

基于双隐变量空间局部粒子搜索的人体运动形态估计

李万益* 孙季丰 王玉龙

(华南理工大学电子与信息学院 广州 510641)

摘要: 该文提出一种双隐变量空间局部粒子搜索(DLVSLPS)算法, 可以从多视角图像序列的轮廓特征较准确地估计出 3 维人体运动形态序列。该算法用高斯过程动态模型(GPDM)降维建立双隐变量空间和低维隐变量数据到高维数据的映射关系后, 然后对双隐变量空间使用近邻权重先验条件搜索(NWPCS), 实现局部低维粒子搜索来生成较优高维数据, 从而估计相应帧的 3 维人体运动形态, 解决传统粒子滤波算法直接在高维数据空间采样较难获取有效正确数据进行估计的问题。经仿真实验验证, 所提出的算法比传统粒子滤波算法在实现多视角非连续帧估计, 克服轮廓图像数据歧义, 减小估计误差有明显优势。

关键词: 人体运动形态估计; 双隐变量空间; 局部粒子搜索; 多视角图像序列; 3 维人体运动形态序列

中图分类号: TP391.4

文献标识码: A

文章编号: 1009-5896(2014)12-2915-08

DOI: 10.3724/SP.J.1146.2013.01915

Human Motion Estimation Based on Dual Latent Variable Spaces Local Particle Search

Li Wan-yi Sun Ji-feng Wang Yu-long

(School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, China)

Abstract: A novel algorithm called Dual Latent Variable Spaces Local Particle Search (DLVSLPS) is proposed. It can estimate the 3D human motion sequence from silhouettes of multi-view image sequence more accurately. Gaussian Process Dynamical Models (GPDM) is used to reduce the dimension to build the dual latent variable spaces and the mapping from low dimensional latent variable data to high dimensional data. Then, the low dimensional particles are searched in these spaces by the method called Neighbor Weight Prior Condition Search (NWPCS). The better high dimensional data are generated from the mapping to estimate the 3D human motion of the corresponding frame. The proposed algorithm aims to solve the problem of traditional particle filters. The problem is that sampling in high dimensional data space can not get the valid and correct data to estimate the 3D human motion. The simulating experiments show the proposed algorithm has better performance than the traditional particle filters. The better performance includes the multi-view and discontinuous frame estimation, overcoming the silhouette ambiguity and reducing the estimation error.

Key words: Human motion estimation; Dual latent variable spaces; Local particle search; Multi-view image sequence; 3D human motion sequence

1 引言

人体运动形态估计是近些年国内外的热点研究课题^[1-4], 尤其是如何从多视角图像序列估计出含有空间位置信息的 3 维人体运动形态序列是研究难点。文献[5,6]用退火粒子滤波(Annealed Particle Filter, APF)算法从高维数据空间直接采样或者部分维多层采样, 选择出有效粒子, 耗费大量的计算时间, 而且估计效果不理想。APF 与其他传统粒子滤波算法如粒子群优化粒子滤波(Particle Swarm

Optimization-Particle Filter, PSO-PF)^[7]及标准粒子滤波(Particle Filter, PF)^[7]等, 估计效果非常依赖于前一时刻的生成粒子, 如果前一时刻生成的粒子较好, 与后一时刻需要估计的粒子较接近, 则估计效果就好, 否则效果较差。文献[8]中, 采用基于主成分分析(Principal Component Analysis, PCA)的退火粒子群优化来估计 3 维人体运动形态, 不能充分利用低维隐变量数据先验信息, 搜索困难且时间很长, 无法估计其人体运动形态的空间位置信息, 并且在用 PCA 降维后, 低维隐变量数据还原得到的高维数据和原高维数据误差较大。从以上研究发现, 要想估计出理想的 3 维人体运动形态很大程度上取决于要有较强先验信息生成有效正确的高维数据。

本文针对以上问题, 提出一种可靠稳定的双隐

2013-12-09 收到, 2014-06-30 改回

国家自然科学基金青年基金(61202292)和广东省自然科学基金(9151064101000037)资助课题

*通信作者: 李万益 luther1212@163.com

变量空间局部粒子搜索算法(Dual Latent Variable Spaces Local Particle Search, DLVSLPS), 对于估计周期性匀速含空间位置信息 3 维人体运动形态很有效。该算法采用高斯过程动态模型(Gaussian Process Dynamical Models, GPDM)^[9,10]分别对少量两套(带空间位置信息和不带空间位置信息)已知的 3 维人体运动形数据降维, 获取低维隐变量数据, 建立低维隐变量数据到高维人体运动形态数据函数映射关系后, 再用近邻权重先验条件搜索(Neighbor Weight Prior Condition Search, NWPCS)来实现。GPDM 降维算法与其他降维算法^[11]相比, 其具有低维到高维可逆性, 并且从低维隐变量数据还原的高维数据与原高维数据差别较小, 它由高斯过程隐变量模型(Gaussian Process Latent Variable Models, GPLVM)^[12,13]演变而来。这两个隐变量空间中的低维隐变量数据对于搜索过程中具有很好的先验信息, 本文提出算法就可以较好利用该信息, 结合相关约束模型, 对两个空间低维粒子进行局部搜索, 生成较有效正确的高维数据进行估计。本文提出的算法与传统几种粒子滤波算法(APF, PSO-PF, PF)相比, 可以估计非连续帧所对应的 3 维人体运动形态, 有效克服轮廓图像数据歧义, 且减小估计误差。最后, 通过仿真实验验证该算法比传统几种粒子滤波算法更有效。文中称描述 3 维人体运动形态的数据为高维数据, 其降维后在隐变量空间中所获取的低维隐变量数据称为低维数据(低维粒子), 隐变量空间也称为低维空间。

2 高斯过程动态模型(GPDM)降维

本文需要估计的 3 维人体运动形态用如图 1(a)所示的人体运动立体模型表示, 里面由圆台体与骨架组成, 该模型可以较近似地描述真实人体运动形态, 肢体大小, 并且该模型是由高维数据转换而来。在对多视角的 2 维人体运动图像处理前, 可以将已知少量立体模型的高维数据组成一个稀疏的数据样本 1, 这组样本序列是覆盖有一个周期空间位置信息的, 如图 1(b)所示。然后, 本文同样可以采集一个走路运动周期少量无空间位置信息稀疏的数据样本 2, 如图 1(c)所示。两组高维数据用 GPDM 降维后就建立了低维空间且获取了如图 1(d)与图 1(e)所示的相应两组低维数据, 及其函数映射关系。GPDM 相关概率关系为

$$\left. \begin{aligned} p(\mathbf{Y} | \mathbf{X}, \bar{\boldsymbol{\beta}}, \mathbf{W}) \\ = \frac{|\mathbf{W}|^N}{\sqrt{(2\pi)^{ND} |\mathbf{K}_Y|^D}} \exp\left(-\frac{1}{2} \text{tr}(\mathbf{K}_Y^{-1} \mathbf{Y} \mathbf{W}^2 \mathbf{Y}^T)\right) \\ k_Y(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\beta_1}{2} \|\mathbf{x} - \mathbf{x}'\|^2\right) + \beta_2^{-1} \delta_{\mathbf{x}\mathbf{x}'} \end{aligned} \right\} (1)$$

