

基于混合跟踪模型的室内步行人体 3D 运动估计

于雪松¹ 赵巍¹ 刘鹏¹ 唐降龙¹

摘要 针对步行人体 3D 运动估计过程中的自遮挡问题, 提出了基于混合跟踪模型的粒子滤波算法. 首先, 利用自遮挡状态检测模型, 将步行人体运动划分为四种自遮挡状态; 其次, 根据混合跟踪模型, 针对不同的自遮挡状态, 算法采用不同的跟踪模型; 最后, 为了估计遮挡状态下的人体运动, 算法提出了基于 M-估计的在线训练方法以训练肢体运动相关系数. 经过实验分析, 算法对处于自遮挡状态下的人体 3D 运动估计有着良好的效果, 人体 3D 运动的估计精度得到了提高.

关键词 混合跟踪模型, 步行人体运动, 自遮挡状态检测模型, M-估计, 肢体运动相关系数

DOI 10.3724/SP.J.1004.2010.00773

Estimating the Pedestrian 3D Motion Indoor via Hybrid Tracking Model

YU Xue-Song¹ ZHAO Wei¹ LIU Peng¹ TANG Xiang-Long¹

Abstract Focusing on self-occlusion of pedestrian 3D motion estimation, the paper proposed a hybrid tracking model particle filter algorithm. First, using the self-occlusion state detecting model, the pedestrian motion can be detected and divided into four self-occlusion states. Second, via hybrid tracking model, different tracking patterns are proposed to track pedestrian motion on different self-occlusion states. And finally, for estimating the pedestrian motion on the self-occlusion state, we proposed the on-line training algorithm based on M-estimate to train the limbs motion correlation coefficient. The result of experiment showed that our algorithm acquires good results of estimating pedestrian 3D motion on the occlusion state and advances the accuracy of estimating pedestrian 3D motion.

Key words Hybrid tracking model, pedestrian 3D motion, self-occlusion state detection model, M-estimator, limbs motion correlation coefficient

近年来, 人体 3D 运动估计成为计算机视觉领域的研究热点之一^[1], 越来越多的研究学者将重点转向了基于单目视频无标记点的人体 3D 运动估计. 然而, 由于运动过程中的人体自遮挡现象所导致的二义性问题严重阻碍了此方向上研究的发展.

本文对 HumanEva^[2] 实验数据集中的室内人体行走视频图像进行了统计分析, 经过分析, 数据集中的室内人体行走视频具有以下特点: 1) 在步行人体运动过程中, 无自遮挡 (正常状态) 行走和处于自遮挡状态的行走交替出现; 2) 通过选取适当的特征, 人体肢体之间的自遮挡具有明显的规律性.

在此基础上, 针对前文所述问题, 本文提出了基于混合跟踪模型的粒子滤波算法 (Hybrid tracking model particle filter, HTMPF), 用于室内步行人体 3D 运动估计:

1) 无遮挡跟踪模式: 本文利用文献 [3] 中的离散关节图模型, 将高维人体状态空间分解为若干个低维状态空间, 从而达到降低状态空间维数的目的. 在此基础上, 基于序列蒙特卡罗 (Sequential Monte

Carlo, SMC) 粒子滤波器 (Particle filter, PF) 被用来估计无遮挡状态下的人体 3D 运动.

2) 有遮挡跟踪模式: 在人体运动过程中, 步行人体的左侧肢体和右侧肢体之间有很强的内在相关性^[4], 而利用这种内在相关性可以缓解因为人体自遮挡所带来的二义性问题. 于是, 根据步行人体运动中的肢体运动相关性系数, 本文利用 Rao-Blackwellised 粒子滤波 (Rao-Blackwellised particle filter, RBPF) 对自遮挡状态下的人体 3D 运动进行估计.

为了获得肢体运动相关性系数, 本文提出了一种基于 M-估计的肢体运动相关性系数在线训练方法. 该方法利用将每一帧的跟踪结果作为训练样本, 利用 M-估计 (M-estimation) 算法在线训练肢体运动相关性系数. 为了统计人体肢体之间自遮挡状态的规律性, 我们对步行人体的肤色特征进行统计分析, 提出了一种基于“由粗到细” (Coarse-to-fine) 策略的自遮挡检测方法, 按照自顶向下 (Top-down) 顺序依次对人体部分肢体进行检测, 并根据检测结果判断是何种跟踪模式.

1 相关工作

目前, 基于单目视频无标记点的人体 3D 运动估计的研究策略主要有两种: 1) 基于模型的方法; 2) 基于学习的方法.

收稿日期 2009-04-23 录用日期 2009-10-23
Manuscript received April 23, 2009; accepted October 23, 2009
国家自然科学基金 (60672090) 资助
Supported by National Natural Science Foundation of China (60672090)

1. 哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001
1. School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001

基于模型的方法^[5-7] 利用人体运动先验知识来建立人体 3D 运动先验模型, 并通过定义似然函数, 将运动估计转化为状态参数空间的优化问题. 在人体 3D 先验模型的基础上, 作为一种基于贝叶斯估计的非线性滤波算法, 粒子滤波^[8], 以及退火粒子滤波^[9] (Annealed particle filter, APF) 和序列蒙特卡罗^[10] 粒子滤波等各种粒子滤波改良算法在人体 3D 运动跟踪领域得到了广泛的应用.

基于学习的方法^[11-13] 通过对低维参数空间中本征变量模型的学习将高维人体状态空间分解为线性子空间, 以提高计算效率. Urtasun 等^[14] 利用非线性高斯过程动态模型 (Nonlinear Gaussian process dynamic model, NGPDM) 来解决由于高维状态参数空间而导致的跟踪效率低下的问题. 而 Agarwal 等^[15] 用矢量量化编码后的图像侧影的形状上下文 (Shape context) 来表示图像特征向量, 并通过机器学习方法学习特征向量与 3D 运动数据之间的关系; 并利用 Poser 软件中的 3D 人体模型库来取代人体运动模型的学习和存储以解决学习阶段所需的大样本量需求^[16]. Xu 等^[4, 17] 利用已获得的步行人体的 3D 真实数据, 通过非线性迭代偏最小二乘算法 (Nonlinear iteratively partial least square, NIPLS) 学习并获得步行人体运动的相关性系数, 并结合 Rao-Blackwellised 粒子滤波^[18], 提出了一种 RBPF-PLS (Rao-Blackwellised particle filter partial least square) 算法来对处于自遮挡状态下人体的行走姿态进行跟踪, 在提高计算效率的同时, 缓解了人体运动过程中的自遮挡问题.

本文的结构安排如下: 在第 2 节中, 详细地介绍了混合跟踪模型中的两种跟踪模式; 第 3 节详细地介绍了基于自顶向下策略的自遮挡检测算法; 第 4 节给出了似然函数的定义; 第 5 节则介绍了基于迭代重加权最小二乘法 (Iteratively reweighted least squares, IRLS) 的肢体运动关系系数的训练; 第 6 节, 给出了算法应用在 HumanEva 室内人体行走实验数据集的结果, 验证了算法的有效性; 最后, 得出结论.

2 混合跟踪模型

在本节中, 详细地定义了用于描述人体运动的人体生理结构约束模型及其在二维图像平面中的表现形式, 并借助于离散关节图模型以及人体运动过程中的肢体之间的相关性, 给出了混合跟踪模型 (Hybrid tracking model, HTM) 所包括的两种跟踪模式的详细定义.

2.1 人体模型

在本文中, 人体模型包括人体生理结构模型和

离散关节图模型. 人体生理结构模型可由人体 3D 骨骼模型^[9] 来表示, 同时, 为了显示跟踪结果, 人体各部分肢体的 3D 表示模型利用锥形圆柱体表示, 相应的人体的 2D 表示模型则利用纸板模型来表示. 若令 φ^i 表示锥形圆柱体参数集合, 则 φ^i 可由下式表示:

$$\varphi^i = \{l^i, \omega^{i,p}, \omega^{i,d}, o^{i,p}, o^{i,d}\} \quad (1)$$

其中, i 为肢体编号; l^i 表示肢体的长度; $\omega^{i,p}, \omega^{i,d}$ 分别为肢体起始端和末端关节所在圆平面的半径; $o^{i,p}, o^{i,d}$ 为肢体起始端和末端关节 3D 坐标的偏移量. 为了减少计算量, φ^i 定义为常数向量.

