doi:10.3772/j.issn.1002-0470.2021.11.011

视觉遮挡下的人体姿态鲁棒估计①

贾晓凌② 张文安③ 杨旭升

(浙江工业大学信息工程学院 浙江省嵌入式系统联合重点实验室 杭州 310023)

摘 要 针对人体姿态估计中的视觉遮挡问题,提出了一种带有鲁棒卡尔曼滤波(RKF) 的人体姿态估计方法。首先,采用随机森林方法(RFM)从深度图像中识别出人体各部 件,并计算出人体各关节点在相机坐标系下的 3D 位置;其次,考虑到视觉自遮挡或遮挡 引起的人体部件误识别,设计了一种鲁棒的卡尔曼滤波器,利用假设检验的方法对视觉遮 挡造成人体姿态信息中包含的复杂噪声进行识别和分类,以提高人体姿态估计对视觉遮 挡的鲁棒性。最后,通过仿真结果以及多个人体关节点的 3D 位置估计实验表明,所提方 法可有效提高人体姿态估计的精度和鲁棒性。

关键词 视觉遮挡; 人体姿态估计; 鲁棒卡尔曼滤波(RKF)

0 引言

近年来,在人机协作、体育竞技、体感娱乐以及 医疗康复等领域,人体姿态估计扮演着越来越重要 的角色^[1-3]。特别地,随着 3D 视觉和人工智能等技 术的发展,3D 人体姿态估计的研究引起了国内外学 者的广泛关注^[4-5]。

在视觉人体姿态估计领域,较为成熟的方法主要有两种,一是利用多相机系统捕捉可穿戴式的反光标记点以实现 3D 人体姿态信息的捕捉,如 Opti-Track、Vicon等人体姿态捕捉系统。该方法利用标记点避免了复杂的人体关节点的识别,具有较高的人体姿态估计的精度,但其对人工标记点的依赖,使得该方法的使用范围受限^[6-7]。二是利用深度传感器信息实现 3D 人体姿态估计。常用的深度传感器有 Microsoft 的 Kinect 系列、ASUS 的 Xtion 系列等传感器。凭借这类传感器获取深度图像的优势,结合随机森林等方法便捷地实现人体关节点的识别,继而获得 3D 人体姿态的估计^[8-9]。

文献[10]提出采用随机森林方法(random forest method, RFM)来实现在单个深度图像中的人体 姿态估计,并由此衍生出一系列在深度图像中人体 姿态估计方法的研究,如近几年提出的 Hough forests^[11]、Random ferns^[12]、Random tree walks^[13]等系 列方法。尽管这类方法已将深度图像领域中的人体 姿态估计做到了很高的水准,但难以克服视觉遮挡 带来的估计精度下降的问题。为了处理视觉遮挡问 题,文献[14]提出了一种基于单深度图像的视点不 变模型,能够做到在自遮挡情况下选择性地预测部 分人体姿态。文献[15]提出了一个语义遮挡模型, 并结合回归森林方法可做到人体被物体部分遮挡时 的姿态估计。此外,还可利用多深度传感器建立人 体姿态的感知系统来应对由于视角问题造成的遮 挡^[16-17]。然而,深度视觉存在红外线(infrared radiation,IR)相互干扰、数据不连续等问题^[18],这使得多 深度传感器系统不仅需要考虑高计算复杂度,还需 充分考虑传感器间干扰问题。另一方面,卡尔曼滤 波方法是一种有效的去噪方法^[19],其不仅在目标跟 踪领域^[20-21]应用广泛,在人体姿态估计领域也发挥

① 国家自然科学基金(61903335,61822311)资助项目。

② 男,1996年生,硕士生;研究方向:3D视觉人体姿态估计,信息融合;E-mail: jjxxll0227@163.com。

③ 通信作者, E-mail: wazhang@ zjut. edu. cn。 (收稿日期:2020-07-17)

重要的作用。例如,文献[22]利用线性卡尔曼滤波 器提高了人体姿态估计的准确性,文献[23]在卡尔 曼滤波的框架下融合来自多 Kinect 传感器的人体 骨骼位置信息以处理视觉自遮挡问题。然而,实际 上采集到的人体姿态信息中包含的噪声并非满足高 斯特性,即所采集到的人体姿态信息中往往会存在 多种不确定的噪声,这些噪声的特性往往难以精确 描述。

本文针对深度图像中 3D 人体姿态估计的视觉 遮挡问题,提出了一种带有鲁棒卡尔曼滤波(robust Kalman filter, RKF)的人体姿态估计方法。本文主 要贡献如下:在深度图像中,通过随机森林方法得到 的人体姿态信息中夹杂着遮挡造成的复杂噪声,所 提出的鲁棒卡尔曼滤波利用假设检验的方法对包含 的噪声进行识别和分类处理,提高了人体姿态估计 的精度和鲁棒性。

