

DOI:10.16356/j.2097-6771.2026.02.021

基于姿态约束的 3D 人体网格细化算法

朱 炜, 吉根林, 王泽宇, 李 宁

(南京师范大学计算机与电子信息/人工智能学院, 南京 210023)

摘要: 3D 人体网格恢复 (Human mesh recovery, HMR) 在虚拟现实、增强现实、自动驾驶和运动科学等领域具有广泛应用前景。然而, 现有方法面临着 3D 模型与 2D 关键点无法精确对齐的问题, 且在恢复过程中常忽视人体固有的对称性和比例约束, 导致恢复的 3D 人体网格模型不符合人体基本特征。本文提出了一种基于姿态约束的人体网格细化 (Pose-constrained human mesh refinement, PC-HMR) 算法, 通过多维约束融合策略优化形状参数和姿态参数。在形状参数优化方面, 引入人体对称性特征和骨骼比例约束; 在姿态参数优化方面, 设计了一种多算法融合的姿态约束建模框架, 通过系统性整合多种互补算法来替代传统的乘法路径选择机制。为了评估本文方法的有效性, 在 3DPW 和 Human3.6M 两个标准数据集上进行了实验。实验结果表明, 相比现有最优方法, PC-HMR 综合性能更优。消融实验进一步验证了各模块对算法性能的影响。

关键词: 人体网格恢复; 姿态约束; 形状参数优化; 姿态参数优化; 姿态对称性

中图分类号: TP391.41; **文献标志码:** A **文章编号:** 1005-2615(2026)02-0449-08

Pose-Constrained 3D Human Mesh Refinement Algorithm

ZHU Wei, JI Genlin, WANG Zeyu, LI Ning

(School of Computer and Electronic Information/Artificial Intelligence, Nanjing Normal University, Nanjing 210023, China)

Abstract: 3D human mesh recovery (HMR) has extensive applications in virtual reality, augmented reality, autonomous driving, and sports science. However, the existing methods suffer from imprecise alignment between 3D models and 2D keypoints and often neglect the inherent symmetry and proportion constraints of the human body, resulting in recovered meshes that violate fundamental human characteristics. Our paper proposes a pose-constrained human mesh refinement (PC-HMR) algorithm that optimizes shape and pose parameters through constraint fusion. For shape parameter optimization, we introduce human body symmetry features and skeletal proportion constraints. For pose parameter optimization, we design a multi-algorithm fusion framework that systematically integrates complementary algorithms to replace traditional single-path selection mechanisms. We evaluate the proposed method in two standard datasets: 3DPW and Human3.6M. Experimental results demonstrate that PC-HMR achieve superior performance compared to existing state-of-the-art methods. Ablation studies further validate the contribution of each module to overall performance.

Key words: human body mesh recovery; pose constraints; shape parameter optimization; pose parameter optimization; skeletal symmetry

3D 人体网格恢复 (Human mesh recovery, HMR) 旨在由 2D 人体图像重建具有精确空间几何的 3D 人体网格模型。该技术在 VR/AR 的人机交互、自动驾驶的行人行为理解, 以及运动科学的动作量化分析中均具有重要应用价值, 因而在计算机视觉与人工智能领域受到广泛关注^[1-2]。

互、自动驾驶的行人行为理解, 以及运动科学的动作量化分析中均具有重要应用价值, 因而在计算机视觉与人工智能领域受到广泛关注^[1-2]。

基金项目: 国家自然科学基金 (41971343, 62102186)。

收稿日期: 2025-07-05; **修订日期:** 2025-10-24

通信作者: 吉根林, 教授, 博士生导师, E-mail: glji@nynu.edu.cn。

引用格式: 朱炜, 吉根林, 王泽宇, 等. 基于姿态约束的 3D 人体网格细化算法 [J]. 南京航空航天大学学报 (自然科学版), 2026, 58(2): 449-456. ZHU Wei, JI Genlin, WANG Zeyu, et al. Pose-constrained 3D human mesh refinement algorithm [J]. Journal of Nanjing University of Aeronautics & Astronautics (Natural Science Edition), 2026, 58(2): 449-456.

现有方法主要分为基于模板^[3-4]与非模板^[5-6]两类。基于模板的方法通过引入预定义的人体模型(如 SMPL^[7])降低问题复杂度,并在 Human3.6M^[8]、3DPW^[9]等数据集上取得较低的经普罗克拉斯提斯分析对齐的平均每关节位置误差(Procrustes analysis mean per joint position error, PA-MPJPE)(如 Tore^[10]、SGRE^[11])。非模板方法近年来亦有推进,但在遮挡与复杂姿态场景下整体仍弱于基于模板的方法(如 GloT^[12]、TCMR^[13]、MPS-net^[14])。同时,随着研究演进,回归与优化、模板与非模板方法的界限趋于模糊,融合范式展示出良好前景,例如 HybrIK^[15]将参数化人体模型与非参数化 3D 关键点估计有效结合。