为了简化计算, 本文引入了离散关节图模型^[3], 如图 1 所示; 利用图的分解, 离散关节图模型可以把人体的整体跟踪简化为按肢体局部独立跟踪. 根据离散关节图模型, 人体运动状态参数空间可被定义为:

$$X = \{x^1, x^2, \dots, x^{10}\} \quad (2)$$

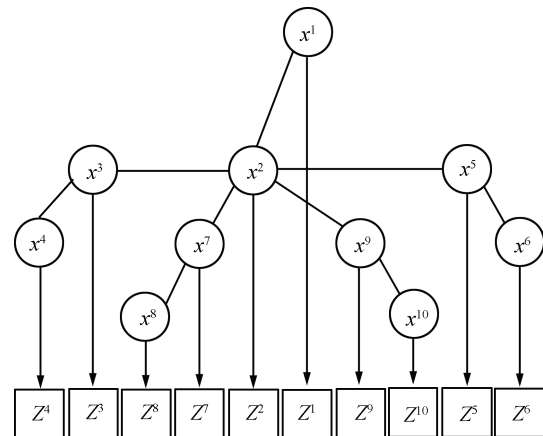


图 1 离散关节图模型

Fig. 1 The decentralized articulated graphical model

2.2 混合跟踪模型

令 X 表示人体运动状态参数空间, Z 为相应的观测状态, 则在任意时刻 t , 混合跟踪模型可被定义如下:

$$P(X_t) = (1 - \lambda_t) \times \varphi(X_t) + \lambda_t \times \psi(X_t) \quad (3)$$

其中, λ_t 为遮挡因子, $\lambda_t = 1$ 时, 步行人体处于自遮挡状态, $\lambda_t = 0$, 则处于正常行走状态, 其计算参见第 3 节; $\varphi(X_t)$ 为无遮挡跟踪模式下概率分布; $\psi(X_t)$ 为自遮挡跟踪模式下概率分布.

2.2.1 无遮挡跟踪模式

在不考虑人体肢体的形状参数 (采用固定值) 的情况下, 在任意时刻 t , 式 (2) 中的 x^i 可由下式定义:

$$x_t^i = \{\rho_t^i, \theta_t^i\}, \quad i = 1, \dots, 10 \quad (4)$$

其中, ρ_t^i 为第 i 个肢体末端关节的三维坐标向量; θ_t^i 为第 i 个肢体末端关节绕三维坐标轴旋转的角度向量. 并且, 根据人体生理结构模型, ρ_t^i 和 θ_t^i 之间可通过计算而相互转换.

在任意时刻 t , 定义人体第 i 个肢体的运动状态为 x_t^i , 其观测状态为 z_t^i ; 于是, 利用 Gibbs 分布和文献 [19] 的推导, 式 (3) 中 $\varphi(X_t)$ 可由下式定义:

$$\varphi(X_t) = P(X_t) = \prod_{i=1}^C p(x_t^i | z_t^i) \quad (5)$$

其中, C 为肢体个数, $C = 10$. 令 $\mathbb{N}(i)$ 表示人体第 i 个肢体的邻近肢体集合, 则根据文献 [3] 中的推导, 式 (5) 中每一个肢体的后验概率 $p(x_t^i | z_t^i)$ 可由式 (6) 计算.

$$p(x_t^i | z_t^i) = p(x_t^i | z_t^i, z_t^{\mathbb{N}(i)}) \propto p(z_t^i | x_t^i) p(x_t^i | x_{t-1}^i) \times p(x_{t-1}^i | z_{t-1}^i, z_{t-1}^{\mathbb{N}(i)}) \prod_{j \in \mathbb{N}(i)} \int p(z_t^j | x_t^j) p(x_t^j | x_t^i) dx_t^j \quad (6)$$

在式 (6) 中, $p(z_t^i | x_t^i)$ 为当前肢体 i 的局部似然概率; $p(x_t^i | x_{t-1}^i)$ 为当前肢体运动状态转移模型; $p(x_{t-1}^i | z_{t-1}^i, z_{t-1}^{\mathbb{N}(i)})$ 为当前肢体 i 的第 $t-1$ 时刻的后验概率; 积分中的 $p(z_t^j | x_t^j)$ 表示当前肢体 i 的邻近肢体 j 的局部似然模型; $p(x_t^j | x_t^i)$ 为相连肢体关系模型, 该模型可由下式定义:

$$p(x_t^j | x_t^i) = p(\rho_t^j | \rho_t^i) p(\theta_t^j | \theta_t^i) \quad (7)$$

其中, $p(\theta_t^j | \theta_t^i)$ 为两个相连肢体的姿态角度模型, 该模型可由人体生理结构先验知识构造, 按照不同的肢体部位, 本文利用先验知识为不同的相连肢体分别建立相应的均匀分布; $p(\rho_t^j | \rho_t^i)$ 为两个相连肢体的位置信息模型^[3, 19], 该模型定义如下:

$$p(\rho_t^j | \rho_t^i) = \frac{\exp\left\{-\frac{(\rho_t^j - \rho_t^i) \Sigma^T (\rho_t^j - \rho_t^i)}{2}\right\}}{2\pi |\Sigma_\rho|^{\frac{1}{2}}} \quad (8)$$

其中, ρ_t^j 和 ρ_t^i 分别为两个相连肢体的特征点, 可由肢体形状参数和末端关节坐标 ρ_t^i 计算获得, 根据文献 [19], ρ_t^j 和 ρ_t^i 关系如图 2 所示; Σ_ρ 为多变量高斯分布的协方差矩阵, 为了简化计算, Σ_ρ 设定为对角阵.

为了计算式 (6), 本文利用序列蒙特卡罗算法进行近似表示. 假设在 t 时刻, 后验概率 $p(x_t^i | z_t^i, z_t^{\mathbb{N}(i)})$ 可由 $N(i)$ 个加权粒子 $\{x_t^{i,k}, \omega_t^{i,k}\}_{k=1}^{N(i)}$ 近似计算; 其中, $N(i)$ 为粒子个数, $\omega_t^{i,k}$ 为 t 时刻第 i 个肢体的第 k 个粒子的权重, 且 $\omega_t^{i,k}$ 可由式 (9) 更新.

$$\omega_t^{i,k} \propto \omega_{t-1}^{i,k} \times \frac{p(z_t^i | x_t^{i,k}) p(x_t^{i,k} | x_{t-1}^{i,k})}{q(x_t^{i,n} | x_{t-1}^{i,k}, z_t^i, z_t^{\mathbb{N}(i)})} \times \prod_{j \in \mathbb{N}(i)} \left\{ \sum_{l=1}^{N_s^j} p(z_t^j | x_t^{j,l}) p(x_t^{j,l} | x_t^{j,k}) \right\} \quad (9)$$

其中, N_s^j 为第 j 个肢体的粒子数目, $q(x_t^{i,n} | x_{t-1}^{i,k}, z_t^i, z_t^{\mathbb{N}(i)})$ 为提议分布函数, 本文选择文献 [19] 中的方法进行计算; $p(x_t^{j,l} | x_t^{j,k})$ 可由式 (8) 计算. 对于当前肢体 i 的 $N(i)$ 个粒子, 则可由肢体运动状态转移模型来产生:

$$x_t^i = x_{t-1}^i + N(\mu^i, \Sigma_P^i) \quad (10)$$

其中, 高斯分布 $N(\mu^i, \Sigma_P^i)$ 的期望 μ^i 为 3×1 的标量, 其值为当前肢体 i 的运动速度; Σ_P^i 为 3×3 的对角阵.

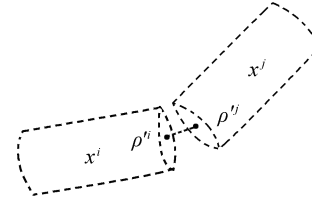


图 2 两个相连肢体之间的位置信息模型

Fig. 2 The position model between two articulated limbs

2.2.2 自遮挡跟踪模式

Xu 等^[4, 17] 认为: 在人体运动过程中, 人体的左部肢体 L (包括左臂和左腿) 和右部肢体 R (包括右臂和右腿) 之间存在着强相关性, 记为 B . 根据这种相关性, 结合 Rao-Blackwellised 离子滤波算法, 可以缓解人体跟踪过程中的自遮挡、二义性等问题. 按照上述思路, 本文提出了自遮挡跟踪模型以用来对处于自遮挡状态下的步行人体运动跟踪.

由于在人体运动过程中, 左侧肢体和右侧肢体都有可能被遮挡, 因此, 我们定义 \varkappa 为处于自遮挡状态下的肢体运动状态集合, Γ 为未遮挡的肢体运动状态集合, 则根据文献 [4], 在任意时刻 t , $\varkappa_t \subset X_t$, $\Gamma_t \subseteq X_t$, 令 Γ_t 为根变量, \varkappa_t 为叶变量. 于是, 式 (3) 中的 $\psi(X_t)$ 可被定义为式 (11).

$$\psi(X_t) = P(X_t | Z_t) = p(x_t | z_t) = p(\Gamma_t, \varkappa_t | z_t, B_t) \quad (11)$$

根据 Rao-Blackwellised 粒子滤波的原理, 在文献 [4, 17] 中, Xu 等利用序列蒙特卡罗算法对 $p(\Gamma_t | z_t)$ 进行逼近, 于是, 有下式成立:

$$p(\Gamma_t, \varkappa_t | z_t, B_t) = p(\varkappa_t | \Gamma_t, z_t, B_t) p(\Gamma_t | z_t) \propto p(z_t | \Gamma_t, \varkappa_t, B_t) \times p(\varkappa_t | \Gamma_t, z_t, B_t) \times \int p(\Gamma_t | \Gamma_{t-1}) p(\Gamma_{t-1} | z_{t-1}) d\Gamma_{t-1} \quad (12)$$