问题与建模 1

1.1 问题描述

考虑一类基于 3D 视觉传感器的人体姿态估计 系统。其中,3D 视觉传感器为 RGB-D 相机,可采集 人体目标的深度信息。通常可将人体目标视为由 头、躯干、臂、手、腿、足等部件相互连接构成的多刚 体系统,人体各关节点如图1所示。这样,人体姿态 估计问题可看作为人体各关节点位置估计问题,并



利用随机森林方法从深度图像中识别出人体各部 件,进而计算出人体各关节点在相机坐标系下的3D 位置。这种深度信息结合随机森林方法得到的人体 姿态估计虽然能够克服光照的影响,但仍难以避免 由视觉遮挡引起的误识别问题。特别地,这使得所 采集到的人体姿态信息中包含的噪声并非满足高斯 特性,即所采集到的人体姿态信息中往往会存在多 种不确定性噪声。

以人机协作中的应用分析为例,人与机器人处 于同一工作空间并同时进行工作,因此精准鲁棒的 人体姿态估计是保障人身安全及机器人正常工作的 重要前提。然而,仅依赖随机森林方法往往难以避 免视觉遮挡引起的误识别问题。如图2所示,自遮 挡或手持物遮挡的存在,都将导致人体重要关节点 的误识别。特别地,如图2(c)所示,人体关节点的 误识别可造成人体姿态量测上的野值问题。如图3



(a) 正常识别

(c) 误识别情形





所示,从采集到的人体腕关节点数据中不难发现,视 觉遮挡可造成腕关节运动轨迹的锯齿状,即人体姿 态的量测噪声统计特性发生了变化。为克服视觉遮 挡等引起人体关节点的误识别,本文设计了带有鲁 棒卡尔曼滤波的人体姿态估计方法来提高人体姿态 估计的精度和鲁棒性。

1.2 系统建模

考虑到视觉遮挡将导致人体姿态量测信息中存 在多种不确定的噪声,对人体姿态的量测模型建模 如下:

$$\boldsymbol{z}_{k} = \boldsymbol{H}\boldsymbol{x}_{k} + \boldsymbol{v}_{k} + \boldsymbol{\alpha}_{k}\boldsymbol{u}_{k} + \boldsymbol{a}_{k}\boldsymbol{\delta}(\boldsymbol{\beta}_{k})$$
(1)
$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

其中,
$$H = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$
为量测矩阵, $z_k =$

 $[\mathbf{z}_{k}^{1} \ \mathbf{z}_{k}^{2} \ \cdots \ \mathbf{z}_{k}^{j}]^{\mathrm{T}}$ 为k时刻人体姿态的量测值, \mathbf{z}_{k}^{j} = $[\mathbf{z}_{x,k}^{j} \ \mathbf{z}_{y,k}^{j} \ \mathbf{z}_{z,k}^{j}]^{\mathrm{T}}$ 为相机坐标系下人体各关节点 的位置量测值, $\mathbf{z}_{x,k}^{j}, \mathbf{z}_{y,k}^{j}, \mathbf{z}_{z,k}^{j}$ 分别为人体各关节点 在x, y和z轴上的量测值, $\mathbf{v}_{k} = [\mathbf{v}_{k}^{1} \ \mathbf{v}_{k}^{2} \ \cdots \ \mathbf{v}_{k}^{j}]^{\mathrm{T}}$ 为零均值且协方差为 R_{k} 的高斯白噪声, $\alpha_{k}\mathbf{u}_{k} =$ $\alpha_{k}[\mathbf{u}_{k}^{1} \ \mathbf{u}_{k}^{2} \ \cdots \ \mathbf{u}_{k}^{j}]^{\mathrm{T}}$ 为不确定性噪声, $a_{k} \delta(\boldsymbol{\beta}_{k}) =$ $a_{k}[\delta^{1}(\boldsymbol{\beta}_{k}) \ \delta^{2}(\boldsymbol{\beta}_{k}) \ \cdots \ \delta^{j}(\boldsymbol{\beta}_{k})]^{\mathrm{T}}$ 为量测野值,其 中 α_{k} 服从参数为 $p_{1}(0 < p_{1} < 1)$ 的伯努利分布, a_{k} 为幅值较大的数值, $\delta(\cdot)$ 为单位脉冲函数, $\boldsymbol{\beta}$ 服从 参数为 $p_{2}(0 < p_{2} < 1)$ 的伯努利分布。

人体各关节点的运动学模型描述如式(2)所示:

$$\begin{aligned} \boldsymbol{x}_{k} &= \boldsymbol{F}\boldsymbol{x}_{k-1} + \boldsymbol{w}_{k} \end{aligned} \tag{2}$$
其中,

$$\boldsymbol{F}_{0} = \begin{bmatrix} 1 & \Delta t & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \Delta t & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{F} = \operatorname{diag} \{ \boldsymbol{F}_{0} , \boldsymbol{F}_{0} \}$$

