基于退火粒子群优化的单目视频人体姿态分析方法

李 毅 [1] 孙正兴 [1] 陈松乐 [1] 李 骞 [1]

摘 要 提出一种基于退火粒子群优化 (Simulated annealing particle swarm optimism, SAPSO) 的单目视频人体姿态分析方法. 该方法具有以下特点: 首先, 利用运动捕获数据采用主成分分析方法 (Principle component analysis, PCA) 得到更能反映人体运动本质的姿态紧致空间, 并在此低维空间中进行姿态分析, 提高了姿态分析的准确性和效率; 其次, 将粒子群优化应用到姿态分析中, 并提出退火粒子群优化姿态分析方法, 该方法具有良好的收敛性和全局最优能力; 再次, 基于退火粒子群优化姿态分析方法, 实现了基于单目视频的人体姿态估计和跟踪. 实验结果表明, 本文方法不仅具有良好的计算效率, 同时具有良好的收敛性和全局搜索能力, 能准确分析单目视频中的人体姿态.

关键词 姿态分析, 主成分分析, 模拟退火, 粒子群优化

引用格式 李毅, 孙正兴, 陈松乐, 李骞. 基于退火粒子群优化的单目视频人体姿态分析方法. 自动化学报, 2012, **38**(5): 732−741

DOI 10.3724/SP.J.1004.2012.00732

3D Human Pose Analysis from Monocular Video by Simulated Annealed Particle Swarm Optimization

LI Yi[1] SUN Zheng-Xing[1] CHEN Song-Le[1] LI Qian[1]

Abstract In this paper we proposed a simulated annealing particle swarm optimism (SAPSO) based method for human pose estimation form monocular image sequences. First, we use principle component analysis (PCA) to learn the low-dimensional compact space of human pose, by which the aim of both reducing dimensionality and extracting the prior knowledge of human motion are achieved simultaneously. Pose is estimated on the compact subspace. In the optimizing step, we introduce particle swarm optimism to human pose estimation, and further, a SAPSO pose estimation method is proposed. And last we use SAPSO to estimate and track human pose in monocular videos separately. Experimental results demonstrate that the proposed method is more convergent and globally optimum, which can estimate and track human pose in monocular images effectively.

Key words Pose estimation, principle component analysis (PCA), simulated annealing, particle swarm optimization (PSO)

Citation Li Yi, Sun Zheng-Xing, Chen Song-Le, Li Qian. 3D Human pose analysis from monocular video by simulated annealed particle swarm optimization. *Acta Automatica Sinica*, 2012, **38**(5): 732−741

视频人体姿态分析是指从视频中获取人体的姿态参数, 包括整体平移位置和各关节的旋转角度等, 该过程可形式化地描述为: 设 $X \in \mathbf{R}^d$ 为姿态矢量空间, $Z \in \mathbf{R}^v$ 为图像特征空间, 则姿态分析即是从图像特征 Z 推理得到姿态矢量 X 的过程. 利用普通单目视频实现三维人体姿态分析在运动捕获、三维动画、视频监控以及人机交互等领域有着重要的应用. 然而由于二维图像特征到三维人体姿态映射的多义性、人体姿态空间的高维度、复杂条件下图像特征提取以及人体自遮挡等因素使得该问题的解决相当困难.

当前, 单目视频人体姿态分析方法可分为判别式方法和生成式方法[1]. 判别式方法利用训练数据集 (通常表示为图像特征和姿态数据对: $\{(x_i, z_i)|x_i \in X,\ z_i \in Z,\ i = 1, 2, \cdots, n\}$) 学习建立二维图像特征 Z 和三维人体姿态 X 之间的映射, 该映射可以是函数拟合[2], 也可以是关系数据库查找表[3]. 判别式方法提供了基于单幅图像直接恢

收稿日期 2011-09-09 录用日期 2012-01-05
Manuscript received September 9, 2011; accepted January 5, 2012
国家高技术研究发展计划 (863 计划) (2007AA01Z334), 国家自然科学基金 (69903006, 60373065, 61021062, 61100110), 教育部新世纪优秀人才资助计划 (NCET-04-04605), 江苏省自然科学基金 (BK2009230, BK2010375), 江苏省科技支撑计划 (BE2010072, BE2011058) 资助
Supported by National High Technology Research and Development Program of China (2007AA01Z334), National Natural Science Foundation of China (69903006, 60373065, 61021062, 61100110), Program for New Century Excellent Talents in University of China (NCET-04-04605), Natural Science Foundation of Jiangsu Province (BK2009230, BK2010375), and Key Technology Research and Development Program of Jiangsu Province (BE2010072, BE2011058)
本文责任编委 周杰
Recommended by Associate Editor ZHOU Jie
1. 南京大学软件新技术国家重点实验室 南京大学计算机科学与技术系 南京 210093
1. State Key Laboratory for Novel Software Technology, Department of Computer Science and Technology, Nanjing University, Nanjing 210093

复人体姿态的有效途径, 但其需要大量的训练样本, 同时能够恢复的姿态受限于训练集中包含的姿态类型. 另外, 由于 $Z \rightarrow X$ 映射的多义性, 使得直接映射关系的建立相当困难. 生成式方法可归结为优化问题, 其显式定义一个人体模型, 通过优化人体模型的投影和图像特征间的适应度进行人体姿态估计, 该过程可表示为: $x^* = \arg\min e(g(x) - z)$, 其中 $e(\cdot)$ 为适应度函数, $g(\cdot)$ 为模型投影函数. 生成式方法的优点是不需要训练数据, 能够分析的姿态种类无限制, 且由于 $X \rightarrow Z$ 的映射不存在一对多关系, 使得姿态分析的精度要优于判别式方法. 但生成式方法面临着两个难点问题: 首先, 高维姿态空间是有效全局搜索的最大障碍, 其不仅使数值计算的复杂度增高, 同时难以保证生成符合人体运动学的合理姿态; 其次, 如何保证姿态优化的收敛性和防止局部最优是姿态优化策略设计必须解决的难题.

针对生成式方法存在的以上难题, 本文提出一种基于粒子群优化思想的生成式人体姿态分析方法. 该方法利用运动捕获数据, 采用主成分分析 (Principle component analysis, PCA) 方法学习建立人体姿态的低维紧致空间, 并在此低维空间中进行姿态优化, 不仅减小了姿态计算的复杂度, 且由于运动数据包含了运动的先验知识, 使得在优化过程中有效避免了不合理姿态的产生, 提高了姿态分析的准确性. 另外, 本文将粒子群优化方法 (Particle swarm optimism, PSO) 应用到姿态分析中, 为了提高算法收敛性和防止局部最优提出退火粒子群优化 (Simulated annealing PSO, SAPSO) 姿态分析方法, 实现了基于单幅图像和图像序列的人体姿态分析. 实验表明: 本文在低维姿态空间中基于退火粒子群优化的姿态分析方法具有良好的收敛性和全局搜索能力, 能准确分析单目视频中的人体姿态.