为了减少计算量, 本文利用离散关节模型计算后验概率 $p(\Gamma_t, \boldsymbol{x}_t | z_t, B_t)$. 令 Λ_t 表示时刻 t 的躯干和头部运动参数空间, \mathfrak{R}_t 表示时刻 t 未被遮挡的肢体运动参数空间, 则 $\Gamma_t = \Lambda_t \cup \mathfrak{R}_t$, 其中, \boldsymbol{x}_t 和 \mathfrak{R}_t 可由下式定义:

$$\begin{aligned} \mathfrak{R}_t &= \lambda_t^S \times R_t + (1 - \lambda_t^S) \times L_t \\ \boldsymbol{x}_t &= \lambda_t^S \times L_t + (1 - \lambda_t^S) \times R_t \end{aligned} \quad (13)$$

其中, λ_t^S 为当前时刻 t 的肢体遮挡状态因子, 其计算过程参见第 3 节, $\lambda_t^S = 1$ 时, 表示左侧肢体被遮挡; L_t 表示在当前时刻 t 左侧肢体运动参数向量; R_t 表示在当前时刻 t 右侧肢体运动参数向量; 并且, 在本文中 \boldsymbol{x}_t 和 \mathfrak{R}_t 仅考虑式 (4) 中的角度向量. 于是, 根据式 (5), 式 (11) 可改写为

$$\begin{aligned} \psi(X_t) &= p(x_t | z_t) = \prod_{i=1}^C p(x_t^i | z_t^i) = \\ & \frac{p(\boldsymbol{x}_t, \mathfrak{R}_t | z_t^{\mathfrak{R}_t}, B_t)}{\text{步骤 1}} \times \frac{\prod_{i=1}^{C_1} p(\Lambda_t^i | z_t^i)}{\text{步骤 2}} \end{aligned} \quad (14)$$

其中, $C_1 = 2$; $z_t^{\mathfrak{R}_t}$ 为 t 时刻 \mathfrak{R}_t 的观测状态, 则 $p(\boldsymbol{x}_t, \mathfrak{R}_t | z_t^{\mathfrak{R}_t}, B_t)$ 表示 t 时刻在肢体运动关系矩阵 B_t 和 $z_t^{\mathfrak{R}_t}$ 条件下的后验概率.

步骤 1. $p(\boldsymbol{x}_t, \mathfrak{R}_t | z_t^{\mathfrak{R}_t}, B_t)$ 可根据文献 [4] 中的方法进行计算:

$$\begin{aligned} p(\boldsymbol{x}_t, \mathfrak{R}_t | z_t^{\mathfrak{R}_t}, B_t) &= p(\boldsymbol{x}_t | \mathfrak{R}_t, z_t^{\mathfrak{R}_t}, B_t) p(\mathfrak{R}_t | z_t^{\mathfrak{R}_t}) \times \\ & \underbrace{p(z_t^{\mathfrak{R}_t} | \boldsymbol{x}_t, \mathfrak{R}_t)}_{\text{Step 3}} \underbrace{p(\boldsymbol{x}_t, \mathfrak{R}_t | z_t^{\mathfrak{R}_t}, B_t)}_{\text{Step 2}} \times \\ & \underbrace{\int p(\mathfrak{R}_t | \mathfrak{R}_{t-1}) p(\mathfrak{R}_{t-1} | z_{t-1}^{\mathfrak{R}_{t-1}}) d\mathfrak{R}_{t-1}}_{\text{Step 1}} \end{aligned} \quad (15)$$

上式中的 Step1 为基于肢体运动相关系数矩阵 B_t 的肢体 \mathfrak{R}_t^i 和 \boldsymbol{x}_t^i 联合局部似然概率; Step2 和 Step3 中的运动状态转移模型, 以及粒子重采样等关键步骤, 本文均采用文献 [4] 中的方法.

步骤 2. 对于 $p(\Lambda_t^i | z_t^i)$, 本文利用离散关节图模型进行计算, 其计算式如下:

$$\begin{aligned} p(\Lambda_t^i | z_t^i) &= p(\Lambda_t^i | z_t^i, z_t^{N'(i)}) \propto p(z_t^i | \Lambda_t^i) p(\Lambda_t^i | \Lambda_{t-1}^i) \times \\ & p(\Lambda_{t-1}^i | z_{t-1}^i, z_{t-1}^{N'(i)}) \prod_{j \in N'(i)} \int p(z_t^j | \Gamma_t^j) p(\Gamma_t^j | \Lambda_t^i) d\Gamma_t^j \end{aligned} \quad (16)$$

其中, $N'(i) \subset \Gamma_t$, 表示在当前时刻 t , 当前肢体 i 的邻域集合 $N'(i)$ 是未被遮挡的集合 Γ_t 的一个子集,

若当前肢体 i 的某个邻域处于被遮挡状态, 那么在计算时该邻域将被忽略, 显然, $N'(i) \subseteq N(i)$. 于是, 上式可按照无遮挡跟踪模型中的方法计算.

3 自遮挡的检测

本节简要地介绍了用于自遮挡检测的人体特征及其提取方法; 利用提取出来的肤色特征, 我们对实验数据进行了分析. 在此基础上, 提出了基于肤色特征的自遮挡检测算法.

3.1 肤色特征的提取

为了排除背景等噪声的干扰以提高肤色特征提取的准确性, 在提取肤色特征之前, 我们利用背景差分法和数学形态学的方法对目标图像进行预处理; 对经过预处理后的目标图像, 本文采用 Hsu 等^[20] 提出的基于 YCbCr 空间的加入亮度补偿的椭圆肤色模型以提取肤色特征. 图 3 即为肤色特征的提取过程.

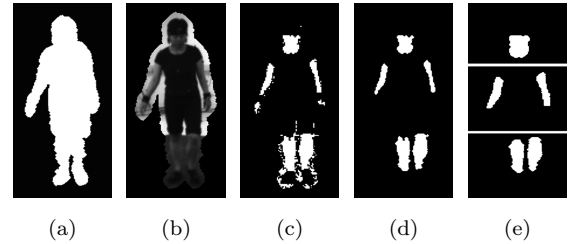


图 3 肤色特征提取过程 ((a) 分割结果; (b) 彩色图像; (c) 肤色提取; (d) 形态学处理; (e) 最终图像)

Fig. 3 The process of skin-color extraction ((a) Segmentation result; (b) Color image; (c) Extracted skin-color; (d) Morphological result; (e) Final results)

从图 3(a) 和 3(b) 可以看出, 经过预处理后的目标图像所包含的噪声已经降到很低的数量, 并且其尺寸已经变得很小, 仅容纳目标运动人体, 因此, 椭圆肤色模型的鲁棒性得到了提高, 其计算量也呈指数级减少. 图 3(c) 即为经过上述步骤处理后所提取的肤色特征, 图 3(d) 为对图 3(c) 进行数学形态学处理后的结果, 图 3(e) 为按照自顶向下的顺序, 对图 3(d) 按照人体部分进行分割后的结果.

3.2 状态分析

在本小节中, 本文利用第 3.1 节中描述的方法对所选取的实验视频进行了处理, 并对头部, 手臂和腿部的局部特征进行了分析.

根据上一小节所描述的方法, 我们可以获得完整的头部二值图像 (见图 4(a)), 以及利用异或操作消除面部肤色区域后的头部二值图像 (见图 4(b)). 在此基础上, 可以分别计算出两个图像中头部区域的重心坐标, 如图 4(a) 中的深色圆点, 图 4(b) 中的面部肤色区域内的白色圆点.

令 Q^H 表示完整的头部二值图像的重心坐标, Q'^H 表示消除面部肤色区域后的头部二值图像的重心坐标, 图 5 为 $Q'^H - Q^H$ 空间内的二维坐标投影散点图. 从图 5 可以看出, 大部分数据点均匀地分布在四个区域内, 仅有少量的数据点存在二义性问题, 见图 5 中的椭圆形区域内的数据点.



图 4 两种情况下的头部重心位置的比较

((a) 完整头部的重心; (b) 去除面部肤色特征的头部重心)
Fig. 4 Comparison of head barycenter between two cases
(a) The barycenter of whole head; (b) The barycenter of head without face)

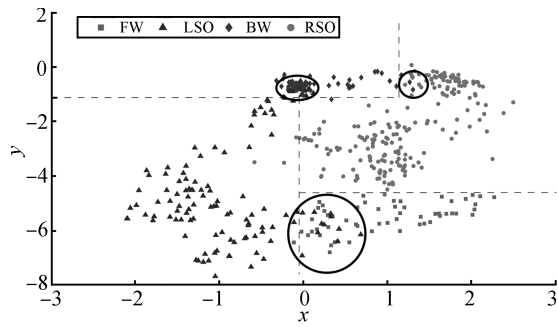


图 5 $Q'^H - Q^H$ 空间内的二维坐标投影散点图

Fig. 5 The scatter chart of projection of head features in $Q'^H - Q^H$ space

根据上一小节所描述的方法, 可以获得的手臂肤色二值图像 I_A , 并且 I_A 仅包括手臂肤色区域. 在此基础上, 我们可以从 I_A 提取两只手臂的轮廓, 从而可以获得两只手臂轮廓之间的重心距离, 记为 D^A . 显然, 如果有一只手臂被遮挡, 重心距离为 D^A 并不存在, 则有:

$$D^A = \begin{cases} \|Q^A - Q'^A\|, & C^A = 2 \\ 0, & C^A = 1 \end{cases} \quad (17)$$

其中, Q'^A 和 Q^A 分别表示两只手臂轮廓的重心二维坐标; $\|\cdot\|$ 为欧氏距离; C^A 表示检测出来的手臂轮廓个数. 图 6 中颜色较深的曲线表示两只手臂重心距离的变化曲线.