 $..., F_{0} \}$ 为状态转移矩阵, Δt 为采样周期, j = 1, 2,..., m为人体各关节点的序号, m = 25, k = 1, 2,... 为离散时间序列, $x_{k} = [x_{k}^{1} \ x_{k}^{2} \ ... \ x_{k}^{j}]^{T}$ 为 k时 刻人体姿态的状态, $x_{k}^{j} = [x_{x,k}^{j} \ x_{x,k}^{j} \ x_{y,k}^{j} \ x_{y,k}^{j}]^{T}$ - 1212 - 为人体各关节点的状态信息, $x_{x,k}^{j}, x_{y,k}^{j}, x_{z,k}^{j}$ 和 $x_{x,k}^{j}, x_{y,k}^{j}, x_{z,k}^{j}$ 分别为人体各关节点在 x , y和 z 轴上的位置和速度, $w_{k} = \begin{bmatrix} w_{k}^{1} & w_{k}^{2} & \cdots & w_{k}^{j} \end{bmatrix}^{T}$ 为零均值且协方差为 Q_{k} 的高斯白噪声。确定人体各关节点的初始状态及协方差分别为

$$E\{\mathbf{x}_{0}^{j}\} = \hat{\mathbf{x}}_{0|0}^{j}$$
(3)

$$E\{(\mathbf{x}_{0}^{j}-\hat{\mathbf{x}}_{0|0}^{j})(\mathbf{x}_{0}^{j}-\hat{\mathbf{x}}_{0|0}^{j})^{\mathrm{T}}\} = \mathbf{P}_{0|0}^{j} \qquad (4)$$

考虑到视觉自遮挡或遮挡造成的人体部件误识 别问题,设计了一种鲁棒卡尔曼滤波器用以提高人 体姿态估计系统对视觉遮挡的鲁棒性,该方法采用 假设检验的方式引入自适应因子,可有效提高系统 的估计精度和鲁棒性。

根据系统量测模型,当 $\alpha_k = 0 \pm \beta_k \neq 0$ 时,则 量测模型描述为

$$\boldsymbol{z}_k = \boldsymbol{H}\boldsymbol{x}_k + \boldsymbol{v}_k \tag{5}$$

根据系统模型式(2)和式(5),人体各关节点的 卡尔曼滤波可描述为

$$\hat{x}_{k|k-1}^{j} = F_{0}\hat{x}_{k-1|k-1}^{j}$$
(6)

$$\boldsymbol{P}_{k|k-1}^{j} = \boldsymbol{F}_{0}\boldsymbol{P}_{k-1|k-1}^{j}\boldsymbol{F}_{0}^{T} + \boldsymbol{Q}_{k}^{j}$$

$$(7)$$

$$\boldsymbol{K}_{k}^{j} = \boldsymbol{P}_{k|k-1}^{j} \boldsymbol{H}^{\mathrm{T}} (\boldsymbol{P}_{z,k|k-1}^{j})^{-1}$$

$$(8)$$

$$\hat{\boldsymbol{x}}_{k|k}^{j} = \hat{\boldsymbol{x}}_{k|k-1}^{j} + \boldsymbol{K}_{k}^{j} \boldsymbol{\eta}_{k|k-1}^{j}$$
(9)

$$\boldsymbol{P}_{k|k}^{j} = (\boldsymbol{I} - \boldsymbol{K}_{k}^{j}\boldsymbol{H})\boldsymbol{P}_{k|k-1}^{j}$$
(10)

其中, $\hat{x}_{k|k-1}^{i}$ 和 $\hat{x}_{k|k}^{i}$ 分别为人体各关节点状态的预测 值和估计值, $P_{k|k-1}^{i}$ 和 $P_{k|k}^{i}$ 分别为预测值和估计值的 协方差, K_{k}^{i} 为滤波增益。 $\eta_{k|k-1}^{i}$ 和 $P_{z,k|k-1}^{i}$ 分别为新息 及其协方差, 可描述为

$$\boldsymbol{\eta}_{k|k-1}^{j} = \boldsymbol{z}_{k}^{j} - \boldsymbol{H} \boldsymbol{\hat{x}}_{k|k-1}^{j}$$
(11)

$$\boldsymbol{P}_{\boldsymbol{z},\boldsymbol{k}|\boldsymbol{k}-1}^{j} = \boldsymbol{H}\boldsymbol{P}_{\boldsymbol{k}|\boldsymbol{k}-1}^{j}\boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R}_{\boldsymbol{k}}^{j}$$
(12)