1 相关工作

视频人体姿态分析已有大量的研究和综述[1, 4−5], 这里仅对和本文相关的生成式姿态分析方法进行了小结. 在姿态分析方法研究中, 高维姿态空间是有效全局搜索的首要问题. 解决这一问题的最直接方法是采用简化的人体模型表示来减少计算量, 但简化人体模型难以实现姿态的准确表示且会影响匹配函数的设计. 当前研究的主流方法是诸如 PCA 等线性降维方法, 如 Urtasun 等[6] 利用运动捕获数据采用 PCA 方法学习姿态的低维表示, 但该方法需要同一类型运动的多段运动数据, 且要对运动数据长度进行归一化和相位的对应, 计算复杂; Zhao 等[7] 采用 PCA 方法针对一段运动数据学习姿态的低维表示, 但由于低维空间是视角相关的, 使得能分析的运动受限于训练数据. 近年来, 非线性降维方法如流形学习也被应用到姿态分析中, 包括局部线性嵌入 (Locally linear embedding, LLE) 和等距映射 (Isomap) 等, 如 Sminchisescu 等[8] 采用高斯混合模型方法将高维姿态空间映射到一个低维流形中, 并采用线性动态模型在低维空间实现姿态跟踪, 最后将跟踪结果通过径向基函数 (Radial basis function, RBF) 映射回高维空间, 该方法的不足是无法有效地确定降维后子空间的维数; Wang 等[9] 采用 Isomap 得到运动姿态的流形空间, 采用基于 K-邻域的线性模型近似建立流形空间到高维姿态空间的映射, 并使用 Condensation 算法在流形空间中采样, 将采样粒子映射到高维状态空间. 如何采用解析方法求得逆映射是流形学习应用到姿态空间降维的最大障碍. 此外, 最近研究中还出现了一类非线性隐变量降维方法, 如高斯过程潜变量模型 (Gaussian process latent variable model, GPLVM), Urtasun 等[10] 采用 GPLVM 学习运动先验模型, 建立起姿态空间和潜在空间的平滑映射, 采用梯度下降方法实现了三维人体跟踪, 然而 GPLVM 难以反映运动数据的空间连续性, 使得学习得到的姿态子空间难以表示优化过程中生成的中间姿态, 因此该类方法多用于判别式姿态分析方法中. 鉴于流形学习方法难以实现高低维姿态空间的有效映射而隐变量模型具有非连续性特性, 本文选择采用主成分分析方法学习得到姿态的低维紧致空间, 并借助于数据选择策略消除其视角相关[7] 的缺点.

生成式方法的另一个重要问题是优化策略, 优化策略不仅要保证姿态优化的收敛性, 同时要得到全局最优解. 当前姿态分析中用到的优化方法主要包括梯度下降法[11], 局部搜索法[12]、类粒子滤波[13] 等. 例如 Wachter 等[11] 采用梯度算法进行基于概率的姿态估计, 梯度方法收敛速度快, 但无法保证全局最优; Gavrila 等[12] 采用局部搜索法进行姿态优化, 通过对人体姿态空间进行分解采用分层方式进行搜索, 该方法的缺点是搜索速度慢, 且易陷入局部极值点; 粒子滤波方法是当前研究和应用最为广泛的一类方法[13−15], 如 Deutscher 等[13] 提出退火粒子滤波算法进行快速姿态优化, 取得良好效果; Peursum 等[15] 提出平滑粒子滤波方法, 利用时序信息提高姿态分析的精度; 粒子滤波已成为当前实验的基准对比方法[13, 15−17], 但粒子滤波方法存在的主要问题是粒子数目随着姿态维数的增加以指数级增加, 使得计算相当复杂. 近年来, 进化计算方法也被应用到姿态优化研究中, 如: Krzeszowski 等[18] 结合粒子群优化方法和粒子滤波方法, 对每一步重采样得到的粒子进行粒子群优化以得到更好的预测, 加快了收敛; Zhao 等[7] 用 PCA 对姿态空间进行降维, 并提出分层的退火遗传算法进行优化搜索, 实现

了单目视频的姿态分析, 但遗传算法的编码和算子设计复杂且涉及很多的参数选择; Vijay 等[19] 采用粒子群优化方法进行姿态估计, 为了加快优化速度, 对人体姿态进行分层优化, 提出层级粒子群优化方法, 实现了从多目视频中进行人体姿态分析, 但该方法直接在高维姿态空间中进行姿态估计, 收敛慢且无法保证解的质量, 如何将进化计算方法更好地应用到姿态分析还有待进一步研究. 本文提出一种基于粒子群优化思想的姿态分析方法, 为了提高算法的收敛速度和全局最优性, 采用退火思想对粒子群进行改进, 在 PCA 学习得到的姿态紧致子空间中实现姿态分析.

1.1 本文工作概述

本文在姿态的低维空间中采用退火粒子群优化方法实现姿态估计, 属于生成式姿态分析方法, 其处理流程如图 1 所示. 首先选择特定类型的一段运动捕获数据, 采用主成分分析方法得到姿态空间的低维表示; 其次学习姿态低维表示的时空约束构建姿态紧致空间; 再次, 使用双向剪影方法计算人体模型投影和图像特征的相似度, 并建立适应度函数; 最后, 采用退火粒子群优化的姿态分析方法实现姿态估计和跟踪.

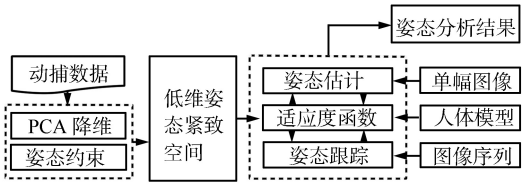


图 1 方法流程图

Fig. 1 The flow chart of our method

2 基于 PCA 的紧致姿态空间学习

本文使用的人体骨架模型由 27 个人体关节点组成, 包括 1 个根节点、21 个中间关节点和 5 个末关节点. 人体的骨架 (姿态) 可表示成: $x = \{x_g, x_k\}$. 其中, $x_g = \{\alpha_x^0, \beta_y^0, \gamma_z^0\}$ 表示根节点的朝向, $x_k = \{\alpha_x^1, \beta_y^1, \gamma_z^1, \cdots, \alpha_x^K, \beta_y^K, \gamma_z^K\}$ 表示中间关节点的朝向, $K = 21$, 人体末关节点自由度为 0, 姿态向量的维度用 M 表示, 本文中 $M = 66$. 图 2 给出了本文使用的人体骨架模型和人体形状模型, 人体形状模型采用圆台体拟合人体的每个骨骼段构成.