同样地, 通过对腿部肤色二值图像 I_L 提取腿部轮廓, 可以获得两个腿部轮廓的重心连线与水平轴之间的夹角, 记为 A^L . 当仅有一个腿部轮廓被检测

到时, 角度 A^L 也不存在, 则有:

$$A^L = \begin{cases} \|Q^L - Q'^L\|, & C^L = 2 \\ -1, & C^L = 1 \end{cases} \quad (18)$$

其中, $0 \leq A^L \leq 90^\circ$; Q^L 和 Q'^L 分别表示两个腿部轮廓的重心二维坐标; A^L 表示检测出来的手臂轮廓个数. 图 6 中颜色较浅的曲线表示两个腿部重心夹角的变化曲线.

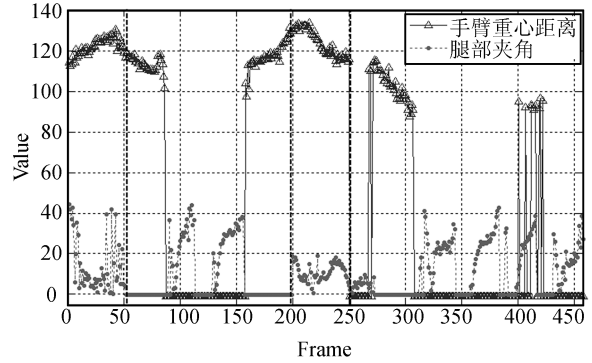


图 6 手臂和腿部肤色轮廓特征变化曲线图

Fig. 6 The variation curves of arm and leg skin features

3.3 概率计算

通过观察, 在本文所选取的实验视频中, 步行人体的手臂仅作常规的摆动, 并没有“甩臂”等现象发生. 因此, 根据步行人体运动连续性的特点, 按照四肢自遮挡情况, 步行人体运动姿态可分为四种状态: 1) 前行状态 (Front walking, FW); 2) 左侧遮挡状态 (Left self-occlusion, LSO); 3) 背行状态 (Back walking, BW); 4) 右侧遮挡状态 (Right self-occlusion, RSO). 于是, 人体运动状态转移过程可以利用上述四种状态进行描述, 见图 7. 在图 7 中, 排除了 4 种实验视频中没有出现的人体运动状态转移: 1) FW \rightarrow BW: 由前行状态直接转换到背行状态; 2) BW \rightarrow FW: 由背行状态直接转换到前行状态; 3) LSO \rightarrow RSO: 由左侧遮挡状态直接转换到右侧遮挡状态; 4) RSO \rightarrow LSO: 由右侧遮挡状态直接转换到左侧遮挡状态.

在自遮挡状态检测过程中, 若在不考虑手臂与腿部之间的运动相关性的情况下, 我们提出假设: 头部、手臂和腿部之间的自遮挡状态相互独立. 于是, 根据上述分析, 令 S 表示自遮挡状态集合, 则有:

$$S \in \{S_1, S_2, S_3, S_4\} \quad (19)$$

其中, 集合中的元素按照下标递增的顺序分别对应前文所述四种自遮挡状态.

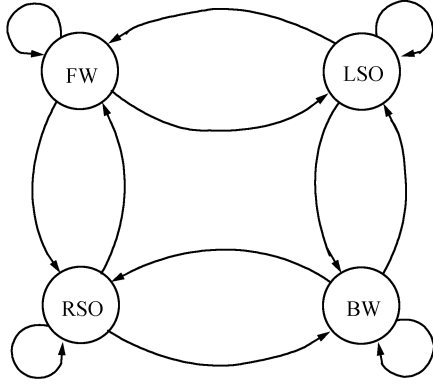


图 7 基于自遮挡状态的人体运动状态转移图

Fig. 7 Pedestrian motion states transition diagram based on self-occlusion states

令 s_t 表示自遮挡状态随时间变化的序列, o_t 为观测状态随时间变化的序列, 且 $o_t = \{o_t^H, o_t^A, o_t^L\}$, 其中, o_t^H 为头部观测状态, $o_t^H = Q'^H - Q^H$; o_t^A 为手臂的观测状态, $o_t^A = D_t^A$; o_t^L 为腿部观测状态, $o_t^L = A_t^L$.

根据上述定义和马尔科夫模型, 对于任意时刻 t , $1 \leq t \leq T$, 自遮挡状态的最大后验概率可以通过下式计算:

$$P(s_t) = \max_{s \in \mathbb{R}(s_{t-1})} P(s_t | s_{t-1}) P(o_t | s) = \max_{s \in \mathbb{R}(s_{t-1})} P(s_t | s_{t-1}) P(o_t^H | s^H) P(o_t^A | s^A) P(o_t^L | s^L) \quad (20)$$

其中, 为了简化计算, 根据图 7 所示, 转移概率 $P(s_t | s_{t-1})$ 定义为常量, 且 $P(s_t | s_{t-1}) = 1/3$; $\mathbb{R}(s_{t-1})$ 为 $t-1$ 时刻的自遮挡状态 s_{t-1} 所能直接达到的邻域; o_t^H, o_t^A, o_t^L 分别为 t 时刻所提取的头部, 手臂和腿部的观测特征; s^H, s^A 和 s^L 为待测状态的头部, 手臂和腿部特征. 根据自遮挡状态的定义, 按照不同部位所建立的均匀分布, 式中的 $P(\cdot)$ 可由这些均匀分布计算, 且每一个均匀分布的具体参数可由经验值确定. 需要注意的是, 手臂的均匀分布参数仅需考虑所有非 0 特征值, 腿部的均匀分布参数仅需考虑所有非负特征值, 如式 (21) 所示.

$$\begin{aligned} U_A^i(s) &= \delta(D_A > 0) \times U_A^i(s) \\ U_L^i(s) &= \delta(D_L \geq 0) \times U_L^i(s) \end{aligned} \quad (21)$$

其中, $\delta(\cdot)$ 为 Delta 函数, 若参数表达式为真时, $\delta(\cdot) = 1$; 否则为 0. $U_A^i(s)$ 和 $U_L^i(s)$ 分别表示第 i 个自遮挡状态下的均匀分布.

从图 5 可以看出, 在 $Q'^H - Q^H$ 空间内, 头部特征在不同的自遮挡状态下满足二维均匀分布, 其中, 在其他三个状态下满足简单的二维均匀分布, 而在

RSO 状态下则可利用两个二维均匀分布进行表示, 因此, 头部的概率计算可由式 (22) 进行表示:

$$U_H(s) = \begin{cases} U_H^S(\hat{x}, \hat{y}), & S \neq \text{RSO} \\ \gamma \hat{U}_H(\hat{x}, \hat{y}) + (1 - \gamma) \tilde{U}_H(\hat{x}, \hat{y}), & \text{其他} \end{cases} \quad (22)$$

其中, $\gamma = \delta(\hat{x} < x_{Th}), (\hat{x}, \hat{y})$ 为 $Q'^H - Q^H$ 空间内的坐标, x_{Th} 为 RSO 状态下的 \hat{x} 坐标阈值, 通过该阈值, 可以将 RSO 状态的概率用两个二维均匀分布表示.

于是, 当前时刻 t 的自遮挡状态 s_t 可由下式确定:

$$s_t = \arg \max_{s \in \mathbb{R}(s_{t-1})} P(s_t | s) \quad (23)$$

根据式 (23), 式 (3) 中的遮挡因子 λ_t 和式 (13) 中的遮挡因子 λ_t^S 可由式 (24) 获得:

$$\begin{aligned} \lambda_t &= \begin{cases} 1, & s_t \in \{S_{FW}, S_{BW}\} \\ 0, & s_t \in \{S_{RSO}, S_{LSO}\} \end{cases} \\ \lambda_t^S &= \begin{cases} 1, & s_t = S_{RSO} \text{ 且 } \lambda_t = 0 \\ 0, & s_t = S_{LSO} \text{ 且 } \lambda_t = 0 \end{cases} \end{aligned} \quad (24)$$

4 观测似然模型

在粒子滤波理论框架中, 观测似然模型被用来建立二维图像中所提取的特征与人体外观模型之间的匹配关系. 在本文中, 我们利用基于区域的加权颜色直方图^[21] 和 Charmfer 距离变换来建立联合观测似然模型.