令新息序列对应的马氏距离
$$\gamma_{j,k|k-1}^i(l)$$
 为

$$\begin{aligned} \gamma_{k|k-1}(l) &= (M_{k|k-1})^{\mathrm{T}} = \\ (\sqrt{(\eta_{k|k-1}^{j}(l))^{\mathrm{T}}E\{(\eta_{k|k-1}^{j}(l))^{\mathrm{T}}\eta_{k|k-1}^{j}(l)\}(\eta_{k|k-1}^{j}(l)))^{2}} \\ &= (\sqrt{(\eta_{k|k-1}^{j}(l))^{\mathrm{T}}(P_{z,k|k-1}^{j}(l,l))^{-1}(\eta_{k|k-1}^{j}(l)))^{2}} \end{aligned}$$

$$(13)$$

其中, $\eta_{k|k-1}^{j}(l)$ 为向量 $\eta_{k|k-1}^{j}$ 的第l个元素, $P_{z,k|k-1}^{j}(l,l)$ 为新息对应的方差, $\gamma_{k|k-1}^{j}(l)$ 为对角阵 γ'_{klk-1} 对角线上的第l个元素,l = 1,2,3。为方便分 析,这里设定 x, y 和 z 3 个方向上的新息序列互不 相关,则新息对应的马氏距离 $\gamma'_{klk-1}(l)$ 服从 χ^2 分 布,有

$$\Pr(\gamma_{k|k-1}^{j}(l) \ge \chi_{\alpha,o}^{j}) = \alpha_{o}^{j}$$
(14)

$$\Pr(\chi^{j}_{\alpha,\alpha} > \gamma^{j}_{k|k-1}(l) \ge \chi^{j}_{\alpha,d}) = \alpha^{j}_{d}$$
(15)

其中, $Pr(\cdot)$ 为某随机事件发生的概率, α_o^i 和 α_d^i 为 显著性水平, $\chi_{\alpha,o}^i$ 和 $\chi_{\alpha,d}^j$ 分别为 1 – α_o^j 和 1 – α_d^j 的置 信界。注意到: $\alpha_o^j < \alpha_d^j$, $\chi_{\alpha,o}^j > \chi_{\alpha,d}^j$ 。

当新息序列的马氏距离大于 $\chi'_{\alpha,o}$ 时,则认定此时出现量测野值,对其进行剔除并用预测值来补偿估计值,即

$$\hat{\boldsymbol{x}}_{k|k}^{j} = \hat{\boldsymbol{x}}_{k|k-1}^{j} \tag{16}$$

$$\boldsymbol{P}_{k|k}' = \boldsymbol{P}_{k|k-1}' \tag{17}$$

当新息序列的马氏距离略小于 $\chi_{\alpha,o}^{i}$ 且大于 $\chi_{\alpha,d}^{i}$ 时,则认为此时出现了不确定性噪声。针对不确定 性噪声的补偿问题,将引入自适应因子 μ_{k}^{j} 到量测噪 声协方差 R_{k}^{i} ,即引入自适应因子后的量测噪声协方 差 \overline{R}_{k}^{i} 为

$$\boldsymbol{R}_{k}^{j} = \boldsymbol{\mu}_{k}^{j} \boldsymbol{R}_{k}^{j} \tag{18}$$

其中, μ_{k}^{j} 和 R_{k}^{j} 为对角阵, $\mu_{k}^{j}(l)$ 和 $R_{k}^{j}(l,l)$ 分别为 对角阵 μ_{k}^{j} 和对角线 R_{k}^{j} 上的第l个元素,l = 1,2,3。 在求解自适应因子 μ_{k}^{j} 之前,引入定理1。

定理 1 当新息的马氏距离大于 $\chi^{i}_{\alpha,d}$ 时,即 $\gamma^{j}_{k|k-1} > \chi^{j}_{\alpha,d}I$,引入自适应因子 μ^{j}_{k} 。若自适应因子满足:

$$\boldsymbol{\mu}_{k}^{j} \geq \boldsymbol{Y}_{k}^{j}(\boldsymbol{R}_{k}^{j})^{-1}$$
(19)
则有:

$$\bar{\boldsymbol{\gamma}}_{k|k-1}^{j} \leq \boldsymbol{\chi}_{\alpha,d}^{j} \boldsymbol{I}$$
(20)

其中, Y_k^i 和 $\vec{\gamma}_{klk-1}$ 为对角阵, $Y_k^i(l)$ 和 $\vec{\gamma}_{klk-1}^j(l)$ 分别 为对角阵 Y_k^i 和 $\vec{\gamma}_{klk-1}$ 对角线上的第l个元素, l = 1, 2,3。

$$Y_{k}^{j}(l) = \frac{(\eta_{k|k-1}^{j}(l))^{2}}{\chi_{\alpha,d}^{j}} - P_{z,k|k-1}^{j}(l,l) + R_{k}^{j}(l,l)$$
(21)