$$\left. \begin{aligned} p(\mathbf{X} | \bar{\boldsymbol{\alpha}}) &= \frac{p(\mathbf{x}_1)}{\sqrt{(2\pi)^{(N-1)q} |\mathbf{K}_X|^q}} \\ &\cdot \exp\left(-\frac{1}{2} \text{tr}(\mathbf{K}_X^{-1} \mathbf{X}_{2:N} \mathbf{X}_{2:N}^T)\right) \\ k_X(\mathbf{x}, \mathbf{x}') &= \alpha_1 \exp\left(-\frac{\alpha_2}{2} \|\mathbf{x} - \mathbf{x}'\|^2\right) \\ &+ \alpha_3 \mathbf{x}^T \mathbf{x}' + \alpha_4^{-1} \delta_{\mathbf{x}\mathbf{x}'} \end{aligned} \right\} (2)$$

$$\begin{aligned} p(\mathbf{X}, \mathbf{Y}, \bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}, \mathbf{W}) \\ = p(\mathbf{Y} | \mathbf{X}, \bar{\boldsymbol{\beta}}, \mathbf{W}) p(\mathbf{X} | \bar{\boldsymbol{\alpha}}) p(\bar{\boldsymbol{\alpha}}) p(\bar{\boldsymbol{\beta}}) p(\mathbf{W}) \end{aligned} \quad (3)$$

$$\begin{aligned} \boldsymbol{\mu}_Y(\mathbf{x}_i) &= \mathbf{Y}^T \mathbf{K}_Y^{-1} [k_Y(\mathbf{x}_1, \mathbf{x}_i), k_Y(\mathbf{x}_2, \mathbf{x}_i), \dots, \\ &k_Y(\mathbf{x}_N, \mathbf{x}_i)]^T = \mathbf{Y}^T \mathbf{K}_Y^{-1} k_Y(\mathbf{x}_i) \end{aligned} \quad (4)$$

其中, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N]^T \in R^{N \times D}$, $\mathbf{y}_i \in R^D$, $i = 1, 2, \dots, N$ 为高维(D 维)数据序列, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]^T \in R^{N \times q}$, $\mathbf{x}_i \in R^q$, $i = 1, 2, \dots, N$ 为低维(q 维)数据序列, \mathbf{W} 为尺度参数且 $p(\mathbf{W}) = \prod_{m=1}^D (2/(\kappa\sqrt{2\pi})) \exp(-w_m^2 / (2\kappa^2))$, $w_m > 0, \kappa = 10^{-3}$, $\mathbf{K}_Y \in R^{N \times N}$, $(\mathbf{K}_Y)_{i,j} = k_Y(\mathbf{x}_i, \mathbf{x}_j)$ 为带核参数 $\bar{\boldsymbol{\beta}} = [\beta_1, \beta_2]$ 核矩阵, $\mathbf{K}_X \in R^{(N-1) \times (N-1)}$, $(\mathbf{K}_X)_{i,j} = k_X(\mathbf{x}_i, \mathbf{x}_j)$ 为带核参数 $\bar{\boldsymbol{\alpha}} = [\alpha_1, \alpha_2, \alpha_3, \alpha_4]$ 核矩阵, 通过 $\mathbf{X}_{1:N-1} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{N-1}]^T$ 构造, 然而 $\mathbf{X}_{2:N} = [\mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_N]^T$ 。 \mathbf{x}_1 服从 q 维高斯分布, 核参数 $\bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}$ 满足 $p(\bar{\boldsymbol{\alpha}}) \propto \prod_i \alpha_i^{-1}$, $p(\bar{\boldsymbol{\beta}}) \propto \prod_i \beta_i^{-1}$ 。由于 $\mathbf{y}_i \sim \mathcal{N}(\boldsymbol{\mu}_Y(\mathbf{x}_i), \sigma_Y^2(\mathbf{x}_i) \mathbf{I})$, 未知高维数据所服从多维高斯分布的均值及方差是低维数据的函数, 低维数据到高维数据就建立了函数映射关系, 如式(4)所示。训练 GPDM 时, 已知高维(D 维)数据序列 \mathbf{Y} , 即 $p(\mathbf{Y})$ 为已知常数, 使式(3)最大就得到相应的最优参数, 及低维(q 维)数据序列, 即: $\mathbf{X}, \bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}, \mathbf{W} = \arg \min_{\mathbf{X}, \bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}, \mathbf{W}} (-\ln p(\mathbf{X}, \bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}, \mathbf{W} | \mathbf{Y})) \Leftrightarrow \arg \min_{\mathbf{X}, \bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}, \mathbf{W}} (-\ln p(\mathbf{X}, \bar{\boldsymbol{\alpha}}, \bar{\boldsymbol{\beta}}, \mathbf{W}, \mathbf{Y}))$ 。

3 双隐变量空间局部粒子搜索

3.1 相关数学模型建立

根据文献[5]的研究以及 Humaneva-I 数据库中的采集数据特点, 每一帧的人体运动形态图像由 3 个不同视角彩色镜头拍摄, 如图 2 第 1 行的 3 幅图所示。那么本文通过图像分割算法^[14,15]对该帧的 3 个视角图像进行处理得到相应视角的轮廓图像, 如图 2 第 2 行的 3 幅图所示, 本文还需要对生成的人体运动立体模型做该 3 个视角的 2 维图像投影, 如图 2 第 3 行的 3 幅图所示。根据式(4), 本文有以下函数及概率关系:

$$R_i^k = \sum_p (M_i^f(p)(1 - M_i^b(p))) \quad (5)$$

$$B_i^k = \sum_p (M_i^b(p)(1 - M_i^f(p))) \quad (6)$$

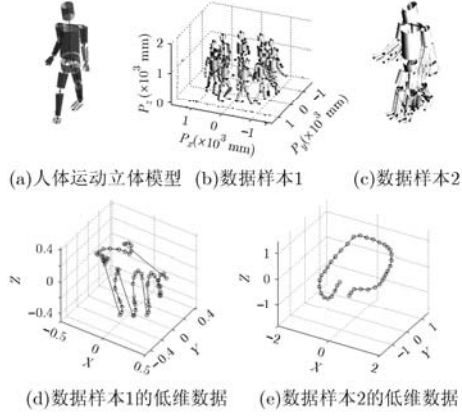
图 1 数据样本示意图与其相应隐变量空间的低维数据($q=3$)

图 2 图像数据

$$\begin{aligned}
 H_t^k &= \sum_p (M_t^f(p) M_t^b(p)) \\
 -\ln p(m_t^k | \mathbf{y}_t^i) &= -\ln p(m_t^k | \mathbf{Y}^T \mathbf{K}_Y^{-1} \mathbf{k}_Y(\mathbf{x}_t^i)) \\
 &= -\ln p(m_t^k | \mathbf{x}_t^i) \propto (1 - \alpha) \\
 &\quad \cdot \frac{B_t^k}{B_t^k + H_t^k} + \alpha \frac{R_t^k}{R_t^k + H_t^k}
 \end{aligned} \quad (7)$$