在任意时刻 t , 对于任意肢体 i 的联合观测似然模型可通过以下步骤计算:

步骤 1. 利用 3D 锥形圆柱体模型, 将运用第 2.2.1 节和 2.2.2 节中描述的算法所生成的粒子投影到二维图像平面上, 从而获得肢体 i 的二维纸版形轮廓 $\tilde{C}_{i,t}^j$, $\tilde{C}_{i,t}^j$ 为二维坐标点集, 且 $\tilde{C}_{i,t}^j = \{\tilde{x}_{i,t}^{j,u}, \tilde{y}_{i,t}^{j,u}\}_{u=1}^K$, 上标 j 表示第 j 个粒子, K 为 $\tilde{C}_{i,t}^j$ 所包含的二维坐标点个数.

步骤 2. 利用文献 [22] 中的加权颜色直方图计算式, 计算 $\tilde{C}_{i,t}^j$ 所形成的区域的加权颜色直方图 $H_{i,t}^j$, $H_{i,t}^j$ 与颜色直方图模版 $H_{i,t}^T$ 之间的相似度则利用 Bhattacharyya 距离^[21] 计算, 见下式:

$$d_{i,j} = \sqrt{1 - \sum_{m=1}^N \sqrt{H_{i,t}^{j,m} \times H_{i,t}^{T,m}}} \quad (25)$$

其中, N 为组成颜色直方图的条柱数目, 且 $N = 512$. 颜色直方图模版 $H_{i,t}^T$ 的更新公式如下:

$$H_{i,t}^T = (1 - \alpha) \times \tilde{H}_i + \alpha \times H_{i,t-1}^T \quad (26)$$

其中, \tilde{H}_i 为当前时刻 t 、任意肢体 i 最优粒子的所形成的颜色直方图, H_{t-1}^T 为 $t-1$ 时刻的模版, 更新权重 $\alpha = 0.95$.

步骤 3. 利用第 3.1 节中所描述的预处理过程, 提取出人体区域, 并存放到一副新的二值图像中, 该图像与原始视频图像具有相同尺寸, 记为 \tilde{I} .

步骤 4. 对 \tilde{I} 进行 Charmfer 距离变换, 变换后的 DT 图记为 \tilde{I}_C , 则 $\tilde{C}_{i,t}^j$ 的边缘匹配系数 ϑ 可用下式计算:

$$\vartheta(\tilde{C}_{i,t}^j) = \sum_{u=1}^C \tilde{I}_C(\tilde{x}_{i,t}^{j,u}, \tilde{y}_{i,t}^{j,u}) \quad (27)$$

其中, C 为 $\tilde{C}_{i,t}^j$ 所包含的像素点个数.

为了计算最大后验概率, 观测似然模型 $P_{\text{Likelihood}}$ 可由下式表示:

$$P_{\text{Likelihood}} = \frac{\exp(-d_{i,j}^2/\sigma^2) \times \exp(-\vartheta(\tilde{C}_{i,t}^j))}{\sigma\sqrt{2\pi}} \quad (28)$$

5 肢体运动相关系数

根据文献 [4, 17], 在步行人体运动过程中, 左侧肢体和右侧肢体之间存在着一定的线性关系, 即通过肢体运动关系系数, 在左侧肢体存在自遮挡现象时, 左侧肢体可由右侧肢体的 3D 位置进行线性预测. 并且, 如前文所述, 步行人体在运动过程中, 左侧肢体被遮挡和右侧肢体被遮挡的情况交替出现. 因此, 为了解决上述两种自遮挡现象, 我们利用回归模型分别对左侧肢体被遮挡时的肢体相关系数 B^L 和右侧肢体被遮挡时的肢体相关系数 B^R .

作为稳健回归 (Robust regression) 的经典算法之一, M-估计 (M-estimation) 算法能够对异常值起到抑制作用, 并可以极小化异常值对正常值的影响, 从而保证了回归方程的精度. 在此基础上, 本文提出了基于 M-估计的在线训练方法.

5.1 回归模型

令人体左侧肢体的角度向量矩阵为 L , 相应地, 右侧肢体的角度向量矩阵为 R , 其中, L 和 R 为 $n \times M$ 阶矩阵, $M = 8$, n 为样本个数, 且 $n \geq M + 1$; 且 L 和 R 采用列向量的形式, 则 $L = [l_1, l_2, \dots, l_M]$, $R = [r_1, r_2, \dots, r_M]$. 定义扩展矩阵 \tilde{R} , $\tilde{R} = [r_0 \ R]$, r_0 为单位列向量. 于是, L 和 \tilde{R} 的线性关系可由如下方程表示:

$$L = \tilde{R}B^L + \varepsilon \quad (29)$$

其中, B^L 即为左侧肢体与右侧肢体之间的肢体运动相关系数矩阵, 为 $M \times (M + 1)$ 维矩阵, 且利用 B^L , 左侧肢体 L 可由右侧肢体 R 进行表示; ε 为 M

$\times n$ 的误差矩阵. 因此, B^L 的训练就是利用已有的样本数据对式 (29) 的求解过程.

根据 M-估计算法, 将式 (29) 展开, 回归模型可由下式表示:

$$l_i = \sum_{k=0}^M \tilde{r}_k b_{i,k}^L + \varepsilon_i = \tilde{R}b_i^L + \varepsilon_i \quad (30)$$

其中, l_i 为第 i 个变量的观测值列向量; b_i^L 为系数矩阵 B^L 的第 i 列; \tilde{r}_k 为矩阵 \tilde{R} 的第 k 行; ε_i 为 ε 的第 i 列, 且 $\varepsilon_i = l_i - \sum_{k=0}^M \tilde{r}_k b_{i,k}^L$.

在 M-估计算法中, 迭代重加权最小二乘法 (Iteratively reweighted least squares, IRLS) 被用来求解式 (30). 定义 $\phi(\varepsilon_i)$ 为误差函数, $\pi(\varepsilon_i)$ 为样本权重函数, 且 $\pi_i = \pi(\varepsilon_i)$, 并且, $\phi(\varepsilon_i)$ 和 $\pi(\varepsilon_i)$ 可利用 Huber 算法进行计算.

于是, 利用 M-估计算法求解回归模型的迭代流程如下:

步骤 1. 利用最小二乘法, 初始化 $b_i^{L(0)}$.

步骤 2. 在第 t 次迭代中, 利用 Huber 算法计算 $t-1$ 次迭代结果的误差 $\varepsilon_i^{(t-1)}$ 和样本权重 $\pi_i^{(t-1)}$.

步骤 3. 利用 IRLS 算法, 计算 $b_i^{L(t)}$:

$$b_i^{L(t)} = (R'\pi^{(t-1)}R)^{-1}R'\pi^{(t-1)}l_i \quad (31)$$

其中, $\pi^{(t-1)} = \text{diag}\{\pi_i^{(t-1)}\}$ 为样本权重矩阵.

步骤 4. 重复步骤 2 和步骤 3, 直至满足迭代终止条件或达到最大迭代次数.

在本文中, 迭代次数 $MaxLoop$ 设置为 20, 定义总平方误差函数如下式:

$$E_{(t)}^2 = \frac{\sum_{i=1}^n \pi(\varepsilon_i)\varepsilon_i^2}{\sum_{i=1}^n \pi(\varepsilon_i)} \quad (32)$$

根据式 (32), 迭代终止条件可被定义为两次迭代之间的总平方误差之间无明显变化.

$$E_{(t)}^2 - E_{(t-1)}^2 \leq E_{th} \quad (33)$$

其中, E_{th} 为常量, $E_{th} = 0.0001$.

显然, 经过上述步骤所求得的 B^L 仅为左侧肢体被遮挡时的肢体相关系数, 右侧肢体被遮挡时的肢体相关系数 B^R , 同样可通过上述步骤进行训练获得.

5.2 在线训练

如前文所述, 在实际跟踪过程中, 初始帧数要小于训练相关系数矩阵 B 所需的样本数据的数目. 本文利用上一小节中的回归模型, 将每一帧的跟踪结果作为样本数据, 采用在线训练的方法训练相关系数矩阵 B .

根据上一小节中的回归模型的定义, 至少需要 $M + 1$ 个样本数据才可通过计算获得 B . 初始样本数据可通过手工标定或者利用无遮挡状态下的跟踪结果获得. 在本文中, 我们利用无遮挡状态下的跟踪结果作为初始样本数据. 为了与视频图像时间顺序相一致, 对于任意时刻 t , 相关系数矩阵记为 B_t , 并且当 $t \leq M$ 时, B_t 并不存在. 在此基础上, 基于在线训练方法的具体步骤如下:

步骤 1. 若初始样本数量 $n \leq M$, 则利用无遮挡跟踪模型对跟踪视频进行跟踪, 并将跟踪结果进行分类并加入到样本集合 L 和 R 中, L 和 R 为 $n \times M$ 阶矩阵, 直至 $n = M + 1$, 且 $t = n$, 利用 M-估计算法, 初始化肢体运动相关系数矩阵 B^L 和 B^R .