在实际应用中,二维关键点与三维网格的对齐是关键挑战。大量工作通过 2D 重投影损失细化 SMPL 参数,如 SMPLify^[16]在优化中引入姿态先验与关节限制,SPIN^[17]通过“回归-优化”闭环互促,Guan 等^[18]基于 SCAPE 从手动标注的 2D 关节估计初始 3D 骨架,并融合轮廓、边缘与阴影线索进行细化且显式进行身体部位间的碰撞检测以确保合理性,EFT^[19]对样本级微调网络权重,CLIFF^[20]在全帧上计算 2D 重投影并直接预测全局旋转以改善全局一致性。这一系列方法依赖 2D 约束,易受深度歧义影响且泛化性受限。最新的研究 KITRO^[21]利用运动学旋转树实现姿态参数的闭式最优解、以 3D 骨长替代 2D 重投影优化形状,显著提升精度,但仍可能出现违背人体基本先验(如左右对称性、整体结构一致性)的情况。

基于上述观察,本文提出基于姿态约束的人体网格细化算法 PC-HMR。以 CLIFF 为基准,本文构建多约束融合框架:在形状参数优化中显式引入人体左右对称性与骨骼长度比例约束;在姿态参数优化中重构优化流程,缓解传统路径选择的局限;并综合多种策略提升整体稳健性。相较仅依赖单一骨长或 2D 重投影的优化,PC-HMR 更好地遵循人体先验,缓解由结构不一致带来的网格不合理问题。

本文的主要贡献如下:

(1) 本文针对现有 3D 人体网格恢复方法中存在的局限性,设计了一种基于多约束融合的优化算法。该算法以 CLIFF 模型为基础,通过引入多层次的人体先验知识约束,有效提升了 3D 人体网格恢复的准确性。

(2) 本文提出将人体固有的左右对称性作为明确的优化约束条件,并结合骨节长度比例约束,构建了更符合人体特性的形状参数优化算法。相比于现有方法仅考虑单一骨长约束或简单的 2D 重投影误差,本方法通过多维约束显著提升了恢复人体模型的合理性,有效缓解了现有方法中存在的问题。

(3) 在 3DPW 和 Human3.6M 两个公开数据集上的广泛实验验证了 PC-HMR 算法的有效性。具体而言,在 3DPW 数据集上,PA-MPJPE、平均每关节位置误差(Mean per joint position error, MPJPE)和平均顶点位置误差(Mean per vertex position error, PVE)分别取得了 0.98、2.51、3.41 mm 的改进。在 Human3.6M 数据集上,PA-MPJPE 改进了 0.49 mm。该算法在维持计算效率的同时,实现了 SOTA 性能。

1 本文方法

1.1 SMPL 模型

SMPL 模型是目前最常用的 3D 人体网格恢复模型。它将姿态参数 $\theta \in \mathbb{R}^{P \times 3}$ 和形状参数 $\beta \in \mathbb{R}^{10}$ 映射到 3D 人体网格 $V \in \mathbb{R}^{N \times 3}$ 。其中 $N=6890$ 代表定点的数量, $P=24$ 代表身体关节的数量。 $J(\beta, \theta) \in \mathbb{R}^{P \times 3}$ 表示为姿态参数 θ 和形状参数 β 映射到关节位置的函数。骨骼长度定义为相邻关节对(父关节 p 与其对应子关节 c)间的欧氏距离度量为