2.1 基于 PCA 的低维姿态空间学习

人体姿态由根节点数据和其他关节点数据构成, 而通常特定运动 (如行走) 包含的姿态在同一视角下具有类似结构, 故本文只选取与视角无关的维度并采用 PCA 方法训练得到与视角无关的姿态子空间. 假设人体姿态空间为 X, 对于特定的运动选取一段人体运动数据作为训练数据, 训练得到其低维的状态空间 X_s. 设训练数据为 $\{x_t | t = 1, \cdots, T\}$, 其中, $x_t = \{x_k\}_t$, T 为运动数据总的关键帧数, 则基于主成分分析的低维姿态空间学习方法如下:

1) 将人体数据 $\{x_t | t = 1, \cdots, T\}$ 表示成以下矩阵形式: $X = [(x_1 - c)(x_2 - c), \cdots, (x_T - c)]$, 其中, c 为所有训练数据的均值, 即: $c = \frac{1}{T}\sum_{t=1}^{T} x_t$;

2) 对矩阵 X 作奇异值分解, 得到主方向 $X = UDV^{\mathrm{T}}$;

3) 将姿态向量投影到子空间, 得到每个姿态的低维表示: $x_s = \mathrm{Projection}(x) = U_m^{\mathrm{T}}(x - c)$; 其中, 矩阵 U_m 是 U 的前 m 维列, 由此得到 m 维的子空间 X_s. 为了保证低维到高维的数据重构信息丢失最少, 通常要求前 m 个特征值的和占总特征值和的 95 % 以上, 实验表明 m 一般不大于 5, 本文统一取 $m = 6$. 同时, 高维姿态也可由低维姿态重构, 即: $x = \mathrm{Recovery}(x_s) = c + U_m x_s$. 由此建立了高维和低维姿态空间之间的映射关系.

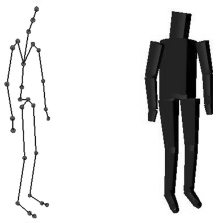


图 2 人体骨架模型和人体形状模型

Fig. 2 The skeleton model and shape model used in this paper

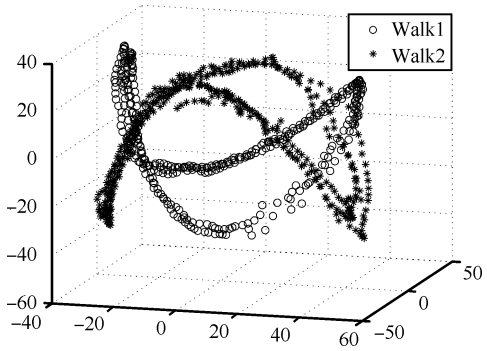
图 3 是分别对两段行走 (图 3 (a)) 和跑步 (图 3 (b)) 运动数据采用本文主成分分析方法学习得到的低维表示, 并将前 3 维绘制出来的结果, 图中每一个点为一个姿态的低维表示. 由图 3 可见, 在低维空间中姿态呈流形分布具有连续性, 相同动作类型的运动具有类似的低维流形结构.

2.2 姿态空间约束

人体姿态在低维空间中呈流形分布, 而不是覆盖整个低维空间. 实际上人体姿态在子空间中覆盖的区域是个紧致空间 (Compact space), 本文通过提出两种姿态约束, 学习得到更符合运动本质的低维紧致姿态空间.

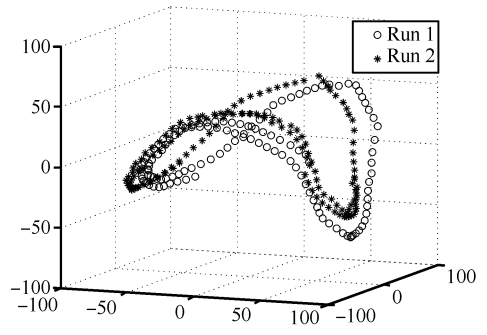
约束 1. 姿态变化范围约束. 主要约束低维姿态每一维的变化范围, 通过运动数据的统计得到. 对每个姿态的低维表示 $x_s = (x_1, x_2, \cdots, x_m)$ 的每个维度 X_i, $i = 1, 2, \cdots, m$ 计算其最大值 $\max(X_i)$ 和最小值 $\min(X_i)$, 同时对姿态速度 $v_s = (v_1, v_2, \cdots, v_m)$ 的每个维度也给出约束, 即每

个低维姿态 $x_s = (x_1, x_2, \cdots, x_m)$ 及其速度 $v_s = (v_1, v_2, \cdots, v_m)$ 需满足: $\min(X_i) \le x_i \le \max(X_i)$ 且 $\min(V_i) \le v_i \le \max(V_i)$, $i = 1, 2, \cdots, m$. 其中 $\min(V_i)$, $\max(V_i)$ 为第 i 维的速度上下限, 通常取 $\min(V_i) = \alpha \min(X_i)$, $\max(V_i) = \alpha \max(X_i)$, 本文实现时取 $\alpha = 0.15$.



(a) 行走数据在三维流形空间中的表示

(a) The manifold of walking data in 3D subspace



(b) 跑步数据在三维流形空间中的表示

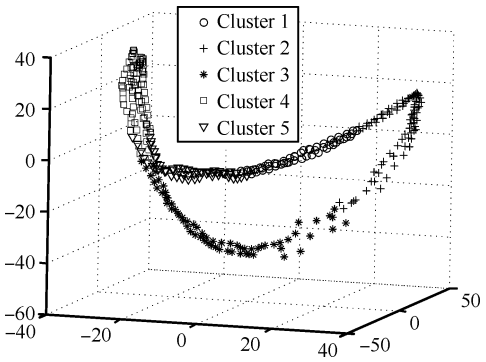
(b) The manifold of running data in 3D subspace

图 3　运动数据的低维流形表示

Fig. 3　The manifolds of motion data in 3D subspace

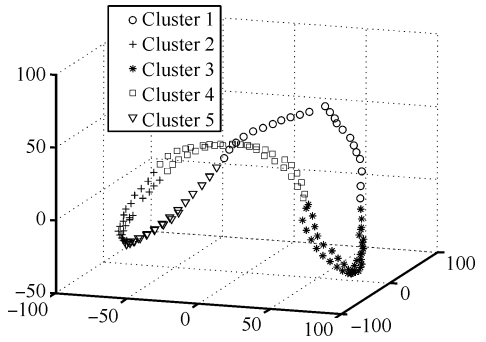
约束 2. 基于 K 均值聚类的姿态空间约束. 仅给出姿态变化范围约束还不能保证每个低维姿态都分布在低维流形上. 本文通过对运动数据进行 K 均值聚类, 对于每个姿态, 计算其到各聚类中心姿态的距离, 大于给定阈值 χ 的姿态被判无效, 由此得到基于 K 均值聚类的姿态空间约束.