步骤 2. 定义矩阵 L 和 R 为固定维数的 $MaxLen \times M$ 阶矩阵, 其中, $MaxLen$ 为样本截断长度, 其值依赖于 M 值以及无遮挡状态下的视频帧数. 若 $t < MaxLen$, 将当前时刻 t 的跟踪结果加入到样本集合 L 和 R 中, 在此基础上, 利用回归模型计算相关系数矩阵 B^L 和 B^R , 重复本步骤, 直至 $t \geq MaxLen$.

步骤 3. 若 $t \geq MaxLen$, 则截断样本数据, 保留最近的 $MaxLen$ 个样本数据. 在此基础上, 利用回归模型计算相关系数矩阵 B^L 和 B^R .

步骤 4. 对于 $t + 1$ 时刻, 若 $t + 1 \leq T$, 利用自遮挡概率模型判断步行人体运动状态处于何种自遮挡状态, 若处于无遮挡状态, 则跳至步骤 5, 若处于无遮挡状态, 则跳至步骤 6. 若 $t + 1 > T$, 则跳至步骤 7.

步骤 5. 利用无遮挡跟踪模型对 $t + 1$ 时刻的步行人体运动进行跟踪, 并将跟踪结果加入到样本集合 S 中, 转至步骤 3.

步骤 6. 利用自遮挡跟踪模型和遮挡因子 λ_t^S , 通过 B^L 和 B^R 对 $t + 1$ 时刻的步行人体运动进行跟踪, 并将跟踪结果加入到样本集合 S 中, 转至步骤 2.

步骤 7. 结束.

由于人体运动的随机性和实时性, 过多的样本数据并不能对提高回归精度有所帮助, 因此, 在步骤 2 中, 我们对样本数据进行截断, 并且截断样本数据还可以提高计算速度, 减少内存消耗. 显然, 对于任意时刻 t , 矩阵 L 和 R 中的样本数据总为时刻 $t - MaxLen + 1$ 到时刻 t 的跟踪结果. 利用上述训练过程, 我们可以分别对左侧遮挡状态和右侧遮挡状态下的肢体运动相关系数进行训练.

6 实验分析

算法的编程环境为 Visual Studio.Net 2003 和 OpenCV 1.0, 全部实验在一台配置为双核 CPU

1.8 GHz, 1G DDR 内存的计算机上完成. 算法的实验视频来源于美国布朗大学的 Leonid 等利用 VICON 系统所采集的 HumanEva 人体运动实验数据集^[2].

本文选取 HumanEva I 数据集中 S1 和 S2 子集中的 Walking 视频作为实验视频, 并分别从 S1 中选取 458 帧连续视频图像, 从 S2 中选取 420 帧连续视频图像. 在两组图像中, 步行目标人体的行走轨迹均为一个完整的圆周, 并且 S1 视频中的步行人体的四肢均有裸露的皮肤, 而 S2 视频中步行人体身穿长裤, 仅有上肢有裸露的皮肤.

前文所描述的所有实验分析及所分析数据均来自所选取 S1 中的 458 帧视频图像及其真实数据. 并且, 这两组数据中的第一帧数据为已知的初始帧数据.

6.1 自遮挡检测误差

自遮挡状态检测是本文算法的核心, 其检测误差将直接影响基于 M-估计的训练误差和跟踪误差.

为了计算自遮挡检测误差, 我们利用人工标定的方法对所选取的 S1 中的 458 帧和所选取的 S2 中的 420 帧连续视频图像进行遮挡状态标定, 标定方法为: 以某一个肢体未被遮挡肤色面积 D_O 大于等于该肢体全部肤色面积 D_A 的 $2/3$ 为标准, 即 $D_O \geq D_A \times 2/3$, n 手工标注四种自遮挡状态. 在所选取的两组实验视频中, 各种遮挡状态出现的先后顺序为 $FW \rightarrow LSO \rightarrow BW \rightarrow RSO$. 面部肤色特征图像标准化后的高度为 $h = 150$, $w = 200$. 我们利用自遮挡状态检测模型对所选 S1 和 S2 的实验视频进行检测, S1 中实验视频的各自遮挡状态检测结果见表 1. S2 中实验视频的各自遮挡状态检测结果见表 2.

从表 1 和表 2 可以看出, S1 实验视频的检测准确率达到 95.6%, S2 实验视频的检测准确率达到 91.7%. 产生错误的主要原因是在 S1 和 S2 的实验视频图像中, 由于光照而导致的阴影和图像分辨率较低等原因 (这些情况与实际应用中所拍摄的图像较类似), 椭圆肤色模型无法准确地对前行状态和背行状态下的肤色进行检测, 并且这种情况会几帧连续图像中持续出现, 这种情况主要集中在前行状态和背行状态中, 此外, 由于 S2 实验视频中的目标人体身穿长裤, 腿部并无裸露的皮肤, 在这种情况下, 我们在提取颜色特征之前就对人体进行部位分割, 以减少分割误差, 尽管如此, 在进行颜色分割时, 仍然极易容易产生粘连的现象, 从而造成了不必要的误差, 如图 8 所示. 产生错误的另外一个原因是计算误差, 主要集中在左侧遮挡状态和右侧遮挡状态的检测过程中, 算法的计算误差约等于这两个状态下

表 1 S1 实验视频的自遮挡检测结果

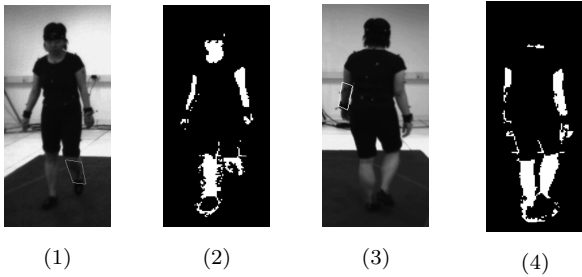
Table 1 The results of S1 task video detection

	总帧数	错误帧数	错误率 (%)
前行状态	55	2	3.6
右侧遮挡状态	145	6	4.1
左侧遮挡状态	53	5	9.4
背行状态	205	7	3.4
总计	458	20	4.4

表 2 S2 实验视频的自遮挡检测结果

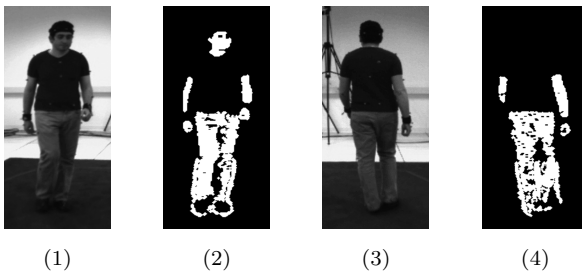
Table 2 The results of S2 task video detection

	总帧数	错误帧数	错误率 (%)
前行状态	21	2	10
右侧遮挡状态	160	10	6.25
左侧遮挡状态	40	8	20
背行状态	199	15	7.5
总计	420	35	8.3



(a) S1 实验视频 ((1) 腿部阴影; (2) 肤色检测结果; (3) 手臂阴影; (4) 肤色检测结果)

(a) S1 task video ((1) The leg shadow; (2) Corresponding skin-color detecting result of (1); (3) The arm shadow; (4) Corresponding skin-color detecting result of (3))



(b) S2 实验视频 ((1) 腿部粘连; (2) 肤色检测结果; (3) 腿部粘连; (4) 肤色检测结果)

(b) S2 task video ((1) The leg conglutination; (2) Corresponding skin-color detecting result of (1); (3) The leg conglutination; (4) Corresponding skin-color detecting result of (3))

图 8 光照不均匀条件下的颜色特征检测结果

Fig. 8 The skin-color detection results using ellipse skin model on poor illumination condition

的错误率.

由于自遮挡状态检测对 HTMPF 算法的重要性, 因此, 本文借助于手工标定的方法纠正这些错误.

6.2 跟踪误差

由于算法利用每一帧的跟踪结果来在线训练肢体运动相关系数, 因此, 基于 M-估计的训练误差和跟踪误差之间相互影响, 跟踪误差对肢体运动相关系数产生影响, 肢体运动相关系数反作用于跟踪误差, 并且, 这两种误差最终的表现形式为跟踪误差, 所以本文将这两种误差以跟踪误差的形式进行统一分析.

在本文所选视频中, 前行状态为初始状态, 并且, 在 S1 和 S2 的实验视频中, 前行状态分别有连续 55 帧图像和 20 帧图像. 考虑到样本阶段长度 $MaxLen$ 必须满足 $MaxLen \geq M + 1$, 定义样本截断长度 $MaxLen = 20$, 在此基础上, 肢体运动相关系数可由第 5.2 节中的方法进行训练.

为了描述跟踪误差, 在每一次跟踪过程中, 根据离散关节图模型算法, 算法对每一个关节点进行粒子采样, 每一个关节点的粒子数目定义为 $N_p = 50$.