$$bl_{3D}^{(p,c)} = \|J^p(\beta, \theta) - J^c(\beta, \theta)\|_2 \quad (1)$$

1.2 PC-HMR 算法框架设计

本文提出的 PC-HMR 算法框架如图 1 所示。图中参数变量含义参见文献[20]。首先利用 CLIFF

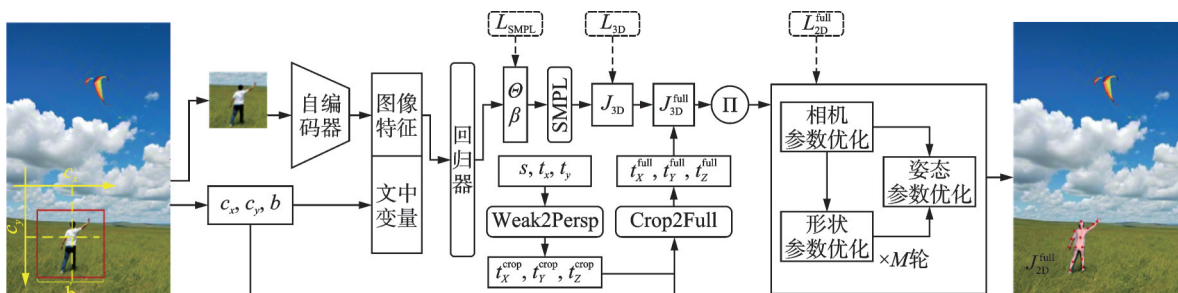


图 1 PC-HMR 整体框架图

Fig.1 Overall architecture of PC-HMR

对相机、形状与姿态进行初始化。具体做法为:以检测框为中心对全图进行正方形裁剪并缩放至固定分辨率,同时将检测框在全图中的中心偏移与尺寸经焦距归一化后编码为附加输入,与裁剪图像的视觉特征一并送入回归器。凭借全图位置信息,CLIFF可在原始相机坐标系下直接预测人体的全局旋转,并回归 SMPL 形状与关节姿态;其在裁剪坐标系下得到的弱透视相机参数再转换为等效透视表述,结合检测框的中心偏移与尺度关系映射回全图坐标系,从而得到与全图内参一致的根部位移。本文据此构建初始化。随后,采用一种交替优化策略,分别对姿态参数和形状参数进行独立更新。在每一类参数的更新迭代中,其余已细化的参数将被固定,以保证优化过程的稳定性和有效收敛。

在单目人体网格恢复任务中,给定输入图像 $X \in \mathbf{R}^{H \times W}$,模型的目标是预测 SMPL 参数 (θ, β) 以及作为相机外参的平移向量 $t \in \mathbf{R}^3$ 。为了进一步提升人体网格恢复的精度,一些基于优化的方法引入了 2D 关键点 $j \in \mathbf{R}^{P \times 2}$ 辅助优化,通过优化框架对初始估计的 SMPL 参数和相机平移参数 $(\hat{\theta}, \hat{\beta}, \hat{t})$ 进行精调。这一优化过程可以表述为

$$(\hat{\theta}', \hat{\beta}', \hat{t}') = \text{Refine}(\hat{\theta}, \hat{\beta}, \hat{t}, j) \quad (2)$$

上述优化过程可采用迭代式策略进行求解。具体而言,将每次优化后得到的参数估计值 $(\hat{\theta}', \hat{\beta}', \hat{t}')$ 作为下一轮优化的初始值的方法,从而实现参数的逐步精调。一个广泛使用的优化目标是最小化估计的 3D 关节 $J(\beta, \theta)$ 投影到 2D 平面后与给定的 2D 关键点 j 之间的重投影误差,有

$$\mathcal{L}_{j2D} = \min_{\theta, \beta, t} \left(\pi(J(\beta, \theta) + t) - j \right)_2 \quad (3)$$

式中 π 表示相机投影函数。值得注意的是,该优化问题支持灵活的求解策略:可以同时优化所有参数,也可以固定部分参数而仅优化其余参数。特别地,当姿态参数 θ 和形状参数 β 保持固定时,相机平移参数 t 的最优解可通过最小二乘法直接求得闭式解。

1.2.1 相机参数优化

遵从之前的工作采用完整的透视相机模型,包括内参和外参。与以往基于弱视角假设的方法不同,采用全视角相机模型,以更好地处理复杂的透视投影场景。

遵循当前已有研究的惯例,采用简化的相机参数设定:通过将相机旋转矩阵固定为单位矩阵 I ,并将旋转信息整合到人体模型的全局朝向参数 θ^0 中。这种处理方式使得相机外参仅需估计平移向量 t 。

基于上述简化策略,相机参数的优化可转化为求解平移向量 t 的最小二乘问题,有

$$t^* = \arg \min_t \left(\pi(J(\beta, \theta) + t) - j \right)_2 \quad (4)$$

式中 t^* 可以用奇异值分解(Singular value decomposition, SVD)表示。从迭代更新的角度来看,可以发现用加权移动平均线更新 t 更有效。对于当前摄像机平移 t ,更新后的 t' 为

$$t' = (t^* + t) / 2 \quad (5)$$

引入上述公式的主要原因是最优平移向量 t^* 容易受到姿态参数 θ 和形状参数 β 中噪声的干扰。从这个角度来看,移动平均策略不仅保留了 HMR 模型历史预测的有效信息,同时也起到了正则化器的作用,能够有效抑制参数优化过程中的噪声扰动,提高预测结果的稳定性。

1.2.2 形状参数优化

传统方法主要依赖单个关节的 2D 重投影损失来优化人体形状参数,这种方法不可避免地丢失了有效的深度信息。为了克服这一局限性,近期的研究方法 KITRO 提出了一种新的优化策略:采用 3D 骨骼长度约束来取代传统的 2D 重投影损失。具体的方法是,考虑索引为 p 的父关节,其 3D 关节位置为 $J^p(\beta, \theta)$,对应的 2D 关节位置为 j^p ;同样地,对于索引为 c 的子关节,其 3D 关节位置为 $J^c(\beta, \theta)$,对应的 2D 关节位置为 j^c 。 (p, c) 关节对之间的投影 2D 骨骼长度定义为两个关节之间的欧几里得距离,有

$$bl_{\text{proj}}^{(p,c)}(\beta) = \left\| \pi(J^p(\beta, \theta) + t') - \pi(J^c(\beta, \theta) + t') \right\|_2 \quad (6)$$

