平均值,表 1 给出了在整个序列上的统计分析结果。

表 1 关节角度误差统计和比较

Tab. 1 Statistic and comparison of joint angle errors (°)

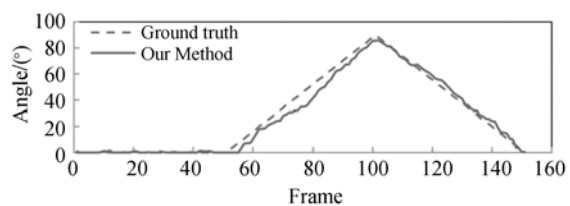
方法	误差均值	标准差	最大值
PF	13.833 2	10.772 9	38.490 4
PSO	5.301 5	4.021 1	13.976 9
本文方法	2.290 3	1.660 3	5.930 1

由图 5 和表 1 可以看出,标准 PF 方法的跟踪误差最大,已不能对高维空间中的人手运动进行有效跟踪。图 5(b)中的 PF 误差曲线出现了峰值,这说明整个跟踪过程中,标准 PF 方法跟踪得到的手指姿态一直在初始姿态附近浮动,致使当合成序列中的手指开始动作后,标准 PF 方法的跟踪误差不断累积,直至达到 38.49° 的最大值。



(a) 无名指 MCP 关节屈伸自由度

(a) MCP flexion of ring finger



(b) 中指 PIP 关节屈伸自由度

(b) PIP flexion of middle finger

图 6 合成序列跟踪结果与真实值的比较

Fig. 6 Comparison of our estimates and ground truth

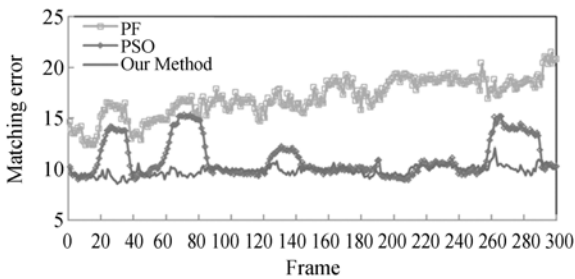
本文方法和标准 PSO 方法的跟踪结果明显优于标准 PF 方法,而本文方法具有比标准 PSO 方法更小的跟踪误差。由表 1 可以看出,本文方法的关节角度误差均值约为 2.3°(标准差约为 1.7°),而标准 PSO 方法关节角度误差均值约为 5.3°(标准差约为 4.0°),本文方法在跟踪精度上较其提高了 3°左右。图 6 所示为部分姿态参数的

跟踪结果,其中,实线为本文方法的跟踪结果,虚线为真实值。由图 6 可以看出,本文方法跟踪结果与真实值之间保持了高度的一致性。

### 5.2 真实序列实验

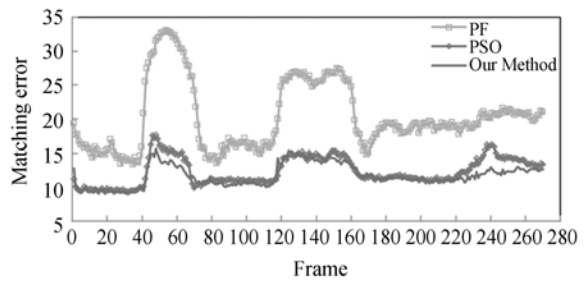
本文也采用真实序列对算法进行评价,使用 Microsoft Kinect 1.0 Beta2 SDK 捕获的人手深度图像序列作为观测输入,图像分辨率为 640 × 480,帧速率为 30 frame/s。实验分为两组,第一组实验共选取 300 frame 图像,第二组实验共选取 270 frame 图像。

由于本部分实验无法获取真实数据,本文只比较了 3 种跟踪方法的匹配误差,如图 7 所示。



(a)第 1 组真实序列匹配误差

(a) Matching errors on the first real sequence



(b)第 2 组真实序列匹配误差

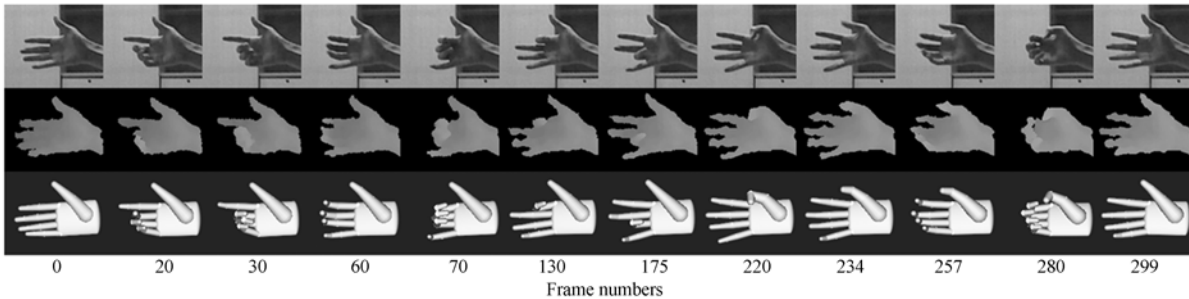
(b) Matching errors on the second real sequence