其中, t 为图像的帧序数, k 为镜头视角序数, α 为权重系数, p 为 2 维图像的像素点, M_t^f 为轮廓图像像素点的集合, M_t^b 为人体运动立体模型在相应视角的投影图像像素点的集合, $M_t^f(p)$ 为像素点 p 是否在轮廓图像里的统计函数 ($M_t^f(p) = 1, p \in M_t^f$ 或 $M_t^f(p) = 0, p \notin M_t^f$), $M_t^b(p)$ 为像素点 p 是否在投影图像里的统计函数 ($M_t^b(p) = 1, p \in M_t^b$ 或 $M_t^b(p) = 0, p \notin M_t^b$)。在第 k 个视角, R_t^k 为像素点 p 只在轮廓图像里而不在投影图像的统计函数值, B_t^k 为像素点 p 只在投影图像里而不在轮廓图像的统计函数值, H_t^k 为像素点 p 同时在轮廓图像和投影图像里的统计函数值, m_t 为目标轮廓图像数据特征, \mathbf{y}_t^i 为高维数据样本, \mathbf{x}_t^i 为相应低维数据, $p(m_t^k | \mathbf{y}_t^i)$ 和 $p(m_t^k | \mathbf{x}_t^i)$ 分别是出现 \mathbf{y}_t^i 和 \mathbf{x}_t^i 在第 k 个视角且符合目标观测轮廓图像数据特征的条件概率。

3.2 低维局部粒子的 NWPCS 及运动形态估计

根据 3.1 节的数学模型, 在 t 时刻(第 t 帧)的 N_V 个视角下及任意一个低维空间里, 设 \mathbf{x}_t^i 为相应 N_e 个高维数据对应的第 i 个低维数据, 即低维粒子, 其归一化权重为

$$\tilde{w}_t^i = w_t^i / \sum_{i=1}^{N_e} w_t^i \quad (9)$$

令 $\mathbf{W}_B = [\tilde{w}_t^1, \tilde{w}_t^2, \dots, \tilde{w}_t^{N_e}]$, $w_t^i = p(\mathbf{x}_t^i, m_t) / q(\mathbf{x}_t^i | \mathbf{X}_t)$ 。如果 $q(\mathbf{x}_t^i | \mathbf{X}_t)$ 为联合均匀分布密度, $w_t^i, i = 1, 2, \dots, N_e$ 在一定程度上说明由 $q(\mathbf{x}_t^i | \mathbf{X}_t)$ 生成的低维

粒子 \mathbf{x}_t^i 是 $p(\mathbf{x}_t^i, m_t)$ 生成的低维粒子 \mathbf{x}_t^i 的接受概率, 也就是说, w_t^i 越大, $p(\mathbf{x}_t^i, m_t)$ 也越大, 已知轮廓图像数据特征 m_t (即 $p(m_t)$ 为常数) 时, 其条件密度 $p(\mathbf{x}_t^i | m_t) = p(\mathbf{x}_t^i, m_t) / p(m_t)$ 也越大, \mathbf{x}_t^i 越能说明是由 $p(\mathbf{x}_t^i | m_t)$ 生成的低维粒子, 因此搜索低维粒子 \mathbf{x}_t^i 就必须搜索权重大的, 才能反映出其就是 $p(\mathbf{x}_t^i | m_t)$ 的真实样本。下面给出近邻权重先验条件搜索 (NWPCS) 原理。

定义 $k_{ps} = \text{Ind}_{ps}^n(\mathbf{W}_B)$ 为寻找 $\mathbf{W}_B = [\tilde{w}_t^1, \tilde{w}_t^2, \dots, \tilde{w}_t^{N_e}]$, $N_e > 3$ 中前 n 个最大元素位列第 ps 的元素序号 k_{ps} 的函数, ${}_r \mathbf{x}_t^i$ 为 t 时刻第 r 次搜索时第 i 个低维粒子, 权重前 3 个较大的粒子间距离为 $l_{21} = \|{}_r \mathbf{x}_t^{k_2} - {}_r \mathbf{x}_t^{k_1}\|$, $l_{31} = \|{}_r \mathbf{x}_t^{k_3} - {}_r \mathbf{x}_t^{k_1}\|$, 令 $L = \max(l_{21}, l_{31})$ 用于调节搜索尺度, $\mathbf{rand}(n_1, n_2)$ 为元素服从均匀分布 $u(-1, 1)$ 的 $n_1 \times n_2$ 矩阵生成函数, 令 $\mathbf{u}_i = \mathbf{rand}(q, 1)$, $\eta_i = \mathbf{rand}(1, 1)$, 在 q 维低维空间里, 有 $N_e - 1$ 个粒子标准化后一定随机长度的搜索向量 $\tilde{\mathbf{u}}_t^i = \eta_i L (\mathbf{u}_i / \|\mathbf{u}_i\|)$, $i = 1, 2, \dots, N_e - 1$, 在各方向均有均等机会出现, 其搜索得到 $N_e - 1$ 个低维空间新的粒子为 ${}_{r+1} \mathbf{x}_t^i = {}_r \mathbf{x}_t^{k_1} + \tilde{\mathbf{u}}_t^i, i = 1, 2, \dots, N_e - 1$, 把先前最大权重粒子 ${}_r \mathbf{x}_t^{k_1}$ 也保留作为一个新样本, 可避免后面搜索得到比原来差的粒子, 即令 ${}_{r+1} \mathbf{x}_t^{N_e} = {}_r \mathbf{x}_t^{k_1}$, 得出的新粒子共 N_e 个。新的粒子概率为 \mathbf{u}_i 和 η_i 服从联合均匀分布概率。

$$\begin{aligned}
 q({}_{r+1} \mathbf{x}_t^i | \mathbf{X}_t) &= q({}_{r+1} \mathbf{x}_t^i | {}_r \mathbf{x}_t^{k_1}, {}_r \mathbf{x}_t^{k_2}, {}_r \mathbf{x}_t^{k_3}) \\
 &= p(\mathbf{u}_i, \eta_i)
 \end{aligned} \quad (10)$$