根据本文提出的优化策略,在优化形状参数时固定相机参数和姿态参数。因此,上述 3D 骨骼长度的计算公式实质上仅受形状参数 β 的影响。基于这一特性,可以类似地定义 2D 关节骨长有

$$bl_{2D}^{(p,c)} = (j^p - j^c)_2 \quad (7)$$

为了估计形状损失,考虑方程中的投影骨骼长度与方程中给定骨骼长度之间的 L1 范数有

$$L(\beta) = \sum_{p,c} \left| bl_{\text{proj}}^{(p,c)}(\beta) - bl_{2D}^{(p,c)} \right| \quad (8)$$

利用人体姿态约束特征,将骨骼对称性约束应用在 3D 人体网格恢复任务中。本文提出的方法直接在三维空间中进行优化,充分利用人体姿态特征的对骨骼应该对称的特性,通过约束对应骨骼对的长度相等性来构建形状损失函数。这种方法提高了恢复后的人体网格的合理性,利用了三维信息,对缓解人体网格恢复固有的深度歧义性起到了一定的作用,即

$$L_{\text{sym}}(\beta) = \sum_{(i,j) \in \delta} \left| \frac{(J^i(\beta, \theta) - J^{p(i)}(\beta, \theta))_2}{(J^j(\beta, \theta) - J^{p(j)}(\beta, \theta))_2} \right| \quad (9)$$

式中: i 和 $p(i)$ 分别表示父关节及其对应的子关节; j 则代表骨架结构中 与关节 i 构成对称关系的对应关节点; (i, j) 为对称的关节对, 本文采用 SPIN 协议中定义的 24 个关键骨骼点作为基准骨架结构, 这些骨骼点能够有效地表征人体的主要关节位置和运动特征, 具体的关节对如表 1 所示。

表 1 SPIN 协议对称关节对

Table 1 Symmetric joint pairs of the SPIN protocol

对称关节对	左侧部位	右侧部位
(1,2)	左髋关节	右髋关节
(4,5)	左膝关节	右膝关节
(7,8)	左踝关节	右踝关节
(16,17)	左肩关节	右肩关节
(18,19)	左肘关节	右肘关节
(20,21)	左腕关节	右腕关节
(22,23)	左手	右手

除此之外, 人体骨骼结构遵循严格的比例约束。以 SPIN 模型的下肢骨骼系统为例, 股骨(从髋关节至膝关节)与胫骨(从膝关节至踝关节)的长度比例在成年人中呈现显著的统计稳定性, 通常保持在 1.26:1 的比值范围内。具体而言, 在 24 关键点骨架模型中, 若定义右侧髋关节至膝关节的欧氏距离为股骨长度, 膝关节至踝关节的欧氏距离为胫骨长度, 这一解剖学比例在人群统计数据中表现出高度一致性。基于以上内容, 本文提出了一种改进的形状参数约束策略, 该策略不仅整合了传统的对称性约束, 更引入了基于解剖学比例范围的约束机制。这种多维度的约束框架为 3D 人体网格恢复提供了可靠的先验知识。

$$L_{\text{ratio}}(\beta) = \sum_{r \in R} \left| \frac{(J^{r_1}(\beta, \theta) - J^{p(r_1)}(\beta, \theta))_2}{(J^{r_2}(\beta, \theta) - J^{p(r_2)}(\beta, \theta))_2} - R_r \right| \quad (10)$$

式中: r_1 为比例约束项 r 中近端骨骼的子关节索引编号; $p(r_1)$ 为关节 r_1 对应的父关节索引编号; r_2 为比例约束项 r 中远端骨骼的子关节索引编号; $p(r_2)$ 为关节 r_2 对应的父关节索引编号。将人体骨骼的标准比例范围定义为 R_r 。根据 Bogin 等^[22] 的研究, 人体大腿与小腿的长度比例通常维持在 1.2 ± 0.07 的范围内。而在上肢比例方面, Wang 等^[23] 的研究表明, 上臂与前臂的长度比值约为 1.18 ± 0.06 。这

些精确的解剖学比例为本文的研究提供了重要参考标准。

$$L_{\text{total}}(\beta) = \lambda_{\text{bone}} L_{\text{bone}}(\beta) + \lambda_{\text{sym}} L_{\text{sym}}(\beta) + \lambda_{\text{ratio}} L_{\text{ratio}}(\beta) + \lambda_{\text{reg}} L_{\text{reg}}(\beta) \quad (11)$$

式中 $L_{\text{reg}}(\beta)$ 为形状参数的 L2 正则化项。与可以通过闭式解直接求得的相机平移参数 t 不同, 最优形状参数 β 的求解过程较为复杂。为此, 采用基于梯度下降的优化策略, 通过迭代最小化公式中定义的损失函数, 从而得到优化后的形状参数 β' 为

$$\beta' = \beta - \eta \nabla_{\beta} L_{\text{total}}(\beta) \quad \text{for } T \text{ steps} \quad (12)$$