图 7 真实序列跟踪误差比较

Fig. 7 Comparison of tracking errors on real sequences

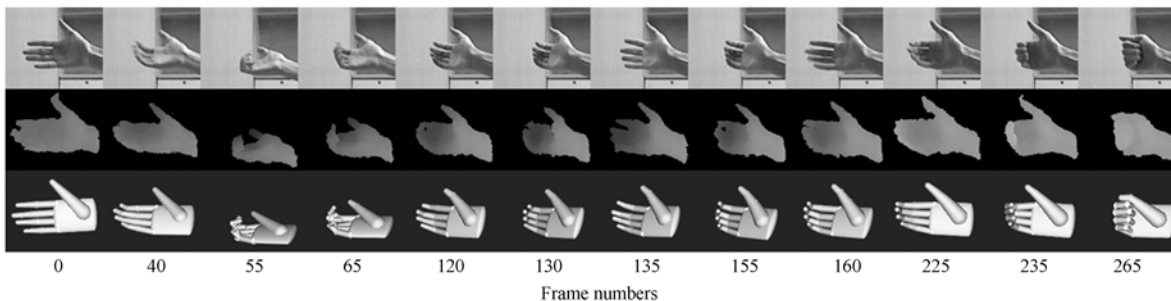
由图 7 可以看出,在对两组真实序列进行跟踪时,标准 PF 方法的匹配误差都最大,而本文方法和标准 PSO 方法的跟踪结果明显优于标准 PF 方法,但在一些手指动作较复杂的帧段,本文方法的表现要优于标准 PSO 方法,如第 1 组真实序列中的 20~40 frame、60~85 frame、260~290 frame 以及第 2 组真实序列中的 230~260 frame。

图 8 为两组真实序列部分帧上的跟踪结果。各组跟踪结果的第一栏为由 Kinect RGB 相机所



(a)第 1 组真实序列跟踪结果

(a) Results of the first real sequence



(b)第 2 组真实序列跟踪结果

(b) Results of the second real sequence

图 8 真实序列跟踪结果

Fig. 8 Results of real sequences

捕获的对应 RGB 图像,第二栏为由 Kinect 深度相机所捕获并进行简单深度分割后的深度图像,第三栏为采用本文方法对深度图像序列进行跟踪所得结果,图底为帧号。由实验结果可以看出,本文方法在 Kinect 比较粗糙的深度数据下仍可以准确鲁棒地跟踪人手运动。

## 6 结 论

人手状态空间的高维特性是三维人手运动跟踪所面临的主要问题之一,由于高维空间中粒子采样具有稀疏性,采用标准 PF 算法无法对多关节人手进行鲁棒跟踪。针对该问题,本文采用 PSOPF 算法进行人手运动跟踪,采用高斯 PSO 优化样本分布。为避免早熟收敛问题,本文同时采用模拟退火和局部随机化方法对算法进行改

进,增强了粒子多样性,改善了算法的全局搜索能力。另外,在 PF 框架下,本文使用 Kinect 传感器获取的深度图像作为观测输入,并通过深度特征与区域特征的融合建立了系统观测模型。实验表明,本文方法的关节角度跟踪误差均值约为  $2.3^\circ$ ,标准差约为  $1.7^\circ$ ,优于标准 PF 和标准 PSO 跟踪方法,可以准确鲁棒地从深度图像跟踪三维人手运动。

由于 Kinect 获取的深度信息比较粗糙,时常会出现由数据缺失造成的“孔洞”,对跟踪系统的性能影响较大。为进一步提高系统的鲁棒性,今后的工作将使用多个深度相机来保证深度图像的数据完整性。另外,针对跟踪系统中最耗时的匹配误差计算,将考虑根据其内在的并行计算特性和 PSOPF 算法的多粒子并行计算特性,使用 cuda 编程对系统进行 GPU 加速。