因 $p(\mathbf{y}_t^i | m_t)$ 分布未知, 由上结论, 当 N_e 较大, 令 $p(\mathbf{x}_t^i) = q(\mathbf{x}_t^i | \mathbf{X}_t)$ 时, 所要估计 \mathbf{y}_t^i 的近似均值和权重计算分别为

$$\begin{aligned}
\mathbf{E}(\mathbf{y}_t^i | m_t) &= \int \mathbf{y}_t^i p(\mathbf{y}_t^i | m_t) d\mathbf{y}_t^i \approx \sum_{i=1}^{N_c} \mathbf{y}_t^i p(\mathbf{y}_t^i | m_t) \\
&= \sum_{i=1}^{N_c} \mathbf{y}_t^i p(\mathbf{x}_t^i, m_t) / p(m_t) \\
&= \sum_{i=1}^{N_c} \mathbf{y}_t^i p(\mathbf{x}_t^i, m_t) / \sum_{i=1}^{N_c} p(\mathbf{x}_t^i, m_t) \\
&= \sum_{i=1}^{N_c} \mathbf{y}_t^i w_t^i q(\mathbf{x}_t^i | \mathbf{X}_t) / \sum_{i=1}^{N_c} w_t^i q(\mathbf{x}_t^i | \mathbf{X}_t) \\
&\approx \frac{1}{N_c} \sum_{i=1}^{N_c} \mathbf{y}_t^i w_t^i / \frac{1}{N_c} \sum_{i=1}^{N_c} w_t^i = \sum_{i=1}^{N_c} \tilde{w}_t^i \mathbf{y}_t^i \\
&= \sum_{i=1}^{N_c} \tilde{w}_t^i \mathbf{Y}^T \mathbf{K}^{-1} \mathbf{k}_Y(\mathbf{x}_t^i) \quad (11)
\end{aligned}$$

$$\begin{aligned}
w_t^i &= p(\mathbf{x}_t^i, m_t) / q(\mathbf{x}_t^i | \mathbf{X}_t) \\
&= p(m_t | \mathbf{x}_t^i) p(\mathbf{x}_t^i) / q(\mathbf{x}_t^i | \mathbf{X}_t) = p(m_t | \mathbf{x}_t^i) \quad (12)
\end{aligned}$$

$$\begin{aligned}
w_t^i &= p(m_t | \mathbf{x}_t^i) = \prod_{k=1}^{N_v} p(m_t^k | \mathbf{x}_t^i) \\
&= \exp \left(- \sum_{k=1}^{N_v} (-\ln p(m_t^k | \mathbf{x}_t^i)) \right) \Leftrightarrow w_t^i \\
&= \exp \left(- \frac{S}{N_v} \sum_{k=1}^{N_v} \left((1-\alpha) \frac{B_t^k}{B_t^k + H_t^k} \right. \right. \\
&\quad \left. \left. + \alpha \frac{R_t^k}{R_t^k + H_t^k} \right) \right) = \mathbf{F}(\mathbf{y}_t^i) \quad (13)
\end{aligned}$$

式(13)中 S, α 为参数, 通常取 $S=100, \alpha=0.5$ 。经过一定次数搜索后, 得到全局低维粒子中较接近最大权重值的局部低维粒子 $\hat{\mathbf{x}}_t^i$, 从而估计所需高维数据 ${}^h \hat{\mathbf{y}}_t = \sum_{i=1}^{N_c} \tilde{w}_t^i \mathbf{Y}^T \mathbf{K}_Y^{-1} \mathbf{k}_Y(\hat{\mathbf{x}}_t^i) = \sum_{i=1}^{N_c} \tilde{w}_t^i \mathbf{y}_t^i$ 。在单峰最高近似对称概率密度分布中, 均值附近的都聚集权重重大粒子, 所以采用接近最大权重值的局部低维粒子计算 \mathbf{y}_t^i 均值会较准确。其中, 在 $p(m_t | \mathbf{x}_t^i)$ 为单峰的情况下, 经过 r 次搜索后, 可得出最优的低维粒子 \mathbf{x}_t^* 为

$$\mathbf{x}_t^* = \lim_{r \rightarrow \infty} {}_r \mathbf{x}_t^{k_1} = \arg \max_{\mathbf{x}_t^i} p(m_t | {}_r \mathbf{x}_t^i) \quad (14)$$

式(14)证明: 因搜索向量 $\tilde{\mathbf{u}}_t^i$ 在 q 维空间每个方向都有均等机会出现, 在 $p(m_t | \mathbf{x}_t^i)$ 为单峰的情况下, 第 r 次搜索, 存在 $\tilde{\mathbf{u}}_t^c, c \in \{1, 2, \dots, N_c - 1\}$, 使得 $\|{}_r \mathbf{x}_t^{k_1} - \mathbf{x}_t^*\| = \|{}_{r-1} \mathbf{x}_t^{k_1} + \tilde{\mathbf{u}}_t^c - \mathbf{x}_t^*\| < \|{}_{r-1} \mathbf{x}_t^{k_1} - \mathbf{x}_t^*\|$, $\|{}_r \mathbf{x}_t^{k_1} - \mathbf{x}_t^*\| < \|{}_r \mathbf{x}_t^i - \mathbf{x}_t^*\|_{i \neq k_1}$, 令 ${}_{r+1} \mathbf{x}_t^{N_c} = {}_r \mathbf{x}_t^{k_1}$ 。第 $r+1$ 次搜索有 ${}_{r+1} \mathbf{x}_t^i = {}_{r+1} \mathbf{x}_t^{N_c} + \tilde{\mathbf{u}}_t^i, i = 1, 2, \dots, N_c - 1$, 同理有 $0 \leq \|{}_{r+1} \mathbf{x}_t^{k_1} - \mathbf{x}_t^*\| < \|{}_r \mathbf{x}_t^{k_1} - \mathbf{x}_t^*\| < \|{}_{r-1} \mathbf{x}_t^{k_1} - \mathbf{x}_t^*\|$, $\|{}_{r+1} \mathbf{x}_t^{k_1} - \mathbf{x}_t^*\| < \|{}_{r+1} \mathbf{x}_t^i - \mathbf{x}_t^*\|_{i \neq k_1}$, 故 $\lim_{r \rightarrow \infty} \|{}_r \mathbf{x}_t^{k_1} - \mathbf{x}_t^*\| = 0$, 即 $\mathbf{x}_t^* = \lim_{r \rightarrow \infty} {}_r \mathbf{x}_t^{k_1}$, 而 $w_t^i = p(m_t | {}_r \mathbf{x}_t^i)$ 与 $1/$

$\|{}_r \mathbf{x}_t^{k_1} - \mathbf{x}_t^*\|$ 单调性一致, 故当 $\lim_{r \rightarrow \infty} {}_r \mathbf{x}_t^{k_1}$ 时, 有 $w_t^i = p(m_t | {}_r \mathbf{x}_t^i)$ 最大, 所以式(14)成立。证毕

通常, 经过首个低维空间的搜索后, 还原得到高维数据的空间位置信息会有偏差, 那么这时需要对还原高维数据的空间位置信息(共6维)进行搜索, 一般是高维数据的前2维数据偏差较大, 搜索前2维即可。搜索方法也使用上述方法进行搜索, 找出权重大的空间位置信息进行保留即可, 之后再进行搜索。对第2个低维空间搜索时, 为了克服轮廓图像数据歧义, 本文建立以下运动形态约束模型, 寻找与首个低维空间搜索后得出的高维数据相对接近的高维数据的相应低维粒子:

$${}^2 \mathbf{y}_t^i < \frac{\gamma}{N_c} \sum_{i=1}^{N_c} \|{}^2 \mathbf{y}_t^i - {}^1 \hat{\mathbf{y}}_t\|^2 \quad (15)$$