式中: η 为学习率, 优化需要 T 轮, $\nabla_{\beta} L_{\text{total}}(\beta)$ 为总形状损失函数的梯度; $L_{\text{total}}(\beta)$ 为对形状参数 β 的梯度, 梯度方向为损失函数上升最快的方向。

1.2.3 姿态参数优化

在优化姿态参数 θ 时, 沿用 KITRO 的方法: 摆动-扭转分解来表示关节旋转, 这与 SMPL 模型中传统的轴角表示形式不同。这种选择在 PC-HMR 的优化框架中具有关键意义, 因为它能够直接从输入的 2D 关键点坐标中通过闭式解高效地估计摆动角度。

在 SMPL 运动学树结构中, 考虑一对具有父子关系的骨骼节点 (p, c) , 其对应的 2D 关键点分别为 J^p 和 J^c 。这些 2D 关键点是其 3D 关节位置 $J^p(\beta, \theta)$ 和 $J^c(\beta, \theta)$ 在图像平面上的投影。由于采用单位旋转矩阵, 相机中心 O 可以表示为相机平移向量的负值, 即 $O = -t'$ 。从相机中心 O 出发, 通过 2D 关键点 J^p 和 J^c 分别形成两条射线 p' 和 c' , 这两条射线必然分别穿过对应的 3D 关节位置。这些射线的方向可以表示为

$$p' = K^{-1} \times j_p^h \quad (13)$$

$$c' = K^{-1} \times j_c^h \quad (14)$$

式中: 下标 h 表示齐次坐标, K 为定义的相机内参矩阵。基于射线 p' 和 c' 的方向, 它们的交角 α 满足

$$\cos \alpha = \frac{p' \cdot c'}{\|p'\| \|c'\|} \quad (15)$$

$$\sin \alpha = \frac{\|p' \times c'\|}{\|p'\| \|c'\|} \quad (16)$$

根据已知的 3D 骨骼长度 $bl_{3D}^{(p,c)}$ 和父关节到相机的深度 $|OJ^p| = (J^p(\beta', \theta) + t')_2$ 来求解子关节。具体的求解方式为通过父关节 J^p 和到相机 O 的深度 $|OJ^p|$ 来计算由向量 $J^p J^c$ 指定的子关节 J^c , 有

$$J^p J^c = \underbrace{|OJ^p| \cdot (\cos \alpha \cdot c' - p')}_{J^p J^c} \pm \underbrace{|FJ_0^c| \cdot c'}_{FJ_0^c} \quad (17)$$

$$|OJ^c| = |OJ^p| \cdot \cos \alpha \pm |FJ_0^c| \quad (18)$$

正如式(17)中所展现的那样,子关节的方向对应着两种可能的方向。一种面向相机,一种背离相机。这两种可能直接展现了在 3D 人体网格恢复中一直存在的深度歧义性。该现象揭示了三维人体网格恢复任务的本质挑战。

如式(18)所示,子关节的空间方向与其父关节的深度值 $|OJ^p|$ 存在显著的依赖关系。由于父关节本身的深度值具有歧义性,这种依赖关系会沿着运动学树逐层传播,最终形成一个具有分支结构的二叉树。如图 2 所示,以人体左侧下肢为例,从骨盆作为根节点出发,经由髋关节、膝关节、踝关节直至脚趾端点,清晰地展现了这种树状结构。在完整的人体姿态表示中,共包含 5 个独立的运动学树,分别对应上肢、下肢和躯干。因此,完整的全身姿态假设可以通过在这 5 棵二叉树中各自选择一条有效路径的组合来表征。

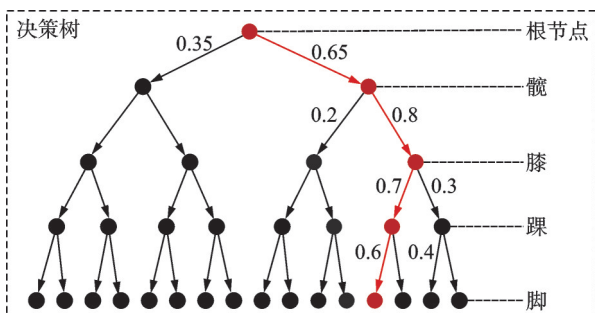


图 2 决策树示例(以左腿为例)

Fig.2 Example of decision tree (Taking the left leg as an example)

为了从二叉树中选择一条最优路径,选择的方法是依赖于初始的 HMR 估计计算两个封闭的解原始 HMR 模型预测的相对骨旋转之间的余弦相似度为

$$\cos_{\text{sim}}(J^p J_k^c | \phi(p)) = \frac{\cos \langle R_{\text{rel}}(J^p J_k^c | \phi(p)), \theta_{\text{HMR}}^p \rangle + 1}{2} \quad (19)$$

式中 $\phi(p) = (\phi_0, \phi_1, \dots, \phi_i, \dots, \phi_p)$ 表示从根节点到当前父节点 p 的路径。除了根节点 ϕ_0 (因为在每次迭代后相机优化后是固定的) 之外,路径中的每个 $\phi_i \in \{0, 1\}$ 表示对应骨骼的两个可能解,其中 J_i 是该骨骼的子关节。如图 2 所示,直接将这两个余弦相似度的 Softmax 作为决策树中边的权重。