图 4 是采用 K 均值聚类分别对行走和跑步低维姿态的聚类结果, 本文取聚类类别数为 5, 设每个类别的中心姿态为 $x_c = (x_{c1}, x_{c2}, \cdots, x_{cm})$, $c = 1, 2, \cdots, 5$, 则每个低维姿态 $x_s = (x_1, x_2, \cdots, x_m)$ 需满足: $\min(D(x_c, x_s)) < \chi$, $c = 1, 2, \cdots, 5$, 即低维姿态离聚类中心距离的最小值需小于阈值 χ. 其中, $D(x_c, x_s) = \sum_{j=1}^{m} \frac{\|x_{cj} - x_j\|}{m}$ 为姿态到聚类中心姿态的距离.



(a) 行走低维姿态聚类

(a) The cluster of walking data



(b) 跑步低维姿态聚类

(b) The cluster of running data

图 4　低维姿态空间中姿态聚类结果

Fig. 4　The cluster results of motion data in low-dimension subspace

给姿态子空间加上以上两条约束即得到低维姿态紧致空间, 与已有方法在姿态原始空间中进行姿态分析不同, 本文在以上学习得到的低维姿态紧致空间中进行姿态分析. 由于该空间更加符合人体运动的本质结构, 同时具有低维度的特点, 大大降低了姿态搜索空间, 能提高姿态分析的精度和效率.

3　退火粒子群优化姿态分析方法

粒子群优化算法是 Kennedy 和 Eberhart 受人工生命研究结果的启发, 通过模拟鸟群觅食过程中的迁徙和群聚行为而提出的一种基于群体智能的全局随机搜索算法. PSO 算法参数少且易实现, 对非线性、多峰问题均具有较强的全局搜索能力和收敛性. 本文将粒子群算法应用到视频人体姿态分析中. 同时, 为了提高粒子群算法的优化能力和有效防止局部最优, 通过引入动态惯性权重和模拟退火思想提出了退火粒子群优化姿态分析方法.

3.1　粒子群算法

粒子群算法首先初始化粒子群, 包括它的随机位置和速度, 并且计算出每个粒子的适应值, 然后

通过迭代搜索最优解. 对于每个粒子通过跟踪两个“极值” 来更新自己的运动方向, 当有足够好的适应值或达到最大迭代次数时, 算法终止.

本文在姿态的低维空间中采用粒子群优化算法进行姿态分析. 在 m 维低维姿态空间 $X_s \subseteq \mathbf{R}^m$ 中, 设群体规模为 N, 群体中每个粒子 $\{p_t^i\}_{i=1}^N$ 有如下属性: 第 n 次迭代时的位置 $x_t^{i,n} = (x_g, x_s)$, 其中, $x_g = \{\alpha_x^0, \beta_y^0, \gamma_z^0\}$ 为根节点方向, $x_s = (x_1, x_2, \cdots, x_m)$, $x_s \in X_s$ 为姿态的低维表示; 飞行速度 $v_t^{i,n} = (v_g, v_s)$, 其中, $v_g = (v_x, v_y, v_z)$ 为根节点方向变化速度, $v_s \in \mathbf{R}^m$ 为低维姿态的变化速度; 粒子的个体极值为 $pbest_t^i$, 以及整个群体的全局极值为 $gbest_t$. 在第 $n+1$ 步, 根据以下公式更新每个粒子的位置和速度:

$$v_t^{i,n+1} = \omega v_t^i + \phi_1(pbest_t^i - x_t^{i,n}) + \phi_2(gbest_t - x_t^{i,n}) \tag{1}$$

$$x_t^{i,n+1} = x_t^{i,n} + v_t^{i,n+1} \tag{2}$$

其中, ω 为惯性权重, $\phi_1 = c_1\text{rand}_1(\cdot)$, $\phi_2 = c_2\text{rand}_2(\cdot)$ 分别为加速系数, 其中, c_1 和 c_2 为常数, 本文取 $c_1 = c_2 = 2$, $\text{rand}_1(\cdot)$ 和 $\text{rand}_2(\cdot)$ 表示 [0, 1] 区间的随机数.

惯性权重对粒子群算法的搜索能力有重要影响, 大的惯性值能使粒子群具有更好的全局搜索能力, 但同时使收敛速度变慢, 小的惯性值使得粒子群具有较好的局部搜索能力, 但算法容易陷入局部最优. 为了使粒子群在初期具有好的全局搜索能力, 在后期能有好的局部搜索能力, 本文对惯性权重作如下设定, 即: $\omega(c) = A/\text{e}^c$, $c \in [0, \ln(10A)]$. 其中, A 为初始惯性权重, 设 C 为最大迭代次数, 则 c 按 $\Delta c = \ln(10A)/C$ 递增.

3.2 适应度函数定义

适应度函数主要用来度量姿态对应的图像特征与真实观测图像特征之间的相似性, 适应度函数设计是粒子群优化的重要步骤. 边缘和剪影是最常用的人体图像特征, 但边缘特征受光线、服饰颜色纹理、背景等影响大而不稳定, 故本文选取图像中的人体剪影特征计算适应度函数.

本文首先采用基于高斯混合模型的背景建模方法从视频中提取人体剪影区域, 并采用双向剪影匹配的相似度计算方法计算适应度. 首先计算图像人体剪影面积 R_t、模型投影面积 B_t、剪影和模型投影重合面积 Y_t, 如式 (3) 所示:

$$\begin{aligned} R_t &= \sum_p \left(M_t^f(p)(1 - M_t^b(p))\right) \\ B_t &= \sum_p \left(M_t^b(p)(1 - M_t^f(p))\right) \\ Y_t &= \sum_p \left(M_t^f(p) M_t^b(p)\right) \end{aligned} \tag{3}$$

其中, M_t^b 和 M_t^f 分别表示 t 时刻人体模型的投影和图像人体剪影. 则基于双向剪影匹配的代价函数为

$$E(y_t, x_t) = S \times \left((1-\beta)\frac{B_t}{B_t + Y_t} + \beta\frac{R_t}{R_t + Y_t}\right) \tag{4}$$

其中, x_t, y_t 分别是 t 时刻姿态和图像特征, S 为常数, β 为权重控制变量, 本文取 $S = 100$, $\beta = 0.5$.