其中, ${}^2 \mathbf{y}_t^i$ 为第2个低维空间相应高维数据, γ 为约束调节因子, 可调节候选运动形态的范围。由于第1个低维空间的低维粒子与带空间位置和动作姿态信息的高维数据已建立函数关系, 而且本文估计的人体运动形态是近规则周期性匀速的, 空间位置信息(共6维)与和该位置的动作姿态及各视角的2维图像轮廓数据是一一对应的, 所以该空间搜索后估计得到 ${}^1 \hat{\mathbf{y}}_t$ 的姿态信息对于第2个低维空间搜索是一个很好的先验信息, 在第2个低维空间中, 通过该约束信息选择离该先验信息较近的相应低维粒子再进行搜索, 之后估计其相应高维数据会较好, 即利用该一一对应关系来避免轮廓数据歧义。故式(15)会影响其后式(16)的计算。

本文在图3中简单描述了NWPCS原理, 低维空间局部粒子搜索与传统粒子滤波算法的采样复制有所不同。在概率权重函数局部含有较小多峰时, 按约束条件寻找前3个权重较大的相邻粒子, 对低维空间从全局到局部逐步定位, 可以较快找到与各视角轮廓图像数据特征最近似(最优)且单峰的相应局部, 在该局部中再通过搜索向量进行空间搜索, 获得新的粒子, 然后在所获粒子中再找出权重较大的前3个粒子(含上次搜索权重最大的粒子), 重复上述的方法, 搜索范围会逐步缩小到权重大的粒子生成空间。相比之下, 传统的粒子滤波算法是多次复制权重大的粒子再来采样, 且每次采样范围是很有限的, 先验信息较不理想时, 较小概率权重值归一化后很有可能出现权重比较大的粒子, 会重复复制了实际概率相对最大值都比较小的粒子进行下一轮粒子的权重计算, 就较难得到实际概率大且和轮廓图像特征匹配的粒子, 相当于陷入局部最优的概率值。由于低维空间含有很多高维数据全局先验信息, 所以在低维空间进行全局到局部定位再局部搜

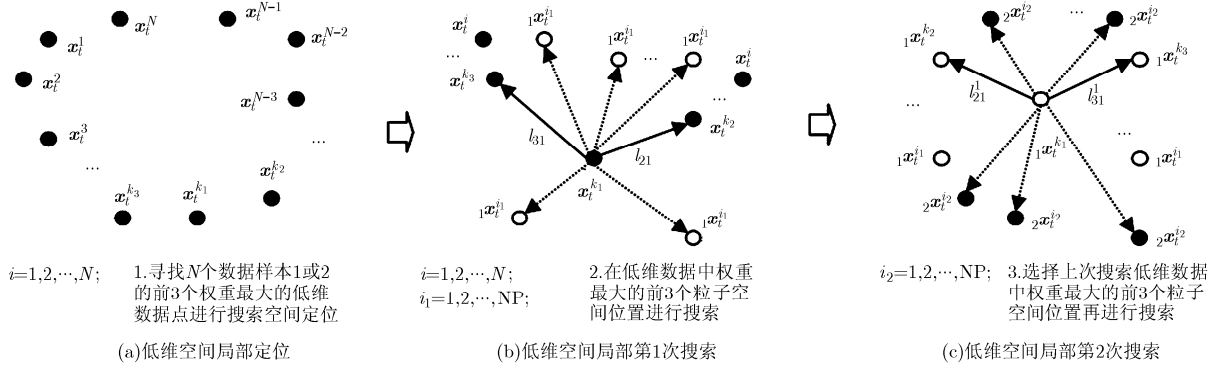


图 3 NWPCS 搜索原理示意图(设 ${}_1x_i^i, {}_2x_i^i$ 分别为 x_i^i 第 1 次, 第 2 次搜索到的粒子, N 为数据样本 1 或 2 的个数, NP 为搜索粒子个数)

索生成的粒子会比传统粒子滤波算法要有效得多, 从而所提算法具有较好的性能。

3.3 双隐变量空间局部粒子搜索(DLVSLPS)算法建立

本文采用双隐变量空间搜索, 可以较准确地估计人体运动立体模型所在的空间位置信息和动作姿态信息。由第 3.2 节的结论, 本文给出算法如下。

已知: 第 t 帧, 镜头视角数为 N_v , 每次搜索粒子数为 NP , 高维数据样本序列 $\mathbf{Y}_1 \in R^{N_1 \times D_1}$ (含空间位置信息), $\mathbf{Y}_2 \in R^{N_2 \times D_2}$ (无空间位置信息, $D_2 = D_1 - 6$), 其相应的低维数据序列 $\mathbf{X}_1 \in R^{N_1 \times q}$, $\mathbf{X}_2 \in R^{N_2 \times q}$, 以及 GPDM 训练好的相关两套参数 $\bar{\alpha}_1, \bar{\beta}_1, \mathbf{W}_1, \bar{\alpha}_2, \bar{\beta}_2, \mathbf{W}_2$, 本文定义 ${}^h x_t^i$ 为第 t 帧中第 h 层搜索过程里第 r 次循环搜索中的第 i 个低维数据, ${}^h r y_t^i$ 为相应高维数据, $\mathbf{y}(n:m)$ 为取向量 \mathbf{y} 的第 n 维到第 m 维数据, end 为最后 1 维数据的序号。本文粒子搜索有 3 层, 每层内循环以 s 次为例, $\hat{\mathbf{y}}_t \in R^{D_1}$, 按式(16)计算:

$$\hat{\mathbf{y}}_t = \left[\begin{array}{c} {}^1 \hat{\mathbf{y}}_t(1:2)^T \quad {}^1 \hat{\mathbf{y}}_t(3:6)^T \quad {}^3 \hat{\mathbf{y}}_t(7:\text{end})^T \\ \left[\begin{array}{c} \arg \max p(m_t | {}^1 \hat{\mathbf{y}}_t(1:2), {}^1 \hat{\mathbf{y}}_t(3:\text{end})) \\ {}^1 \hat{\mathbf{y}}_t(1:2) \end{array} \right]_{\text{位置校正}} \\ = {}^1 \hat{\mathbf{y}}_t(3:\text{end})^T \quad \mathbf{E}({}^1 \mathbf{y}_t^i | m_t)_{\text{第1层计算}}(3:6)^T \\ \mathbf{E}({}^3 \mathbf{y}_t^i | {}^3 \mathbf{y}_t^i(1:6) = {}^1 \hat{\mathbf{y}}_t(1:6), m_t)(7:\text{end})^T \Big]_{\text{第2,3层计算}}^T \quad (16)$$

具体搜索步骤如表 1 所示。

4 仿真实验

本文用 DLVSLPS 算法在估计非连续帧, 克服轮廓图像数据歧义以及估计人体运动形态的误差 3 个方面和其他粒子滤波算法进行比较, 其中所用视角为 3 个视角, 测试数据选用 HumanEva-I 数据。

4.1 估计非连续帧

本文选用任意间隔 20 帧的两帧图像进行测试, 从任意 1 帧开始估计其后的第 20 帧, 且用 4 种不同的算法比较, 其后的第 20 帧估计效果如图 4(a)~图 4(d)所示。图 4 中可以看出所提出的 DLVSLPS