 $\overline{\gamma}_{k|k-1}^{j}(l) = (\sqrt{(\eta_{k|k-1}^{j}(l))^{\mathrm{T}}(\overline{P}_{z,k|k-1}^{j}(l,l))^{-1}(\eta_{k|k-1}^{j}(l))})^{2}$ (22)

 $\vec{P}_{z,k|k-1}$ 和 $\vec{\gamma}_{k|k-1}$ 分别为引入自适应因子后的新息协 方差和马氏距离, I为三维单位矩阵。

证明 当
$$\boldsymbol{\gamma}_{k|k-1}^{j} \geq \boldsymbol{\chi}_{\alpha,d}^{j} \boldsymbol{I}$$
时,由式(19)可得:

$$\boldsymbol{Y}_{k}^{j}(\boldsymbol{R}_{k}^{j})^{-1} - \boldsymbol{\mu}_{k}^{j} \leq 0$$
(23)

将式(23)两边同时乘以 Rⁱ, 可得:

$$\left[\boldsymbol{Y}_{k}^{j}(\boldsymbol{R}_{k}^{j})^{-1} - \boldsymbol{\mu}_{k}^{j}\right]\boldsymbol{R}_{k}^{j} \leq 0$$
(24)

即:

$$\boldsymbol{Y}_{k}^{j} - \boldsymbol{\mu}_{k}^{j} \boldsymbol{R}_{k}^{j} \leqslant 0 \tag{25}$$

记 W_k^j 为对角阵,其对角元素为 $W_k^j(l) = \frac{(\eta_{k|k-1}^j(l))^2}{\chi_{n,d}^j}, l = 1,2,3$ 。则有:

$$Y_{k}^{j} - \boldsymbol{\mu}_{k}^{j}\boldsymbol{R}_{k}^{j} = W_{k}^{j} - N_{k|k-1}^{j} + \boldsymbol{R}_{k}^{j} - \boldsymbol{\mu}_{k}^{j}\boldsymbol{R}_{k}^{j} \qquad (26)$$
注意到:

$$\overline{\boldsymbol{P}}_{z,k|k-1}^{j} = \boldsymbol{P}_{z,k|k-1}^{j} - \boldsymbol{R}_{k}^{j} + \boldsymbol{\mu}_{k}^{j} \boldsymbol{R}_{k}^{j}$$
(27)

将式(26)和式(27)整合代人式(25)中,可得到:
$$W_{k}^{j}(\overline{P}_{z,k|k-1}^{j})^{-1} \leq I$$
 (28)

将对角元素
$$W_k^j(l)$$
 带入式(27)可得:
 $(\eta_{k|k-1}^j(l))^{\mathrm{T}}(\overline{P}_{z,k|k-1}^j(l,l))^{-1}(\eta_{k|k-1}^j(l)) \leq \chi_{\alpha,d}^j$
(29)

即:

$$\overline{\gamma}^{j}_{k|k-1}(l) \leq \chi^{j}_{\alpha,d} \tag{30}$$

则有:

$$\overline{\boldsymbol{\gamma}}_{k|k-1}^{j} \leqslant \boldsymbol{\chi}_{\alpha,d}^{j} \boldsymbol{I}$$
(31)

证毕。

 $\boldsymbol{\mu}_{k}^{j} =$

当 $\boldsymbol{\gamma}_{k|k-1}^{j} > \chi_{\alpha,d}^{j} I$ 时,根据定理 1,引入自适应因 子:

$$\boldsymbol{\mu}_{k}^{j} = \boldsymbol{Y}_{k}^{j} (\boldsymbol{R}_{k}^{j})^{-1}$$
(32)

当 $\boldsymbol{\gamma}_{kk} \leq \chi_{\alpha,d}^{i} I$ 时,则认为其满足高斯假设前提, 不需要引入自适应因子。于是,有:

$$\begin{cases} \left[\frac{(\boldsymbol{\eta}_{k|k-1}^{j})^{2}}{\boldsymbol{\chi}_{\alpha,d}^{j}} - \boldsymbol{P}_{z,k|k-1}^{j} + \boldsymbol{R}_{k}^{j}\right] (\boldsymbol{R}_{k}^{j})^{-1} \quad \boldsymbol{\gamma}_{k|k-1}^{j} > \boldsymbol{\chi}_{\alpha,d}^{j}\boldsymbol{I} \\ \boldsymbol{I} \qquad \boldsymbol{\gamma}_{k|k-1}^{j} \leqslant \boldsymbol{\chi}_{\alpha,d}^{j}\boldsymbol{I} \end{cases}$$

$$(33)$$

基于鲁棒卡尔曼滤波的人体姿态估计算法描述 及其流程图分别如算法1和图4所示。

算法1 基于鲁棒卡尔曼滤波的人体姿态估计 算法。

— 1213 —

(1) 初始化,获取传感器下的初始化状态 \hat{x}_{010}^{j} 及其协方差 P_{010}^{j} ;