3.3 退火粒子群优化姿态分析算法

基本粒子群算法是一个正反馈过程, 当本身信息和个体极值信息占优势时, 该算法容易陷入局部最优解. 为了使粒子群算法有更好的全局搜索能力, 本文采用模拟退火的思想对粒子群算法进行改进. 模拟退火的基本思想是: 对更新得到的新的粒子, 计算两个位置所引起的适应值的变化量 ΔE; 若 $\Delta E \le 0$, 接受新值; 否则若 $\exp(-\Delta E/T) > \text{rand}(\cdot)$ ($\text{rand}(\cdot)$ 表示 [0, 1] 之间的随机数) 也接受新值; 否则就拒绝, 即 $x_t^{i,n+1}$ 仍为 $x_t^{i,n}$.

综上, 本文模拟退火粒子群优化姿态分析算法具体计算如下:

算法 1. 退火粒子群优化姿态分析算法

输入. 粒子数 N, 迭代次数 C, 初始温度 T, 终止温度 T_0, 退火速度 γ;

输出. 姿态估计 $\widehat{x} = \{x_g, x_k\}$.

算法描述.

1) 初始化

随机产生 N 个粒子 $\{p_t^i\}_{i=1}^N$, 初始化其位置 $x_t^{i,0}$ 和速度 $v_t^{i,0}$, 且粒子的位置和速度满足约束 (1) 和约束 (2);

2) 迭代

While((迭代次数 $< C$) 且 ($T > T_0$)) do

a) 将低维姿态表示映射到高维, 并采用式 (4) 计算每个粒子新的适应值.

b) 对每个粒子, 如粒子的适应值优于原来的个体极值 $pbest_t^i$, 设置当前适应值为个体极值; 并根据每个粒子的个体极值找出全局极值 $gbest_t$.

c) 根据式 (1) 和式 (2) 更新每个粒子的速度和位置, 并检测是否满足约束 (1) 和约束 (2); 若不满足则重新计算粒子的速度和位置.

d) 计算两个位置所引起的适应值的变化量 ΔE; 若 $\Delta E \le 0$, 接受新值; 若 $\exp(-\Delta E/T) > \text{rand}(\cdot)$ 也接受新值; 否则就拒绝, 即 $x_t^{i,n+1}$ 仍为 $x_t^{i,n}$.

e) 若接受新值, 降温 $T \leftarrow \gamma T$; 否则不降温.

End

3) 输出

将收敛的姿态映射回高维姿态空间, 得到姿态估计 $\widehat{x} = \{x_g, x_k\}$.

4　基于退火粒子群优化的姿态估计和跟踪

姿态分析通常包括姿态估计 (Pose estimation) 和姿态跟踪 (Pose tracking) 两部分, 分别指基于单幅图像和基于图像序列的姿态分析, 本文将退火粒子群优化姿态分析算法用于姿态估计和跟踪.

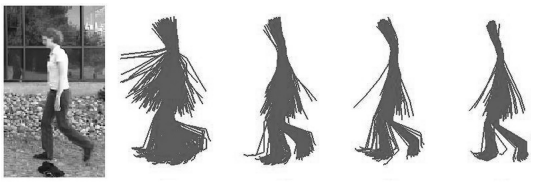
4.1　基于 SAPSO 的姿态估计

姿态估计指的是从单幅图像中分析人体姿态, 其姿态分析过程可描述为以下优化过程:

$$\hat{x}_t = \arg\min_x E(y_t, \text{Recovery}(x_t)) \tag{5}$$

式 (5) 是指在低维姿态空间 X_s 中采用本文退火粒子群优化算法进行姿态估计, 并将结果映射回高维空间, 实现单幅图像的姿态估计.

图 5 给出了采用本文退火粒子群姿态优化方法实现姿态估计的过程. 图 5 (a) 是行走视频的某一帧, 图 5 (b) 是初始化的姿态, 粒子数 $N = 100$, 图 5 (c) 和图 5 (d) 分别是迭代 10 次和 40 次的结果, 图 5 (e) 是迭代 60 次的结果. 由图 5 可见, 初始姿态能有效覆盖行走的姿态空间, 随着迭代次数的增加粒子逐渐收敛到正确的姿态.



(a)　(b)　(c)　(d)　(e)

图 5　基于退火粒子群优化的姿态估计过程 ((a) 某帧视频; (b) 初始化; (c), (d), (e) 分别为迭代 10, 40, 60 次结果)

Fig. 5　The process of human pose estimation ((a) One video frame; (b) The initialized particles; (c), (d), and (e) are results with different numbers of iterations 10, 40, and 60.)

4.2　基于 SAPSO 的姿态跟踪

姿态跟踪指的是基于图像序列的姿态分析, 一般采用分析合成方法, 具体分为预测、匹配、修正三步. 其中, 预测是姿态跟踪的关键, 即根据当前帧姿态预测下一帧的人体姿态. 由于本文退火粒子群方法具有良好的全局搜索能力和收敛性, 所以本文根据 t 时刻的分析结果预测 $t+1$ 时刻的状态, $t+1$ 时刻粒子的初始位置 $x_{t+1}^{i,0}$ 和速度 $v_{t+1}^{i,0}$ 分别满足以下分布:

$$x_{t+1}^{i,0} \sim \text{N}(gbest_t, \Sigma) \tag{6}$$

$$v_{t+1}^{i,0} \sim \text{U}[0, v_t^{\text{pred}}] \tag{7}$$

粒子位置 $x_{t+1}^{i,0}$ 满足以 t 时刻的分析结果, 即 $gbest_t$ 为期望值, 以 Σ 为协方差矩阵的正态分布. 其中, Σ 的对角元素的值与 v_t^{pred} 相关, $v_t^{\text{pred}} = gbest_{t-1} - gbest_{t-2}$ 表示了 t 时刻姿态每一维的速度. 粒子速度 $v_{t+1}^{i,0}$ 满足 $0 \sim v_t^{\text{pred}}$ 的均匀分布.

本文采用姿态估计方法计算视频第一帧的姿态, 并作为跟踪的初始化, 后继帧通过以上预测模型进行预测, 并采用退火粒子群方法进行优化, 实现了基于单目视频的姿态跟踪.

5　实验与结果分析

5.1　实验数据和实验评价机制

实验数据: 本文选取 CMU (Carnegie Mellon University) 的运动捕获数据, 利用本文的人体模型合成虚拟人体运动视频, 并基于 CMU 的 Ground truth 数据分析算法的性能. 运动种类包括不同视角的行走、跑步等. 同时本文在真实视频上进行了实验, 使用的视频包括 CSC (http://www.csc.kth.se/∼hedvig/data.html)、CMU 运动数据库 (http://mocap.cs.cmu.edu/) 和 HumanEva 的测试视频.