### 参考文献:

- [1] KESKIN C, KIRAÇ F, KARA Y E, *et al.*. Hand pose estimation and hand shape classification using multi-layered randomized decision forests [C]. *Proceedings of the 12th European Conference on Computer Vision, Berlin Heidelberg: Springer*, 2012, 6: 852-863.
- [2] KESKIN C, KIRAC F, KARA Y E, *et al.*. Real time hand pose estimation using depth sensors [C]. *Consumer Depth Cameras for Computer Vision, London: Springer*, 2013:119-137.
- [3] DOLIOTIS P, ATHITSOS V, KOSMOPOULOS D, *et al.*. Hand shape and 3D pose estimation using depth data from a single cluttered frame [C]. *Advances in Visual Computing, Berlin Heidelberg: Springer*, 2012, 1: 148-158.
- [4] PRISACARIU V A, REID I. 3D hand tracking for human computer interaction [J]. *Image and Vision Computing*, 2012, 30(3): 236-250.
- [5] KENNEDY J, EBERHART R. Particle swarm optimization [C]. *Proceedings of the IEEE International Conference on Neural Networks, Piscataway: IEEE Press*, 1995, 4:1942-1948.
- [6] BRAY M, KOLLER-MEIER E, VAN GOOL L. Smart particle filtering for 3D hand tracking [C]. *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, Los Alamitos: IEEE Computer Society Press*, 2004: 675-680.
- [7] 吕治国,李焱,徐昕. 快速的三维人手运动跟踪方法研究[J]. *计算机研究与发展*, 2012, 49(7): 1398-1407.
- [8] LÜ ZH G, LI Y, XU X. Research on fast 3D hand motion tracking system [J]. *Journal of Computer Research and Development*, 2012, 49(7): 1398-1407. (in Chinese)
- [9] ZHANG Z, SEAH H S, QUAH C K, *et al.*. GPU-accelerated real-time tracking of full-body motion with multi-layer search [J]. *IEEE Transactions on Multimedia*, 2013, 15(1): 106-119.
- [10] ZHANG Z, SEAH H S. Skeleton body pose tracking from efficient three-dimensional motion estimation and volumetric reconstruction [J]. *Applied Optics*, 2012, 51(23): 5686-5697.
- [11] OIKONOMIDIS I, KYRIAZIS N, ARGYROS A A. Efficient model-based 3D tracking of hand articulations using Kinect [C]. *Proceedings of the 22nd British Machine Vision Conference, Guildford: BMVA Press*, 2011: 101.1-101.11.
- [12] LIANG H, YUAN J S, THALMANN D, *et al.*. Model-based hand pose estimation via spatial-temporal hand parsing and 3D fingertip localization [J]. *The Visual Computer*, 2013, 29(6-8): 837-848.
- [12] 龚俊亮,何昕,魏仲慧,等. 采用改进辅助粒子滤

- 波的红外多目标跟踪[J]. 光学精密工程, 2012, 20(2): 413-421.
- GONG J L, HE X, WEI ZH H, *et al.*. Multiple infrared target tracking using improved auxiliary particle filter [J]. *Opt. Precision Eng.*, 2012, 20(2): 413-421. (in Chinese)
- [13] 吴新杰, 黄国兴, 王静文. 粒子滤波算法在 ECT 图像重建中的应用[J]. 光学精密工程, 2012, 20(8): 1824-1830.
- WU X J, HUANG G X, WANG J W. Application of particle filtering algorithm to image reconstruction of ECT [J]. *Opt. Precision Eng.*, 2012, 20(8): 1824-1830. (in Chinese)
- [14] 王国良, 刘金国. 基于粒子滤波的多自由度运动目标跟踪[J]. 光学精密工程, 2011, 19(4): 864-869.
- WANG G L, LIU J G. Moving object tracking with multi-degree-of-freedom based on particle filters [J]. *Opt. Precision Eng.*, 2011, 19(4): 864-869. (in Chinese)
- [15] 方正, 佟国峰, 徐心和. 粒子群优化粒子滤波方法[J]. 控制与决策, 2007, 22(3): 273-277.
- FANG ZH, TONG G F, XU X H. Particle swarm optimized particle filter [J]. *Control and Decision*, 2007, 22(3): 273-277. (in Chinese)
- [16] KROHLING R A. Gaussian swarm: a novel particle swarm optimization algorithm [C]. *Proceedings of the 2004 IEEE Conference on Cybernetics and Intelligent Systems*, Piscataway: IEEE Press, 2004, 1: 372-376.
- [17] YASUDA T, OHKURA K, MATSUMURA Y. Extended PSO with partial randomization for large scale multimodal problems [C]. *Proceedings of World Automation Congress*, Albuquerque: TSI Press, 2010: 1-6.
- [18] EROL A, BEBIS G, NICOLESCU M, *et al.*. A review on vision-based full DOF hand motion estimation [C]. *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Los Alamitos: IEEE Computer Society Press, 2005: 75-82.

#### 作者简介:



李东年(1987—),男,山东东营人,博士研究生,2009年于山东大学获得学士学位,主要从事机器视觉、图像处理、计算机图形学等方面的研究。E-mail: dongnianli@yeah.net

#### 导师简介:



周以齐(1957—),男,江苏扬州人,博士,教授,博士生导师,2002年于山东大学获得博士学位,主要从事自动控制、虚拟现实、虚拟工程等方面的研究。E-mail: yqzhou@sdu.edu.cn

(本栏目编辑:李自乐)  
(版权所有 未经许可 不得转载)