表 1 搜索步骤

(1) 第 1 层在第 1 个隐变量空间局部搜索
步骤 1 如果 $r=1$, 选择 \mathbf{Y}_1 , 计算 $\tilde{w}_t^i, i=1,2,\dots,N_1$; 如果 $1 < r \leq s$, 计算 ${}^1 r y_t^i$ 的权重 $\tilde{w}_t^i, i=1,2,\dots,NP$, 令 $k_{ps} = \text{Ind}_{ps}^3(\mathbf{W}_B), ps=1,2,3$, 有 ${}^1 r x_t^{k_1}, {}^1 r x_t^{k_2}, {}^1 r x_t^{k_3}$, 可得 l_{21}^1, l_{31}^1 , 令 $L_1 = \max(l_{21}^1, l_{31}^1)$, 若 $r=s$, 执行步骤 3。
步骤 2 令 $\tilde{\mathbf{u}}_t^i = \eta_i L_1(\mathbf{u}_t^i / \ \mathbf{u}_t^i\)$, 获取新 NP-1 个粒子 $\hat{x}_t^i = {}^1 r x_t^{k_1} + \tilde{\mathbf{u}}_t^i, i=1,2,\dots,NP-1$, 令 $\hat{x}_t^{NP} = {}^1 r x_t^{k_1}$, 使 $r=r+1$, ${}^1 r y_t^i = \hat{\mathbf{y}}_t^i = \mathbf{Y}_1^T \mathbf{K}_{Y_1}^{-1} \mathbf{k}_{Y_1}(\hat{x}_t^i), {}^1 r x_t^i = \hat{x}_t^i$, 返回步骤 1。
步骤 3 令 $\mathbf{E}({}^1 r y_t^i m_t) = {}^1 \hat{\mathbf{y}}_t = \sum_{i=1}^{NP} \tilde{w}_t^i {}^1 r y_t^i$, 重置 $r=1$ 。
提取 ${}^1 \hat{\mathbf{y}}_t, {}^1 r y_t^{k_2}, {}^1 r y_t^{k_3}$ 的前 2 维空间位置信息(共 6 维), 以类似的搜索方法, 搜索该 2 维空间位置信息, 使 ${}^1 \hat{\mathbf{y}}_t(1:2)^T = \arg \max p(m_t {}^1 \hat{\mathbf{y}}_t(1:2), {}^1 \hat{\mathbf{y}}_t(3:\text{end}))^T$ 。
(2) 第 2 层在第 2 个隐变量空间局部搜索
步骤 1 如果 $r=1$, 选择 \mathbf{Y}_2 , 令 ${}^2 r y_t^i = [{}^1 \hat{\mathbf{y}}_t(1:6)^T, \mathbf{y}_{2i}^T]^T$, \mathbf{y}_{2i} 为 \mathbf{Y}_2 序列的一个样本, 计算 $\tilde{w}_t^i, i=1,2,\dots,N_2$, 及 ${}^1 \hat{\mathbf{y}}_t$ 的 \tilde{w}_a , 若 $\tilde{w}_a > \tilde{w}_t^i, i=1,2,\dots,N_2$, 则令 $\hat{\mathbf{y}}_t = {}^1 \hat{\mathbf{y}}_t$, 终止算法, 否则根据式(15), 淘汰不满足约束的粒子及权重; 如果 $1 < r \leq s$, 令 ${}^2 r y_t^i = [{}^1 \hat{\mathbf{y}}_t(1:6)^T, ({}^2 r y_{tU}^i)^T]^T$, 计算 $\tilde{w}_t^i, i=1,2,\dots,NP$ 。令 $k_{ps} = \text{Ind}_{ps}^3(\mathbf{W}_B), ps=1,2,3$, 有 ${}^2 r x_t^{k_1}, {}^2 r x_t^{k_2}, {}^2 r x_t^{k_3}$, 得出 l_{21}^2, l_{31}^2 , 令 $L_2 = \max(l_{21}^2, l_{31}^2)$, 若 $r=s$, 则执行第 3 层的搜索, 并且重置 $r=1$ 。
步骤 2 令 $\tilde{\mathbf{u}}_t^i = \eta_i L_2(\mathbf{u}_t^i / \ \mathbf{u}_t^i\)$, 获取新 NP-1 个粒子 $\hat{x}_t^i = {}^2 r x_t^{k_1} + \tilde{\mathbf{u}}_t^i, i=1,2,\dots,NP-1$, 令 $\hat{x}_t^{NP} = {}^2 r x_t^{k_1}$, 使 $r=r+1$, ${}^2 r y_{tU}^i = \hat{\mathbf{y}}_t^i = \mathbf{Y}_2^T \mathbf{K}_{Y_2}^{-1} \mathbf{k}_{Y_2}(\hat{x}_t^i), {}^2 r x_t^i = \hat{x}_t^i$, 返回步骤 1。
(3) 第 3 层在第 2 个隐变量空间再局部搜索
步骤 1 如果 $r=1$, 使 ${}^3 r x_t^{k_1} = {}^2 r x_t^{k_1}$, 令 $L_3=L_2$, 执行步骤 2; 如果 $1 < r \leq s$, 令 ${}^3 r y_t^i = [{}^1 \hat{\mathbf{y}}_t(1:6)^T, ({}^3 r y_{tU}^i)^T]^T$, 计算 $\tilde{w}_t^i, i=1,2,\dots,NP$ 。令 $k_{ps} = \text{Ind}_{ps}^3(\mathbf{W}_B), ps=1,2,3$, 有 ${}^3 r x_t^{k_1}, {}^3 r x_t^{k_2}, {}^3 r x_t^{k_3}$, 得出 l_{21}^3, l_{31}^3 , 令 $L_3 = \max(l_{21}^3, l_{31}^3)$, 若 $r=s$, 执行步骤 3。
步骤 2 令 $\tilde{\mathbf{u}}_t^i = \eta_i L_3(\mathbf{u}_t^i / \ \mathbf{u}_t^i\)$, 获取新 NP-1 个粒子 $\hat{x}_t^i = {}^3 r x_t^{k_1} + \tilde{\mathbf{u}}_t^i, i=1,2,\dots,NP-1$, 令 $\hat{x}_t^{NP} = {}^3 r x_t^{k_1}$, 使 $r=r+1$, ${}^3 r y_{tU}^i = \hat{\mathbf{y}}_t^i = \mathbf{Y}_2^T \mathbf{K}_{Y_2}^{-1} \mathbf{k}_{Y_2}(\hat{x}_t^i), {}^3 r x_t^i = \hat{x}_t^i$, 返回步骤 1。
步骤 3 计算高维数据 $\hat{\mathbf{y}}_t = \mathbf{E}({}^3 r y_t^i {}^3 r y_t^i(1:6) = {}^1 \hat{\mathbf{y}}_t(1:6), m_t) = \sum_{i=1}^{NP} \tilde{w}_t^i {}^3 r y_t^i$, 输出该帧带空间位置信息的人体运动立体模型, 算法结束。