我们利用人体姿态总体误差 (Full pose error, FPE)^[22], 对本文所提出的基于混合跟踪模型的粒子滤波 (HTMPF) 算法的跟踪结果与真实值之间的误差进行分析. 在文献 [22] 中, 在每一次跟踪过程中, FPE 可以通过计算每一个关节点和肢体的每一个粒子和真实值之间的欧氏距离的平均值获得. 在本文中, 我们直接利用跟踪结果计算 FPE, 则在任意时刻, FPE 的计算式如下:

$$FPE(S_t, V_t) = \frac{\sum_{i=1}^H \|S_t^i - V_t^i\|}{H} \quad (34)$$

其中, S_t 为人体 3D 姿态跟踪结果, S_t^i 表示第 i 个关节点的跟踪结果, V_t 为人体 3D 姿态的真实值, V_t^i 为第 i 个关节点的真实值. H 为重构人体运动姿态的关节点数目, 根据人体 3D 骨骼模型, $H = 15$. 为了与真实数据相对应, FPE 的单位为毫米 (mm).

根据式 (34), 图 9 为利用 HTMPF 算法对 S1 和 S2 的实验视频的跟踪结果的二维投影效果图, 图 10 为算法跟踪误差 FPE 随时间变化曲线.

6.3 实验对比

本文利用 HTMPF 算法, RBPF^[4] 算法, PF^[8] 算法, LLPT^[22] 算法分别对前文所选取的 S1 实验视频图像序列进行跟踪. 为了方便比较, 四种算法的实验所采用相同的观测似然模型, 包括颜色直方图和 Charmfer 距离, 相同的粒子产生机制以及相同的

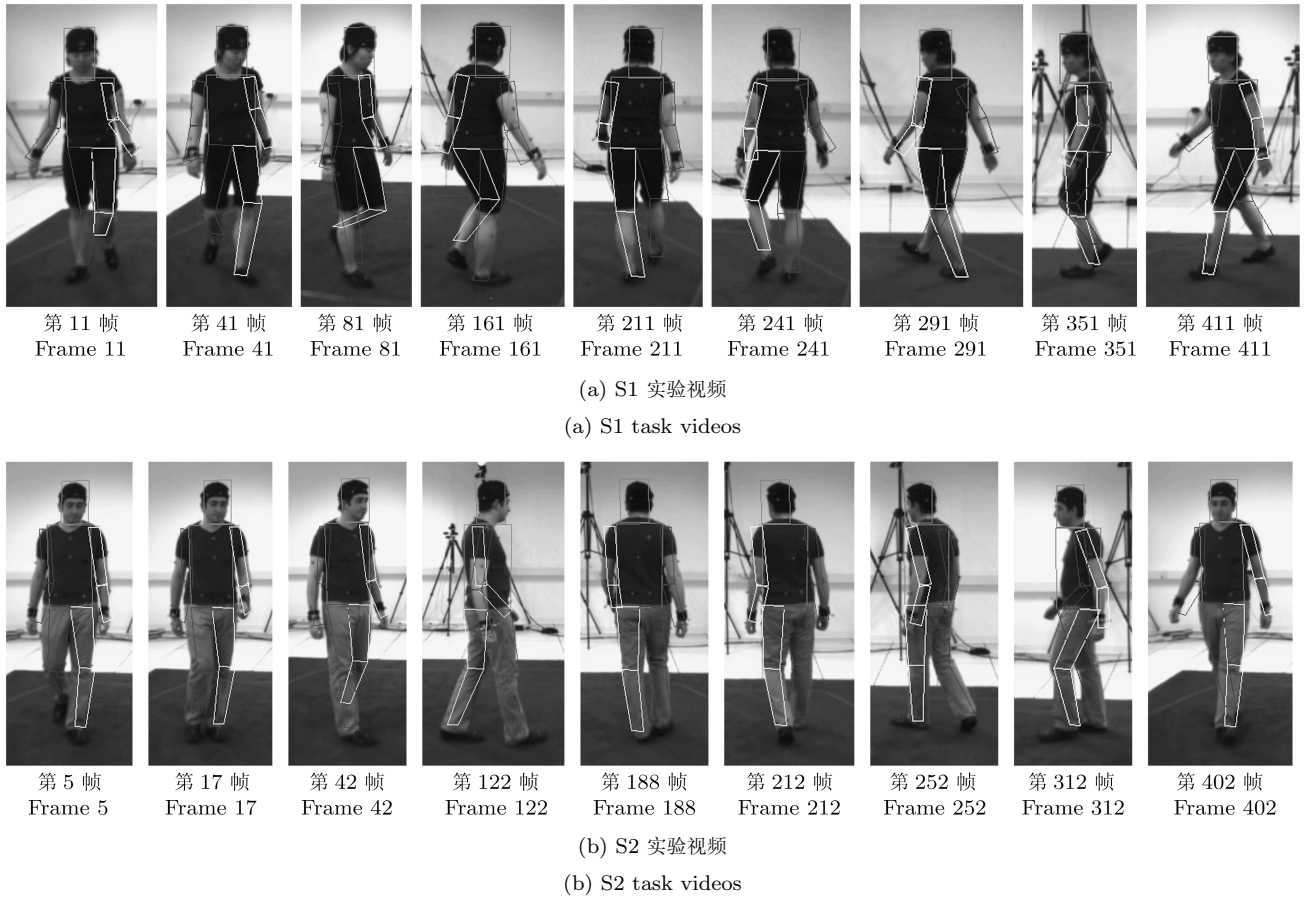


图 9 HTMPF 算法跟踪结果二维投影效果图

Fig. 9 The projection of HTMPF tracking results in image plane

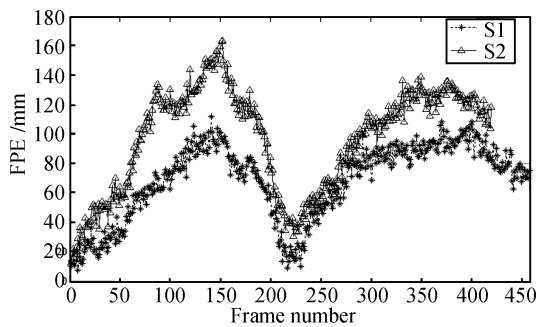


图 10 HTMPF 算法 FPE 随时间变化曲线

Fig. 10 The variation curves of FPE for HTMPF with time

运动转移模型. 对于 RBPF, 我们按照文献 [4] 中所描述的方法训练肢体相关性系数. 在每一次跟踪过程中, 对于 PF 算法和 RBPF 算法, 产生 1200 个粒子用来进行跟踪; 对于 LLPT 算法和 HTMPF 算法, 为每一个跟踪器产生 200 个粒子, 其中, 每一个跟踪器对应构成人体的一个肢体部位, 总计 6 个肢

体部位, 则需产生 $6 \times 200 = 1200$ 个粒子. 图 11 为四种算法跟踪误差 FPE 随时间变化对比曲线. 表 3 为 4 种算法的误差均值, 单帧跟踪时间的比较. 其中, 单帧跟踪时间为连续跟踪 50 帧所需时间的均值, 而误差均值可由下式计算:

$$Mean_{FPE} = \sum_{i=1}^T FPE(S_i, V_i) T \quad (35)$$

其中, T 为实验视频帧数, $T = 458$.

从表 3 可以看出, HTMPF 算法单帧跟踪时间比 LLPT 算法和 PF 算法优越, 而与 RBPF 算法相比, HTMPF 算法的时间消耗要多. 但是, 与 LLPT 算法和 PF 算法相比, HTMPF 算法的误差均值要小得多, 跟踪精度得到了提高. 与 RBPF 算法相比, HTMPF 算法尽管在跟踪精度和跟踪所需的时间消耗方面并没有得到改善, 甚至单帧跟踪时间要超过 RBPF 算法很多; 然而, 由于 HTMPF 算法采用在线训练的方法以获得肢体运动相关系数, 而 RBPF 则必须利用真实数据训练并获得肢体运动相关系数,

然而在实际的跟踪过程中, 真实数据是无法获得的. 因此, 与 RBPF 算法相比, HTMPF 算法具有更高实用性.

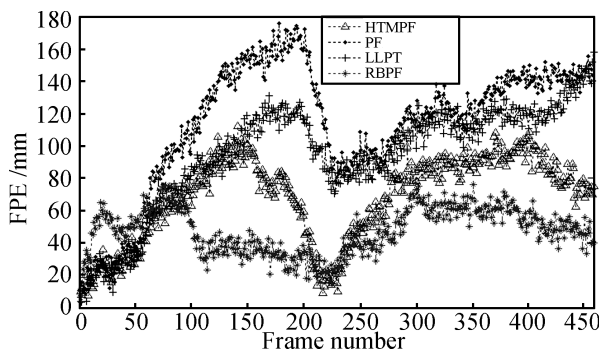


图 11 四种跟踪算法的 FPE 随时间变化对比曲线

Fig. 11 Variation curves of FPE compared among four algorithms with time

表 3 四种算法的性能指标比较

Table 3 The performance compared among four algorithms

算法名称	跟踪器 (个)	每秒跟踪帧数 (fps)	误差均值 (mm)
LLPT	6	2	94.4218
PF	1	1	113.4987
RBPF	1	7	47.4866
HTMPF	10	4	68.0583

7 结论

根据室内步行人体运动的特点, 本文利用肤色特征建立自遮挡检测模型, 并根据该模型将室内步行人体运动对应四种自遮挡状态进行划分. 在此基础上, 提出了基于混合跟踪模型的粒子滤波算法 (HTMPF). 按照不同自遮挡状态选择不同的粒子滤波算法对室内步行人体运动进行估计. 在前行状态和背行状态下, 室内步行人体运动可利用基于离散关节图模型的序列蒙特卡洛粒子滤波算法进行估计; 在左侧遮挡状态和右侧遮挡状态下, 则可利用 RBPF 算法进行估计. 根据实验对比分析, 尽管 HTMPF 算法单帧跟踪时间消耗要多于 LLPT 算法和 RBPF 算法, 但是, 与 PF 算法, LLPT 算法和 RBPF 算法相比, HTMPF 算法的跟踪精度得到了提高. 然而, 在实际跟踪过程中, 由于无法提前获得真实数据, 肢体运动相关系数无法提前获得. 因此, 利用 M-估计算法, HTMPF 算法利用每一帧的跟踪结果在线训练肢体运动相关系数, 从而保证了 HTMPF 算法的实用性.