(2) 由式(6)和式(7)计算状态预测值 $\hat{x}_{k|k-1}^{j}$ 及 其协方差 $P_{k|k-1}^{j}$;

(3) 读取 3D 视觉传感器下的深度图像,并基于随机森林方法计算得到 z_k;

(4) 由式(11)和式(12)计算新息 $\eta_{k|k-1}^{j}$ 及其协 方差 $P_{z,k|k-1}^{j}$;

(5) 由式(13)计算新息的马氏距离 $\gamma_{k|k-1}$;

(6) if $\boldsymbol{\gamma}_{k|k-1}^{j} < \boldsymbol{\chi}_{\alpha,d}^{j} \boldsymbol{I}$ then

(7) 跳到式(13)继续执行;

(8) else if $\chi^{j}_{\alpha,d} I \leq \gamma^{j}_{k|k-1} < \chi^{j}_{\alpha,o} I$ then

(9) 由式(31)计算自适应因子 μ_k^j ;

(10)由式(27)计算引入自适应因子后的新息 协方差 **P**ⁱ_{k|k-1};

(11) else 计算式(16)和式(17);





(12) end if;

(13) 由式(8)计算滤波增益 K_k;

(14) 由式(9) 和式(10) 计算状态估计值 $\hat{x}_{k|k}^{j}$ 及其协方差 $P_{k|k}^{j}$ 。

3 实验分析

在本节中,将分别从人体姿态估计的仿真和实 验两方面来验证所提方法的有效性。

3.1 仿真实验

在仿真中,以人体右臂腕关节点为例进行分析, 其在相机坐标系下做匀速运动。设定采样周期为 $\Delta t = 1 \text{ s}$,过程噪声协方差为 $Q_k = \text{diag} \{0.04 \text{ cm}^2, 0.025(\text{ cm/s})^2, 0.04 \text{ cm}^2, 0.03(\text{ cm/s})^2, 0.04 \text{ cm}^2, 0.028(\text{ cm/s})^2 \}$,不确定性噪声设为零均值且协方差 为 diag $\{2.25 \text{ cm}^2, 2.52 \text{ cm}^2, 2.35 \text{ cm}^2 \}$ 的高斯白噪 声,其出现概率为 $Pr(\alpha = 1) = 0.1$; 野值设为幅值大 小为 30 cm 的噪声,即 a = 30 cm,其出现概率为 $Pr(\beta = 1) = 0.05$ 。获取人体的右臂腕关节点初始化状态 及其协方差, $\mathbf{x}_{010} = [50.00 \text{ cm}, 5.00 \text{ cm/s}, 0.00 \text{ cm}, 25.00 \text{ cm/s}, 150 \text{ cm}, 20.00 \text{ cm/s}]$

 $P_{010} = \text{diag} \{0.50 \text{ cm}^2, 0.36(\text{ cm/s})^2, 0.45 \text{ cm}^2, 0.45 \text{ cm}^2\}$

 $0.23 (\text{cm/s})^2$, 0.37 cm^2 , $0.30 (\text{cm/s})^2$

为了更好地体现所提方法的有效性,定义估计误差衡量标准为均方根误差(root mean square error, RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^{n} (\hat{\mathbf{x}}_{k|k}^{j} - \mathbf{x}_{k|k}^{j})^{2}}$$
(34)

其中, $h = 1, 2, \dots, n$ 为仿真实验的序号, n 为蒙特卡 罗仿真次数, $\hat{x}_{k|k}^{j}$ 为 k 时刻的状态估计值, $x_{k|k}^{j}$ 为 k 时刻的状态真实值。

分别在量测(Meas)、卡尔曼滤波器(KF)、鲁棒卡 尔曼滤波器(RKF)下做 200 次蒙特卡罗仿真,仿真的 对比结果如图 4~图 6 所示。从中可看出,相比于 Meas 和 KF,基于 RKF 方法下得到的 RMSE 更小,表 明该方法具有更好的估计精度。进一步,表 1 中给出 了 3 种方法下的 RMSE 均值对比,其中 *e_x、e_y、e_z*分别 为在 *X、Y* 和 *Z* 轴上的位置估计 RMSE 均值。



图 5 X 轴方向上不同方法下的 RMSE 对比



图 6 Y 轴方向上不同方法下的 RMSE 对比

表1 仿具中谷方法卜的位置估计误差5	匀值
--------------------	----

方法	e_x /cm	e_y / cm	e_z /cm
Meas	23.78	24.43	24.05
KF	19.32	19.94	19.35
RKF	6.18	8.13	7.16