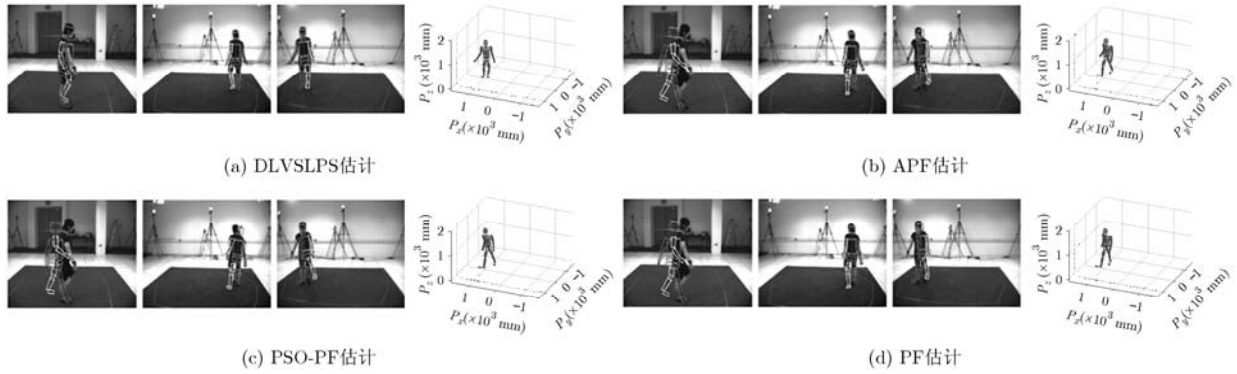


图 4 各算法在 3 个视角下估计非连续帧的人体运动立体模型及其 2 维图像显示

算法可以很好地估计出非连续帧的人体运动立体模型及其空间位置，并且其各肢体和该非连续帧各视角的 2 维图像也很匹配，然而 APF, PSO-PF 和 PF 算法估计出非连续帧的人体运动立体模型及其空间位置均有误，其各肢体与相应各视角 2 维图像明显不匹配。

4.2 克服轮廓图像数据歧义估计

一组有限视角如图 2 第 2 行的 3 幅图这样的 2 维轮廓图像数据是带有数据歧义的，也就是说一组轮廓数据无法判别人体运动立体模型肢体的具体情况。例如，一组有限视角轮廓图像数据无法判别人体运动形态的脚到底是左脚在前还是右脚在前的类似情况。DLVSLPS 算法利用低维空间的先验信息，加上运动形态约束，可克服轮廓图像数据歧义估计，给出更接近真实数据的人体运动立体模型。如图 5(a)~图 5(e)所示，DLVSLPS 算法估计效果和真实数据较接近，而 APF, PF 算法估计左脚(白色)及右脚(灰色)的位置相反了，手有和 2 维图像不匹配的现象，并且与真实数据相差很大。PSO-PF 算法更无法克服轮廓图像数据歧义，估计的效果和真实数

据差别更大，得出人体运动立体模型完全不成人形。因此，说明 DLVSLPS 算法估计过程中可以克服轮廓数据歧义进行估计。

4.3 估计误差

本文采用文献[16]的计算误差方法，计算镶嵌在圆台里的骨架各关节点的平均误差作为每一帧的误差，然后采用 APF, PSO-PF, PF 3 种粒子滤波算法与所提算法 DLVSLPS 比较。由于 APF, PSO-PF, PF 算法估计非连续帧不是很好，故采用 50 帧连续图像，比较估计误差。本文采用 HumanEva-I 里空间位置信息变化较大的人体走路和跑步形态的数据进行测试比较，其中走路 1~3 与跑步 1~3 两类运动形态数据段都是空间位置信息不同的数据，如图 6 所示，各子图中可以分别看出所提算法 DLVSLPS 估计连续帧的平均误差，标准差，走路 1~3 和跑步 1~3 每帧的误差及最大误差(图 6(b)~图 6(g)的算法图例标识所示)，都比 APF, PSO-PF, PF 算法总体明显要小，所以 DLVSLPS 算法比以上 3 种算法有很大优势。