然而, 作为 HTMPF 算法的核心, 自遮挡检测

模型所要检测的视频图像必须在较好的光照条件下拍摄, 因此, HTMPF 算法带有一定的局限性, 即室内步行人体运动, 并且无法跟踪多人运动. 此外, 由于在实际应用中, 经常会出现以下情况: 1) 人体运动过程中皮肤裸露面积较少, 有的视频中甚至仅有面部皮肤和手部皮肤能够检测到; 2) 由于人体的动作的随意性, 可能会出现甩臂等现象, 从而导致状态是别的二义性问题. 针对肤色裸露的问题, 根据 S2 实验视频的实验结果和分析, 若目标人体没有裸露的皮肤区域, 可通过在四肢衣服上放置标志物, 或者在四肢位置的衣物上标志不同颜色加以解决, 或者四肢的衣物的颜色互不相同, 并且最好与背景颜色不相同, 此时, 可利用其他颜色特征或标志物代替人体肤色模型. 对于检测过程中可能出现的二义性问题, 可结合人体动作识别算法获得解决, 如基于上下文的人体动作识别等算法, 另外, 还可考虑将自遮挡检测模型结合人体 3D 运动模型对人体动作进行运动约束, 从而对人体剧烈运动情况进行剪枝处理, 以解决二义性问题. 在本文中, 自遮挡检测模型而产生的二义性问题情况非常少, 因此, 可以采用手工标定的方法进行处理.

经过上述分析, HTMPF 算法可以应用于医疗康复训练的运动估计, HumanEva 实验数据集; 并且, 通过与人体动作识别等算法的结合, HTMPF 算法还可以得到进一步的应用, 这将是我们的下一步的工作目标.

References

- 1 Moeslund T B, Hilton A, Kruger V. A survey of advances in vision based human motion capture and analysis. *Computer Vision and Image Understanding*, 2006, **104**(2): 90–126
- 2 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, USA, 2006
- 3 Qu W, Schonfeld D. Real-time decentralized articulated motion analysis and object tracking from videos. *IEEE Transactions on Image Processing*, 2007, **16**(8): 2129–2138
- 4 Xu X Y, Li B X. Learning motion correlation for tracking articulated human body with a rao-blackwellised particle filter. In: *Proceedings of the 11th International Conference on Computer Vision*. Rio de Janeiro, Brazil: IEEE, 2007. 1–8
- 5 Balan A, Black M J. An adaptive appearance model approach for model based articulated object tracking. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. New York, USA: IEEE, 2006. 758–765
- 6 Sminchisescu C, Kanaujia A, Metaxas D. BM³E: discriminative density propagation for visual tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, **29**(11): 2030–2044
- 7 Urtasun R, Fleet D J, Fua P. 3D people tracking with Gaussian process dynamical models. In: *Proceedings of IEEE*

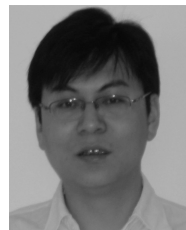
- Computer Society Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE, 2006. 238–245
- 8 Isard M, Blake A. Condensation-conditional density propagation for visual tracking. *International Journal of Computer Vision*, 1998, **29**(1): 5–28
- 9 Deutscher J, Blake A, Reid I. Articulated body motion capture by annealed particle filtering. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Hilton Head, South Carolina: IEEE, 2000. 126–133
- 10 Chen Rui, Liu Guo-Yi, Zhao Guo-Ying, Zhang Jun, Li Hua. 3D human motion tracking based on sequential Monte Carlo method. *Journal of Computer-Aided Design and Computer Graphics*, 2005, **17**(1): 85–92
(陈睿, 刘国翌, 赵国英, 张俊, 李华. 基于序列蒙特卡罗方法的3D人体运动跟踪. 计算机辅助设计与图形学学报, 2005, **17**(1): 85–92)
- 11 Sidenbladh H, Black M J, Sigal L. Implicit probabilistic models of human motion for synthesis and tracking. In: Proceedings of the 7th European Conference on Computer Vision. Copenhagen, Denmark: Springer, 2002. 784–800
- 12 Urtasun R, Fleet D J, Fua P. Monocular 3D tracking of the golf swing. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE, 2005. 932–938
- 13 Sidenbladh H, Black M J, Fleet D J. Stochastic tracking of 3D human figures using 2D image motion. In: Proceedings of the 6th European Conference on Computer Vision. Berlin, Germany: Springer, 2000. 702–718
- 14 Urtasun R, Fleet D J, Hertzmann A, Fua P. Priors for people tracking from small training sets. In: Proceedings of the 10th International Conference on Computer Vision. Beijing, China: IEEE, 2005. 403–410
- 15 Agarwal A, Triggs B. Monocular human motion capture with a mixture of regressors. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE, 2005. 72–86
- 16 Agarwal A, Triggs B. Recovering 3D human pose from monocular images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(1): 44–58
- 17 Xu X Y, Li B X. Adaptive Rao-Blackwellized particle filter and its evaluation for tracking in surveillance. *IEEE Transactions on Image Processing*, 2007, **16**(3): 838–849
- 18 Doucet A, de Freitas N, Murphy K, Russell S. Rao-blackwellised particle filtering for dynamic Bayesian networks. In: Proceedings of the 16th Conference in Uncertainty in Artificial Intelligence. Stanford, USA: Morgan Kaufmann, 2000. 176–183
- 19 Wu Y, Hua G, Yu T. Tracking articulated body by dynamic Markov network. In: Proceedings of the 9th Conference on Computer Vision. Nice, France: IEEE, 2003. 1094–1101
- 20 Hsu R L, Abdel-Mottaleb M, Jain A K. Face detection in color images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(5): 696–706
- 21 Zhuang Yan, Zhan Hong-Bin, Wang Wei, Wang Ke. Weighted color histogram based particle filter for visual target tracking. *Control and Decision*, 2006, **21**(8): 868–873
(庄严, 战洪斌, 王伟, 王珂. 基于加权颜色直方图和粒子滤波的彩色物体跟踪. 控制与决策, 2006, **21**(8): 868–873)
- 22 Sigal L, Bhatia S, Roth S, Black M J, Isard M. Tracking loose-limbed people. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE, 2004. 421–428



于雪松 博士. 分别于2004年和2009年在哈尔滨工业大学获硕士学位和博士学位. 主要研究方向为计算机视觉和人体运动分析. 本文通信作者.
E-mail: yyz001@hit.edu.cn
(YU Xue-Song Ph.D.. He received his master and Ph.D. degrees from Harbin Institute of Technology in 2004 and 2009, respectively. His research interest covers computer vision and human motion analysis. Corresponding author of this paper.)



赵巍 博士, 讲师. 分别于2000年和2006年在哈尔滨工业大学获硕士学位和博士学位. 主要研究方向为模式识别, 图像处理和计算机视觉.
E-mail: zhaowei@hit.edu.cn
(ZHAO Wei Ph.D., lecturer. She received her master and Ph.D. degrees from Harbin Institute of Technology in 2000 and 2006, respectively. Her research interest covers pattern recognition, image process, and computer vision.)



刘鹏 博士, 讲师. 分别于2003年和2007年在哈尔滨工业大学获硕士学位和博士学位. 主要研究方向为模式识别, 图像处理和机器视觉.
E-mail: pengliu@hit.edu.cn
(LIU Peng Ph.D., lecturer. He received his master and Ph.D. degrees from Harbin Institute of Technology in 2003 and 2007, respectively. His research interest covers pattern recognition, image process, and computer vision.)



唐降龙 博士, 教授. 分别于1986年和1995年在哈尔滨工业大学获硕士学位和博士学位. 主要研究方向为字符识别, 生物特征识别, 图像处理和模式识别. E-mail: txl60@public.hr.hl.cn
(TANG Xiang-Long Ph.D., professor. He received his master and Ph.D. degrees from Harbin Institute of Technology in 1986 and 1995, respectively. His research interest covers optical character recognition, human biometric identification, image process, and pattern recognition.)