3.2 人体姿态估计实验

如图 7 所示,人体姿态估计实验平台由一台微软 公司的 Kinect V2 视觉传感器、一台工控机和一位人 体姿态估计对象组成。Kinect V2 视觉传感器包括彩 色摄像头和深度摄像头,采集到的彩色图像分辨率为 1920×1080,深度图像分辨率为 512×424,拍摄速度 为 30 fps。在工控机上,编写了一个基于 Visual Studio 2013 集成 Qt 5.7 的开发环境,利用随机森林方法得 到在深度相机空间下人体骨骼关节点的空间位置信 息。同时,将深度相机空间下的人体骨骼关节点投影 到彩色图像中,并利用计算机视觉库 OpenCV 对人体 骨架进行绘制以及彩色图像显示。



实验场景设置如下:实验环境位于室内,人体位 于 Kinect V2 正前方 1.5 m 左右的位置进行运动,用 Kinect V2 来完成对人体关节点的捕捉。这里需要补 充说明的是,人体关节点对应的实际人体位置并不明 确,即人体关节点的物理意义是不明确的。故以高精 度定位系统 OptiTrack(精度0.5 mm)来获取人体关节 点的真实轨迹,如图 8 所示。该定位系统由 8 个 Prime 13 相机组成,能够实时捕捉运动目标的位姿, 以追踪到的光学标记点的位置视为真值,即视为真实 人体关节点位置进行对比。



图 8 人体姿态估计系统

在本实验中,在随机森林方法识别并获取到人体 关节点位置信息的基础上分别采用 KF 和 RKF,捕捉 对象为人体右臂,其包括肩关节、肘关节和腕关节。 以人体右臂腕关节点分析为例,图9~图11 分别为其 在运动过程中的 X、Y 和 Z 坐标信息在不同方法处理 下的累积误差分析图。进一步,表 2 所示为腕关节点 以及肘关节点的具体误差均值,从中可看出,RKF 方 法下得到的误差更低。由此说明 RKF 方法能有效提 高人体姿态估计的精度和鲁棒性。另外,从 3 组关节 点误差均值的整体对比中,可看出腕关节点的误差相 对更大,肩关节点的误差相对更小,表明机动性更强 的关节点存在的误差也更大。



图9 OptiTrack 定位系统



图 10 X 轴方向上不同方法的累积误差对比







图 12 X 轴方向上不同方法的累积误差对比

表 2 实验中各方法下的累积误差均值

关节部位	实验方法	e_x / cm	e_y / cm	e_z / cm
右手肩部	RFM	2.78	4.20	3.12
	RFM + KF	2.42	3.15	2.55
	RFM + RKF	2.21	2.76	2.38
右手肘部	RFM	3.15	4.33	3.28
	RFM + KF	2.46	3.30	2.84
	RFM + RKF	2.40	3.16	2.68
右手腕部	RFM	6.09	8.76	5.54
	RFM + KF	4.98	7.63	4.88
	RFM + RKF	4.90	7.11	4.75

4 结论

为了处理人体姿态估计中的视觉遮挡问题,提出 了一种带有鲁棒卡尔曼滤波的人体姿态估计方法。 首先,采用随机森林方法从深度图像中识别出并得出 人体各关节点在相机坐标系下的 3D 位置;其次,考虑 到视觉自遮挡或遮挡造成的人体部件误识别,设计了 一种鲁棒的卡尔曼滤波器用以提高人体姿态估计对 视觉遮挡的鲁棒性。最后,仿真结果和实验结果表 明,所提方法能够有效提高人体姿态估计的精度和鲁 棒性。

本文在室内环境下进行自遮挡情形实验,从结果 可看出所提方法的有效性,然而复杂环境下对遮挡的 应对要求更高,在后续的研究中将测试复杂环境下 (如障碍物遮挡)该方法的鲁棒性。此外将在所获得 人体姿态数据的基础上进行一些人机协作实验,以验 证在实际应用中所提方法的有效性与实时性。