误差评价机制: 对于单幅图像, 通过计算姿态分析的结果和真实数据之间的距离进行误差评价, 如式 (8) 所示, $x = (x_1, x_2, \cdots, x_M)$, $\hat{x} = (\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_M)$ 分别表示了 Ground truth 数据和分析得到的姿态数据, M 为姿态向量的维度, 本文中 $M = 66$.

$$D(x, \hat{x}) = \sum_{m=1}^{M} \frac{\|x_m - \hat{x}_m\|}{M} \tag{8}$$

对于帧数为 T 的视频序列, 计算其误差的均值和方差, 方法如下:

$$\mu_{\text{seq}} = \frac{1}{T} \sum_{t=1}^{T} D(x_t, \hat{x}_t) \tag{9}$$

$$\sigma_{\text{seq}} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} [D(x_t, \hat{x}_t) - \mu_{\text{seq}}]^2} \tag{10}$$

5.2　算法收敛性验证

本文退火粒子群姿态优化算法的收敛性主要受粒子数 N 和迭代次数 C 的影响, 本文通过实验给出

了收敛性同粒子数 N 和迭代次数 C 的关系. 如图 6 所示, 不同曲线表示了不同的粒子数目, 横坐标为迭代次数, 纵坐标为全局最优的粒子的适应度值. 由图 6 可见: 迭代次数在 30 次后适应度函数即趋向收敛, 而不同的粒子数目对收敛亦影响不大. 这说明了本文改进的退火粒子群算法具有很好的收敛性. 另一方面, 粒子群算法的计算开销随着粒子数据和迭代次数的增加而增加. 为有效折中精度和计算代价, 本文选取粒子数目为 40, 迭代次数为 60 次. 另外取初始温度 $T = 1\,000$, 终止温度 $T = 1$, 退火速度 $\gamma = 0.95$.

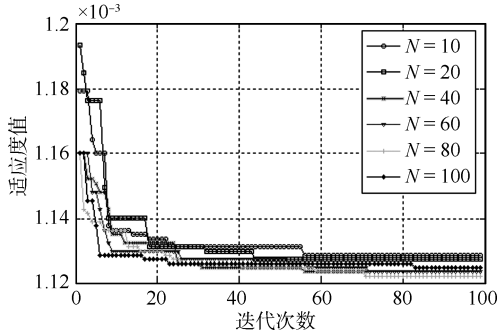


图 6 算法收敛性验证

Fig. 6 Convergence process of SAPSO

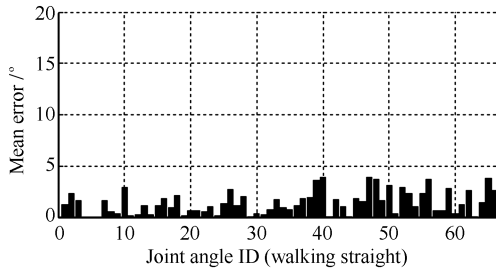
5.3 姿态估计结果

遮挡 (Occlusion)、左右歧义 (Left-right ambiguity)、视角 (View-point) 问题是影响单目视频姿态分析的重要因素, 本文在直线行走、带转向行走、跑步的视频上进行实验, 验证本文方法对以上影响因素的鲁棒性.

实验中用于 PCA 学习姿态子空间的训练数据来自 CMU 运动数据库, 其中步行数据量为 386 帧, 跑步为 198 帧. 图 7 给出了分别在 100 帧直线行走 (图 7 (a))、带转向行走 (图 7 (b))、跑步 (图 7 (c)) 模拟图像上进行姿态估计得到的每一维的误差. 由图 7 可见, 大部分维度的误差在 5 度以内. 另外, 左右手臂和左右膝的夹角是人体运动中变化最显著的关节点, 表 1 给出了直线行走、带转向行走和跑步模拟图像上实验得到的左右手臂和左右膝的夹角误差. 由表 1 可见, 姿态估计结果与真实姿态数据接近. 图 8 给出了在真实直线行走、带转向行走、跑步视频上的实验结果, 得到的姿态估计结果与真实姿态接近.

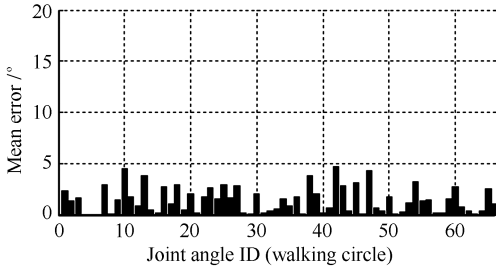
由实验结果可知, 本文算法对遮挡、左右歧义、视角等问题具有较好的处理能力, 具体分析如下:

1) 遮挡问题: 本文实验所采用的视频存在人体



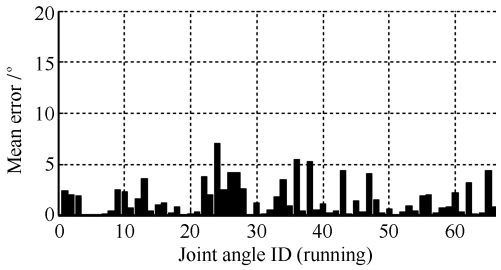
(a) 直线行走视频姿态估计误差分析

(a) Mean error of pose estimation on walk-straight video



(b) 转向行走视频姿态估计误差分析

(b) Mean error of pose estimation on walk-circle video



(c) 跑步视频姿态估计误差分析

(c) Mean error of pose estimation on running video

图 7 姿态估计每一维误差分析

Fig. 7 Mean errors of individual joint angles

表 1 部分关节点姿态真实数据和估计数据比较

Table 1 Comparison of ground truth and estimated data of some joint angles for different motions

		关节点姿态数据			
		LFemure	RFemure	LKnee	RKnee
直线行走	真实数据	(−13.015, 46.717, 10.994)	(−0.127, 6.481, 23.928)	(−2.302, 57.572, 7.482)	(−0.454, 31.281, 4.379)
	估计结果	(−11.397, 48.776, 5.114)	(−0.632, 2.766, 24.117)	(−3.180, 60.346, 10.051)	(0.346, 34.422, 1.443)
转向行走	真实数据	(−12.624, 45.177, 9.155)	(−1.243, 2.548, 19.946)	(−3.693, 64.398, 8.325)	(−0.165, 31.91, 3.480)
	估计结果	(−9.331, 43.254, 2.798)	(−0.538, −0.045, 23.227)	(−2.193, 62.504, 7.542)	(0.960, 34.945, −1.518)
跑步	真实数据	(−12.348, 46.922, 11.044)	(0.591, 4.265, 27.956)	(−1.221, 48.237,6.975)	(−0.513, 30.113, 5.298)
	估计结果	(−11.144, 45.707, 8.435)	(1.127, 1.374, 27.919)	(−1.458, 46.142,7.313)	(0.254, 32.668, 3.628)

遮挡, 但实验结果表明, 在遮挡情况下本文方法也得到了较好的结果. 这主要是因为本文的姿态分析在姿态紧致子空间中进行, 由于姿态紧致子空间包含了人体运动的先验知识, 因而本文方法能根据运动约束估计出被遮挡部分的姿态. 在图 8 中, 尽管手臂被部分遮挡, 但也能得到正确的手臂姿态.