式(20)通过对候选骨骼方向与当前骨骼方向加权融合,实现姿态参数的更新,有

$$\mathbf{b}_{\text{new}}^{(p,c)} = \lambda^{(p,c)} \cdot \mathbf{J}^p \mathbf{J}_c^c + (1 - \lambda^{(p,c)}) \cdot \mathbf{b}^{(p,c)} \quad (20)$$

以上公式构建了一个量化评估框架,用于度量计算结果与初始 HMR 预测值之间的一致性。

式(21)将两个候选解与初始 HMR 预测之间的余弦相似度经过 Softmax 归一化为边权重,有

$$\omega(e_k^{(p,c)} | \phi(p)) = \frac{\exp(\cos_{\text{sim}}(J^p J_k^c | \phi(p)))}{\sum_{k=0}^1 \exp(\cos_{\text{sim}}(J^p J_k^c | \phi(p)))} \quad (21)$$

通过构建二叉树结构并进行系统的权值估计,建立了一个完整的姿态可行性评估框架。该框架能够有效地表征姿态空间中的全局特征分布。在此基础上,采用节点积与累积加权值的综合评分作为路径优化准则,以确保在考虑局部特征的同时保持全局一致性。最终的最优姿态由具有最大综合评分的路径唯一确定,这种方法既保证了计算效率,又确保了结果的稳定性。

$$S_{\text{mult}} = \prod_{(p,c)} \omega(e_{\phi^*}^{(p,c)} | \phi(p)) \quad (22)$$

$$S_{\text{add}} = \sum_{(p,c)} \frac{\cos_{\text{sim}}(J^p J_k^c | \phi(p))}{N} \quad (23)$$

$$S_{\text{combined}} = \alpha S_{\text{mult}} + (1 - \alpha) S_{\text{add}} \quad (24)$$

$$\phi^* = \arg \max_{\phi} S_{\text{combined}} \quad (25)$$

式中: $\phi = \{\phi_0, \phi_1, \dots, \phi_{23}\}$ 为人体所有 24 块骨头的姿势假设, α 为 0.9。

在 PC-HMR 的框架中,姿态参数调整依据式(25)中的选定链 ϕ^* 进行。根据式(25)的选择链 ϕ^* ,对当前姿态参数进行更新。更新过程通过边权重 $\omega(e_{\phi^*}^{(p,c)} | \phi^*(p))$ 加权,以表示选中链的置信度。加权因子为

$$\lambda^{(p,c)} = \omega(e_{\phi^*}^{(p,c)} | \phi^*(p)) \quad (26)$$

骨骼的最终方向向量 $\mathbf{b}_{\text{new}}^{(p,c)}$ 为当前骨骼方向 $\mathbf{b}^{(p,c)}$ 和选中解的加权和,公式为

$$\mathbf{b}_{\text{new}}^{(p,c)} = \lambda^{(p,c)} \cdot \mathbf{J}^p \mathbf{J}_c^c + (1 - \lambda^{(p,c)}) \cdot \mathbf{b}^{(p,c)} \quad (27)$$

然后,父关节 p 的姿态参数 θ^p 根据 Rodrigues 公式旋转 $\mathbf{b}^{(p,c)}$ 到新的方向 $\mathbf{b}_{\text{new}}^{(p,c)}$ 。对于特殊情况(如根关节或第三颈椎关节),可通过 SVD 确定最佳旋转矩阵 $R_{\text{sw}}^{(p,c)}$,优化调整骨骼方向,使 $\mathbf{b}_{\text{new}}^{(p,c_0)}$ 、 $\mathbf{b}_{\text{new}}^{(p,c_1)}$ 和 $\mathbf{b}_{\text{new}}^{(p,c_2)}$ 对齐。优化更新后,父关节 P 的旋转矩阵形式的姿态参数为

$$\theta_r^p = \left(\prod_{i \in \text{KC}(\bar{p})} \theta_r^i \right)^T \cdot R_{\text{sw}}^{(p,c)} \cdot \prod_{i \in \text{KC}(p)} \theta_r^i \quad (28)$$

式中 θ_r^i 为集合 $\text{KC}(p)$ 中第 i 个祖先关节的旋转矩阵。

式(5, 12, 28)描述了对参数 t 、 β 和 θ 进行单次优化的过程。这一优化框架可进一步扩展为多次迭代,即将每轮优化后得到的参数值用作下一轮计算的初始条件,如此重复 M 次。PC-HMR 的流程如算法 1 所示。

算法 1 PC-HMR 人体网格细化过程

输入: 初始姿态 θ_0 , 形状 β_0 , 相机平移 t_0 , 2D 关

键点 J , 迭代次数 M , 运动学树 KT;

输出: 精细化的 θ_M, β_M, t_M ;