参考文献

- [1]张鋆豪,何百岳,杨旭升,等.基于可穿戴式惯性传感器的人体运动跟踪方法综述[J].自动化学报,2019,45(8):1439-1454
- [2] Zhang J H, Li P, Jin C C, et al. A novel adaptive Kalman filtering approach to human motion tracking with magneticinertial sensors[J]. *IEEE Transactions on Industrial Electronics*, 2020,67(10): 8569-8669
- [3] Aggarwal J K, Xia L. Human activity recognition from 3D data: a review[J]. Pattern Recognition Letters, 2014, 48: 70-80
- [4] Chan K C, Koh C K, George Lee C S. A 3-D-point-cloud system for human-pose estimation [J]. *IEEE Transactions* on Systems, Man, and Cybernetics: Systems, 2014, 44 (11): 1486-1497
- [5] Schwarz L A, Mkhitaryan A, Mateus D, et al. Estimating human 3d pose from time-of-flight images based on geodesic distances and optical flow[C] // Proceedings of the 2011 IEEE Conference on Face and Gesture Recognition, Santa Barbara, USA, 2011: 700-706
- [6] Kolahi A, Hoviattalab M, Rezazian T, et al. Design of a marker-based human motion tracking system [J]. Biomedical Signal Processing and Control, 2007, 2(1): 59-67
- [7] Canton-ferrer C, Casas J R, Pardas M. Towards a low cost multi-camera marker based human motion capture system
 [C] // Proceedings of the 2009 IEEE International Conference on Image Processing (ICIP), Cairo, Egypt, 2009: 2581-2584
- [8] Liu Z, Huang J, Han J, et al. Human motion tracking by multiple RGBD cameras [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 27 (9): 2014-2027
- [9] Yang L, Yang B, Dong H, et al. 3-D markerless tracking of human gait by geometric trilateration of multiple Kinects
 [J]. *IEEE Systems Journal*, 2016, 12(2): 1393-1403
- [10] Shotton J, Fitzgibbon A, Cook M, et al. Real-time human pose recognition in parts from single depth images [C] // Proceedings of the 24th IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, USA, 2011: 1297-1304
- [11] Girshick R, Shotton J, Kohli P, et al. Efficient regression of general-activity human poses from depth images [C] // Proceedings of the IEEE 2011 International Conference on

Computer Vision, Barcelona, Spain, 2011: 415-422

- [12] Hesse N, Stachowiak G, Breuer T, et al. Estimating body pose of infants in depth images using random ferns [C] // Proceedings of the 2015 IEEE International Conference on Computer Vision Workshops, Santiago, Chile, 2015: 35-43
- [13] Yub Jung H, Lee S, Seok Heo Y , et al. Random tree walk toward instantaneous 3D human pose estimation [C] // Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 2467-2474
- [14] Haque A, Peng B, Luo Z, et al. Towards viewpoint invariant 3D human pose estimation [C] // Proceedings of the European Conference on Computer Vision, Amsterdam, the Netherlands, 2016; 160-177
- [15] Rafi U, Gall J, Leibe B. A semantic occlusion model for human pose estimation from a single depth image [C] // Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, USA, 2015: 67-74
- [16] Shuai L, Li C, Guo X, et al. Motion capture with ellipsoidal skeleton using multiple depth cameras [J]. IEEE Transactions on Visualization and Computer Graphics, 2016, 23(2): 1085-1098
- [17] Fabrizio F, De Luca A. Real-time computation of distance to dynamic obstacles with multiple depth sensors [J].
 IEEE Robotics and Automation Letters, 2016, 2(1); 56-63
- [18] Mallick T, Das P P, Majumdar A K. Characterizations of noise in Kinect depth images: a review [J]. *IEEE Sensors Journal*, 2014, 14(6): 1731-1740
- [19] 吕东辉, 王炯琦, 熊凯, 等. 适用处理非高斯观测噪声
 的强跟踪卡尔曼滤波器[J]. 控制理论与应用, 2019, 36(12): 1997-2004
- [20] 杨旭升,张文安,俞立.适用于事件触发的分布式随机 目标跟踪方法[J].自动化学报,2017,43(8):1393-1401
- [21] 杜华臻,张文安,杨旭升. 基于激光雷达的移动机器人 人体目标跟随[J]. 高技术通讯, 2019, 29(12): 1239-1246
- [22] Pathirana P N, Li S, Trinh H M, et al. Robust real-time bio-kinematic movement tracking using multiple kinects for tele-rehabilitation [J]. *IEEE Transactions on Industrial Electronics*, 2015, 63(3): 1822-1833

— 1217 —

[23] Moon S, Park Y, Ko D W, et al. Multiple kinect sensor fusion for human skeleton tracking using Kalman filtering [J]. International Journal of Advanced Robotic Systems, 2016, 13(2): 1-10

Robust estimation of human pose under visual occlusion

Jia Xiaoling, Zhang Wenan, Yang Xusheng

(College of Information Engineering, Zhejiang Provincial United Key Laboratory of Embedded Systems,

Zhejiang University of Technology, Hangzhou 310023)

Abstract

A human pose estimation method with robust Kalman filter (RKF) is presented aiming at the problem of visual occlusion in human pose estimation. Firstly, the random forest method is used to identify each part of the human body from the depth image and calculate the 3D position of each node of the human body in the camera coordinate system. Secondly, considering the false recognition of human body parts caused by visual self-occlusion or occlusion, a robust Kalman filter is designed to identify and classify the complex noises in the human pose information caused by visual occlusion by using the method of hypothesis testing, so as to improve the robustness of human pose estimation against visual occlusion. Finally, by the simulation results and 3D position estimation experiments of multiple human body joints, it demonstrates that improved accuracy and robustness of the human pose estimation can be achieved by the proposed method.

Key words: visual occlusion, human posture estimation, robust Kalman filter (RKF)