2) 左右歧义问题: 左右歧义指的是左右手 (腿) 识别错误的情况, 但以上实验结果表明, 本文方法得到的结果不存在左右歧义问题. 这主要是因为左右歧义的姿态不符合人体运动约束, 而本文姿态紧致子空间包含了人体运动约束, 能有效避免不符合人体运动姿态的产生.

3) 视角无关性: 实验结果表明, 本文方法能处理不同视角的运动视频, 方法具有视角无关性. 这主要是因为本文提出的数据选择策略, 即只选取与视角无关的姿态维度进行姿态子空间学习, 得到的姿态表示与视角无关. 同时在姿态分析时, 将人体姿态分成姿态和视角两个部分, 对视角进行了单独处理, 这种设计方案保证了能对视角进行全局优化, 因此具有处理多视角人体运动的能力.

实验证明了本文姿态估计方法能有效分析单幅图像中的人体姿态. 不同运动类型视频上的实验表明本文方法的可扩展性.

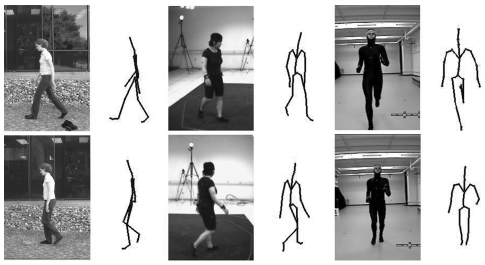


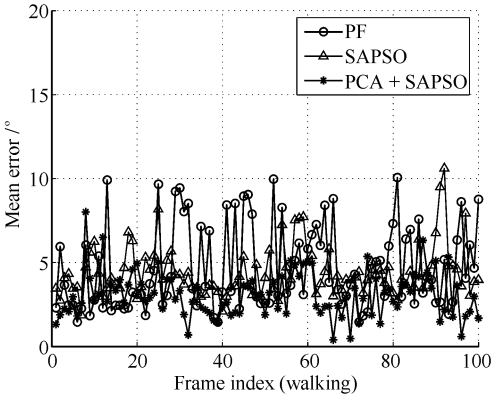
图 8　姿态估计结果

Fig. 8　Results of pose estimation

5.4　姿态跟踪结果

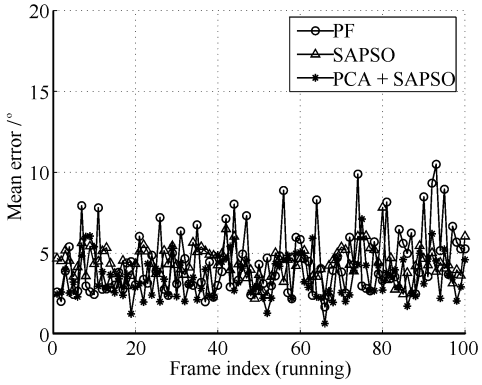
本文在行走和跑步视频序列上进行实验验证本文姿态跟踪算法的有效性, 同时将本文姿态紧致空间中退火粒子群优化姿态跟踪方法 (PCA + SAPSO) 和粒子滤波方法 (Particle filter, PF)、高维姿态空间中退火粒子群优化方法 (SAPSO) 进行了实验比较.

图 9 给出了分别在 100 帧行走和跑步视频序列上采用 PF、SAPSO、PCA + SAPSO 方法的实验结果, 同时表 2 给出了实验误差分析. 由图 9 和表 2 可见, SAPSO 方法实验结果要优于 PF 方法, 而本文方法要优于 SAPSO 方法. 本文方法的平均误差在 4 度左右. 本文方法在部分帧上的误差较高, 但这主要是因为该帧上个别关节点的误差很大, 导致整体误差的增加, 例如头部姿态分析的不准确导致的整体误差增加. 表 2 给出了在粒子数目为 40, 迭代次数为 60 次, 100 帧视频上实验得到的平均每帧处理时间. 由表 2 可见, 本文在低维子空间上进行姿态估计的方法相比于在原始空间中进行姿态估计能提高计算效率.



(a) 行走序列姿态跟踪结果对比

(a) Comparison of tracking results on walking video



(b) 跑步序列姿态跟踪结果对比

(b) Comparison of tracking results on running video

图 9　姿态跟踪结果比较

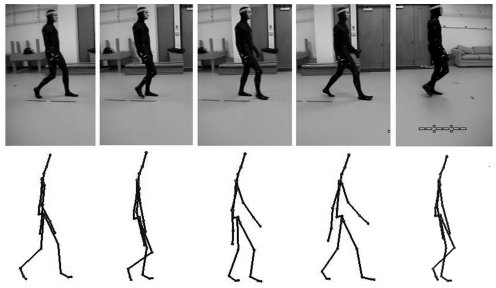
Fig. 9　Comparison of different tracking methods

表 2　姿态跟踪误差统计和对比

Table 2　Comparison of tracking results for different methods

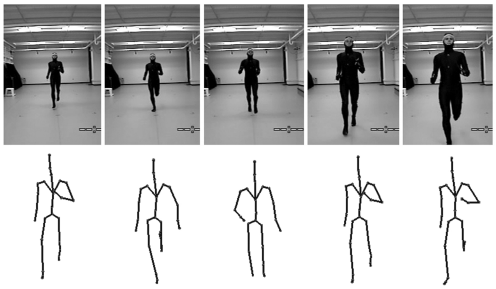
	步行			跑步		
	误差 (度)	方差	耗时 (秒)	误差 (度)	方差	耗时 (秒)
PF	4.6713	2.5157	5.236	4.4669	2.0289	4.965
SAPSO	4.4369	1.5181	4.384	4.3949	0.9821	3.962
PCA + SAPSO	3.5123	1.6324	2.541	3.8769	1.3789	2.365

图 10 给出了在真实行走和跑步视频上的实验结果, 实验结果表明了本文方法能有效实现单目视频的姿态跟踪, 且本文方法具有扩展到多种运动类型视频跟踪的能力. 姿态初始化是姿态跟踪的难题, 由于本文姿态优化方法具有良好的收敛性, 因此本文采用视频第一帧的姿态估计结果作为跟踪的初始化, 姿态跟踪实验结果证明了本文姿态初始化方法的有效性.