(1) for $m = 0$ to $M-1$ do

(2) $t^* = \text{Best } t \text{ aligned } j \text{ and } J(\beta_m, \theta_m)$

(3) $t_{m+1} = (t^* + t_m)/2$

(4) $\beta_{m+1} = \text{AdamW optimize } \beta_m \text{ by } L_{\text{total}}$

(β)//通过 $L_{\text{total}}(\beta)$ 优化 β_m

(5) DecisionTree = Binary solutions along KT//沿运动学旋转树的二元解

(6) $\varnothing^* = \text{Optimal path in Decision Tree}$ //决策树的最优路径

(7) $\theta_{m+1} = \text{Update bone rotation base on } \varnothing^*$

(8) end for

2 实验与结果分析

2.1 实验设置

本文采用了 3D 人体网格恢复领域的两个标准基准数据集 (3DPW 和 Human3.6M。) 进行实验验证。3DPW 数据集是首个提供复杂场景下准确 3D 人体姿态的数据集。该数据集包含 60 个视频序列, 总计 51 000 多帧 (约 1 700 s, 30 Hz 采样率)。数据集记录了各种自然环境下的日常活动, 如购物、运动、拥抱、讨论、自拍、乘坐公交车、弹吉他和休息等。该数据集的主要特点是具有挑战性的拍摄环境, 包括移动摄像头、杂乱背景、遮挡和多人场景等。在 TotalCapture 数据集上的验证显示, 其平均每关节位置误差为 26 mm。Human3.6M 数据集包含 360 万帧人体动作数据, 由 11 位不同体型的演员在 4 个固定视角摄像头下完成 15 类基础动作。

该数据集提供了精确的 3D 动作捕捉标注, 采样频率为 50 Hz。

2.2 评价指标

MPJPE 是评估 3D 人体姿态估计准确性的重要指标。核心思想是计算预测的关节坐标与真实关节坐标之间的 L2 距离。具体公式为

$$\text{MPJPE} = \frac{1}{N} \sum_{j=1}^N (p_j - p_j^*)_2 \quad (29)$$

式中: N 为骨架中的关节数量; p_j 为第 j 个关节的真实坐标; p_j^* 为预测坐标。与 2D 姿态估计中使用像素作为单位不同, 3D 人体网格恢复的误差以毫米 (mm) 为单位。

经 Procrustes 对齐的 PA-MPJPE 是 MPJPE 的一个重要变体, 主要用于评估 3D 人体姿态估计和人体网格恢复的准确性。与 MPJPE 不同, PA-MPJPE 在计算误差之前会先进行 Procrustes 分析对齐, 这样可以消除因全局旋转和缩放带来的误差, 使评估更加关注姿态本身的准确性。

MPVPE 是一个用于评估人体网格重建的指标, 它通过计算预测网格点和真实网格点之间的 L2 距离来衡量人体网格恢复的精度。

2.3 实验结果分析

本文采用当前最先进的 CLIFF 作为基准 HMR 模型, 这一选择基于其在现有研究中展现的优异性能 and 广泛认可度。为了进行严格且全面的性能评估, 在两个最具代表性的大规模数据集上开展实验: 包括在真实场景下具有较高挑战性的 3DPW 数据集, 以及在人体姿态估计领域被广泛用作基准的 Human3.6M 数据集。通过与当前最先进的人体网格细化方法进行定量对比, 构建了一个系统的评估框架。实验结果如表 2 所示。

表 2 在 3DPW 和 Human3.6M 数据集上的 SOTA 比较

Table 2 Comparison of state-of-the-art methods on 3DPW and Human3.6M datasets

方法	3DPW			Human3.6M	
	PA-MPJPE	MPJPE	PVE	PA-MPJPE	MPJPE
CLIFFb*	43.0	69.0	81.2	—	—
CLIFFb ⁺	43.76	73.67	91.58	36.16	55.18
DynaBOA	40.4	65.5	82.0	—	—
Pose2Mesh	34.6	65.1	—	35.3	51.1
CLIFFb+CLIFFr*	32.8	52.8	61.5	—	—
CLIFFb+SMPLify	36.11	66.67	79.91	28.07	45.19
CLIFFb+CLIFFr ⁺	32.04	55.83	71.95	25.88	42.79
CLIFFb+KITRO	27.67	43.53	53.44	21.04	34.50
CLIFFb+PC-HMR(ours)	26.69	41.02	50.03	20.55	35.56

注: 结果中的⁺表示本文的复现结果; *表示在数据预处理中使用额外性别信息的不同 3DPW 评估协议。

从实验数据可以观察到,PC-HMR 在 3DPW 数据集上展现出显著的性能优势。特别值得注意的是,在关键评估指标 PA-MPJPE 上,PC-HMR 达到了 26.69 mm 的好成绩,这一结果与数据集发布时设定的约 26 mm 误差标准已经非常接近,体现了本方法的高精度和稳定性。同时,在 Human3.6M 数据集的评估中,PC-HMR 法也取得了优于现有方法的表现,进一步验证了所提出方法的有效性和泛化能力。这些实验结果充分证明了 PC-HMR 在人体网格恢复任务中的先进性。