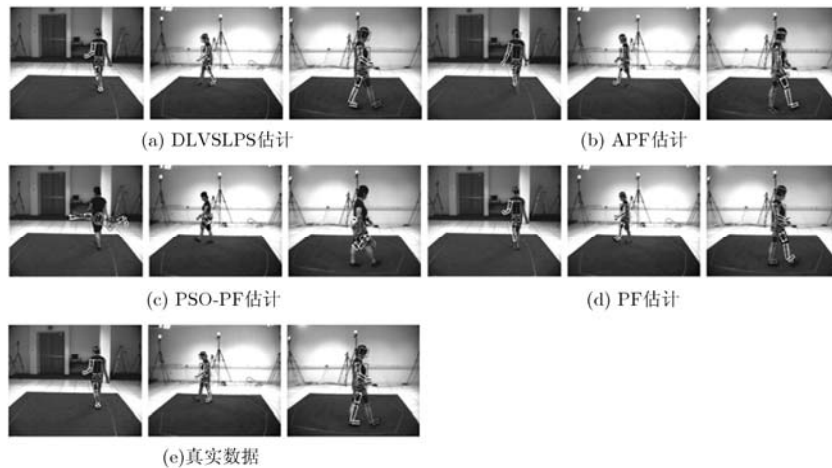


图 5 各算法在 3 个视角下克服轮廓数据歧义估计人体运动立体模型的 2 维图像显示

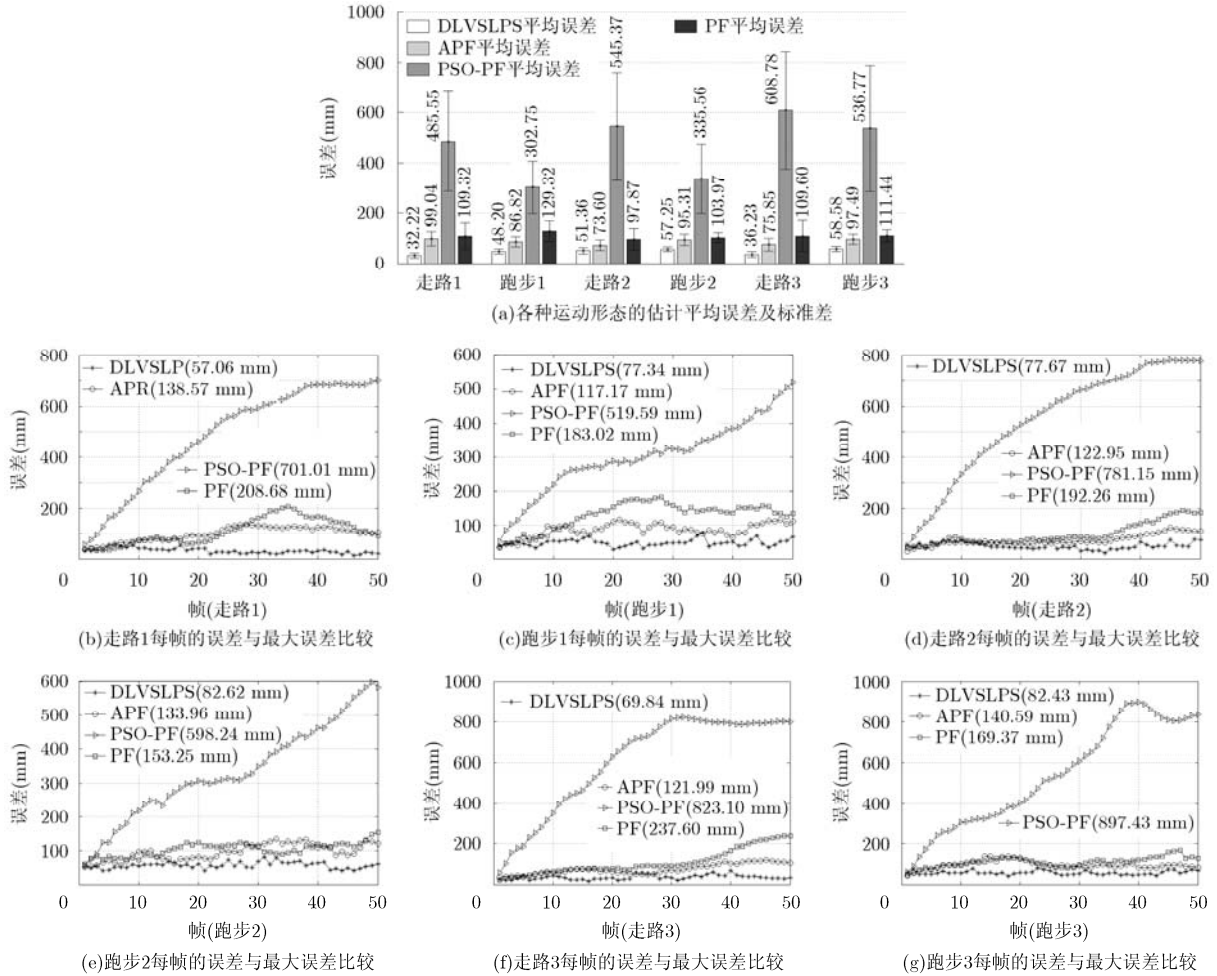


图 6 估计误差比较

5 结束语

从以上实验结果可知，所提 DLVSLPS 算法可以从多视角图像序列中的人体运动形态轮廓特征估计出相应的 3 维人体运动形态，并且可以较好地估计出其所含的相应空间位置信息。和其他传统粒子滤波算法 APF, PF, PSO-PF 相比，DLVSLPS 算法还有非连续帧估计，克服轮廓图像数据歧义及减小估计误差的优势。另外，对多视角的图像序列进行图像分割提取轮廓图像时，如果轮廓图像提取的质量较差，会直接对实验结果产生较大影响。本文的研究还局限于人体运动形态及空间位置变化较规则，且有标记人体运动立体模型的估计，将来本文的研究还会拓展到人体运动空间位置变化不规则，无标记且多角色交互复杂的人体运动立体模型估计^[7]。如何在低维空间中进行最优搜索，与高维数据相互建立函数关系，仍是研究重点。

参考文献

[1] Yao A, Gall J, Gool L V, et al. Learning probabilistic

non-linear latent variable models for tracking complex activities[C]. Proceedings of Neural Information Processing Systems, Vancouver, Canada, 2011: 1-8.

[2] Luo Wei-lan, Yamasaki T, and Aizawa K. Cooperative estimation of human motion and surfaces using multiview videos[J]. *Computer Vision and Image Understanding*, 2013, 117(11): 1560-1574.

[3] Gall J, Rosenhahn B, Brox T, et al. Optimization and filtering for human motion capture[J]. *International Journal of Computer Vision*, 2010, 87(1/2): 75-92.

[4] Bernard J, Wilhelm N, Krüger B, et al. Motionexplorer: exploratory search in human motion capture data based on hierarchical aggregation[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2013, 19(12): 2257-2266.

[5] Sigal L, Balan A O, and Black M J. Humaneva: synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion[J]. *International Journal of Computer Vision*, 2010, 87(1/2): 4-27.

[6] Deutscher J and Reid I. Articulated body motion capture by stochastic search[J]. *International Journal of Computer*

- Vision*, 2005, 61(2): 185–205.
- [7] 朱志宇. 粒子滤波算法及其应用[M]. 北京: 科学出版社, 2010: 78–83.
Zhu Zhi-yu. Particle Filter Algorithm and Its Application[M]. Beijing: Science Press, 2010: 78–83.
- [8] 李毅, 孙正兴, 陈松乐, 等. 基于退火粒子群优化的单目视频人体姿态分析方法[J]. 自动化学报, 2012, 38(5): 732–741.
Li Yi, Sun Zheng-xing, Chen Song-le, *et al.*. 3D human pose analysis from monocular video by simulated annealed particle swarm optimization[J]. *Acta Automatica Sinica*, 2012, 38(5): 732–741.
- [9] Wang J M and Fleet D J. Gaussian process dynamical models for human motion[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(3): 283–298.
- [10] Urtasun R, Fleet D J, and Fua P. 3D people tracking with gaussian process dynamical models[C]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, USA, 2006: 238–245.
- [11] Rosman G, Bronstein M M, Bronstein A M, *et al.*. Nonlinear dimensionality reduction by topologically constrained isometric embedding[J]. *International Journal of Computer Vision*, 2010, 89(1): 56–68.
- [12] Lawrence N D. Probabilistic non-linear principal component analysis with Gaussian process latent variable models[J]. *Machine Learning Research*, 2005, 6(11): 1783–1816.
- [13] Ek C H, Torr P H S, and Lawrence N D. Gaussian process latent variable models for human pose estimation[C]. Proceedings of 4th Workshop on Machine Learning for Multimodal Interaction, Brno, Czech Republic, 2007: 132–143.
- [14] Dong Fang-fang, Chen Zeng-si, and Wang Jin-wei. A new level set method for inhomogeneous image segmentation[J]. *Image and Vision Computing*, 2013, 31(10): 809–822.
- [15] Kasaiezadeh A and Khajepour A. Multi-agent stochastic level set method in image segmentation[J]. *Computer Vision and Image Understanding*, 2013, 117(9): 1147–1162.
- [16] Sigal L and Black M J. Humaneva: synchronized video and motion capture dataset for evaluation of articulated human motion[R]. Report of Brown University, Providence: USA, 2006.
- [17] Liu Ye-bin, Gall J, Stoll C, *et al.*. Markerless motion capture of multiple characters using multiview image segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(11): 2720–2735.
- 李万益: 男, 1985年生, 博士生, 研究方向为机器学习、智能计算方法、通信系统优化.
- 孙季丰: 男, 1962年生, 教授, 博士生导师, 研究方向为图像理解、计算机视觉.
- 王玉龙: 男, 1989年生, 硕士生, 研究方向为图像与视频处理.