(a) 行走序列姿态跟踪结果

(a) Tracking results on walking video



(b) 跑步序列姿态跟踪结果

(b) Tracking results on running video

图 10 姿态跟踪结果比较

Fig. 10 Comparison of different tracking methods

以上实验证明了本文姿态跟踪方法能有效分析单目视频中的人体姿态. 但本文方法也存在以下不足: 首先, 针对不同的运动需要利用运动捕获数据学习其低维姿态空间, 因此针对复杂的运动视频能够有效自动选择低维姿态空间的机制有待研究; 其次, 视角对姿态分析影响较大, 如何鲁棒处理视角对姿态分析带来的影响需要研究; 另外, 由于本文方法依赖于从图像中提取人体剪影, 当背景复杂时剪影提取不准确使得姿态分析误差变高, 有效的图像特征计算方法和非基于剪影检测的相似度计算方法值得研究; 最后, 姿态分析的实时性应用和基于姿态分析的姿态识别等还有待更多的研究, 这都将成为我们下一步的工作目标.

6 结论

基于单目视频的人体姿态分析是当前计算机视觉的难点问题之一, 本文提出一种基于退火粒子群优化思想的单目视频人体姿态分析方法, 主要贡献体现在以下两点: 首先, 针对高维人体姿态空间提出采用主成分分析方法学习得到更符合运动本质的姿态紧致空间, 姿态分析在低维空间中进行. 该方法在降低姿态空间的同时利用了运动数据包含的运动先验知识, 提高了姿态分析的效率和精度; 其次, 将粒子群算法引入姿态分析中, 为了提高算法的收敛性和全局搜索能力, 提出退火粒子群优化姿态分析方法, 实现了基于单目视频的姿态估计和姿态跟踪. 本文方法具有良好的计算效率, 同时退火粒子群算法具有良好的收敛性和全局搜索能力, 在模拟和真实视频上的实验证明了本文方法能准确分析单目视频中的人体姿态.

References

1 Sminchisescu C. 3D human motion analysis in monocular video: techniques and challenges. In: Proceedings of the IEEE International Conference on Video and Signal Based Surveillance (AVSS′06). Washington D. C., USA: IEEE, 2006. 76

2 Agarwal A, Triggs B. Recovering 3D human pose from monocular images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(1): 44−58

3 Howe N R. Silhouette lookup for monocular 3D pose tracking. *Image and Vision Computing*, 2007, **25**(3): 331−341

4 Moeslund T B, Hilton A, Krüger V. A survey of advances in vision-based human motion capture and analysis. *Computer Vision and Image Understanding*, 2006, **104**(2–3): 90−126

5 Ronald P. Vision-based human motion analysis: an overview. *Computer Vision and Image Understanding*, 2007, **108**(1–2): 4−18

6 Urtasun R, Fleet D J, Fua P. Monocular 3D tracking of the golf swing. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, CA, USA: IEEE, 2005. 932−938

7 Zhao X, Liu Y C. Generative tracking of 3D human motion by hierarchical annealed genetic algorithm. *Pattern Recognition*, 2008, **41**(8): 2470−2483

8 Sminchisescu C, Jepson A. Generative modeling for continuous non-linearly embedded visual inference. In: Proceedings of the 21st International Conference on Machine Learning. New York, NY: ACM, 2004. 759−766

9 Wang Q, Xu G Y, Ai H Z. Learning object intrinsic structure for robust visual tracking. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Madison, USA: IEEE, 2003. 227−233

10 Urtasun R, Fleet D J, Hertzmann A, Fua P. Priors for people tracking from small training sets. In: Proceedings of the 10th IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Computer Society, 2005. 403−410

11 Wachter S, Nagel H H. Tracking persons in monocular image sequences. *Computer Vision and Image Understanding*, 1999, **74**(3): 174−192

12 Gavrila D M, Davis L S. Tracking of humans in action: a 3D model-based approach. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA: IEEE, 1996. 73−80

13 Deutscher J, Blake A, Reid I. Articulated body motion capture by annealed particle filtering. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Hilton Head, SC, USA: IEEE, 2000. 126−133

14 Sigal L, Balan A O, Black M J. HumanEva: synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. *International Journal of Computer Vision*, 2010, **87**(1): 4−27

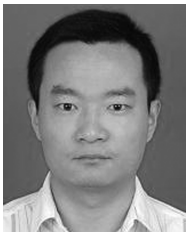
15 Peursum P, Venkatesh S, West G. A study on smoothing for particle-filtered 3D human body tracking. *International Journal of Computer Vision*, 2010, **87**(1-2): 53−74

16 Daubney B, Xie X H. Tracking 3D human pose with large root node uncertainty. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Colorado Springs, CO, USA: IEEE, 2011. 1321−1328

17 Wang X Y, Wan W G, Zhang X Q. Annealed particle filter based on particle swarm optimization for articulated three-dimensional human motion tracking. *Optical Engineering*, 2010, **49**(1): 017204−11

18 Krzeszowski T, Kwolek B, Wojciechowski K. Articulated body motion tracking by combined particle swarm optimization and particle filtering. In: Proceedings of the 2010 International Conference on Computer Vision and Graphics: Part I. Warsaw, Poland: LNCS, 2010. 147−154

19 Vijay J, Emanuele T, Spela I. Markerless human articulated tracking using hierarchical particle swarm optimisation. *Image and Vision Computing*, 2010, **28**(11): 1530−1547



李 毅 南京大学计算机科学与技术系博士研究生. 主要研究方向为图像处理与计算机视觉.
E-mail: njulanty@gmail.com
(**LI Yi** Ph. D. candidate in the Department of Computer Science and Technology, Nanjing University. His research interest covers image process and computer vision.)

孙正兴 南京大学计算机科学与技术系教授. 主要研究方向为多媒体计算与计算机视觉. 本文通信作者.
E-mail: szx@nju.edu.cn
(**SUN Zheng-Xing** Professor in the Department of Computer Science and Technology, Nanjing University. His research interest covers multimedia computing and computer vision. Corresponding author of this paper.)

陈松乐 南京大学计算机科学与技术系博士研究生. 主要研究方向为图像处理与计算机视觉.
E-mail: flychen21@gmail.com
(**CHEN Song-Le** Ph. D. candidate in the Department of Computer Science and Technology, Nanjing University. His research interest covers image processing and computer vision.)

李 骞 南京大学计算机科学与技术系博士研究生. 主要研究方向为图像处理与计算机视觉.
E-mail: liqian0267@gmail.com
(**LI Qian** Ph. D. candidate in the Department of Computer Science and Technology, Nanjing University. His research interest covers image processing and computer vision.)