2.4 消融实验

为量化各模块贡献,本文进行了系统的消融实验,评估了相机、形状及姿态参数在孤立优化与协同优化配置下的效果(见表 3)。结果明确显示,单一参数优化无法达到理想的细化效果,这凸显了本方法中实现 2D 关键点与 3D 人体网格精确对齐的必要性。进一步地,实验证实只有通过多参数协同优化才能实现稳健的人体姿态恢复,从而验证了本文所提出方法的有效性。

表 3 在 3DPW 数据集上的参数优化消融实验

Table 3 Ablation study for parameter optimization in 3DPW dataset

相机参数	形状参数	姿态参数	PA-MPJPE	MPJPE	PVE
×	×	×	43.76	73.67	91.58
√	×	×	43.76	73.67	91.58
×	√	×	43.35	68.93	81.46
×	×	√	45.92	87.35	100.74
√	√	×	43.47	78.17	92.83
√	×	√	35.01	69.03	84.26
×	√	√	27.27	43.20	53.28
√	√	√	26.69	41.02	50.03

3 结 论

针对当前 3D 人体网格恢复领域中存在的人体骨骼对应关系不一致的问题,本文提出了一种基于先验知识约束的网格恢复算法 PC-HMR。该算法创新性地将人体固有对称性特征和骨骼比例约束引入形状参数优化过程,并通过改进的姿态参数优化决策树选择机制,有效提升了算法的鲁棒性和准确性。实验结果表明,PC-HMR 算法在公开基准数据集上的综合性能显著优于现有方法,为 3D 人体网格恢复领域提供了参考。

参考文献:

[1] LIU Yang, QIU Changzhen, ZHANG Zhiyong. Deep learning for 3D human pose estimation and mesh re-

covery: A survey [J]. Neurocomputing, 2024, 596: 128049.

- [2] TIAN Yating, ZHANG Hongwen, LIU Yebin, et al. Recovering 3D human mesh from monocular images: A survey [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45 (12): 15406-15425.
- [3] CHO H, CHO Y, AHN J, et al. Implicit 3D human mesh recovery using consistency with pose and shape from unseen-view [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: [s.n.], 2023: 21148-21158.
- [4] KANAZAWA A, BLACK M J, JACOBS D W, et al. End-to-end recovery of human shape and pose [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: [s.n.], 2018: 7122-7131.
- [5] LIN Kevin, WANG Lijuan, LIU Zicheng. Mesh graphormer [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, Virtual Event, 2021. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers, 2021: 12939-12948.
- [6] LIN Kevin, WANG Lijuan, LIU Zicheng. End-to-end human pose and mesh reconstruction with transformers [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Virtual Event, 2021. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers, 2021: 1954-1963.
- [7] LOPER M, MAHMOOD N, ROMERO J, et al. SMPL: A skinned multi-person linear model [C]//Proceedings of Seminal Graphics Papers: Pushing the Boundaries, Volume 2. New York, NY, USA: Association for Computing Machinery, 2023: 851-866.
- [8] IONESCU C, PAPAVALA D, OLARU V, et al. Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 36(7): 1325-1339.
- [9] VON MARCARD T, HENSCHER R, BLACK M J, et al. Recovering accurate 3d human pose in the wild using imus and a moving camera [C]//Proceedings of the European Conference on Computer Vision. Munich, Germany: [s.n.], 2018: 601-617.
- [10] DOU Zhiyang, WU Qingxuan, LIN Cheng, et al. Tore: Token reduction for efficient human mesh recovery with transformer [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris, France: [s.n.], 2023: 15143-15155.
- [11] WANG Dongkai, ZHANG Shiliang. 3D human mesh recovery with sequentially global rotation estimation

- [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris, France: [s.n.], 2023: 14953-14962.
- [12] SHEN Xiaolong, YANG Zongxin, WANG Xiaohan, et al. Global-to-local modeling for video-based 3D human pose and shape estimation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: [s.n.], 2023: 8887-8896.
- [13] CHOI H, MOON G, CHANG J Y, et al. Beyond static features for temporally consistent 3D human pose and shape from a video[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Virtual Event, 2021. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers, 2021: 1964-1973.
- [14] WEI Wenli, LIN Jenchun, LIU Tyngluh, et al. Capturing humans in motion: Temporal-attentive 3D human pose and shape estimation from monocular video [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: [s.n.], 2022: 13211-13220.
- [15] LI Jiefeng, XU Chao, CHEN Zhicun, et al. Hybrik: A hybrid analytical-neural inverse kinematics solution for 3D human pose and shape estimation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Virtual Event, 2021. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers, 2021: 3383-3393.
- [16] BOGO F, KANAZAWA A, LASSNER C, et al. Keep it SMPL: Automatic estimation of 3d human pose and shape from a single image[C]//Proceedings of the European Conference on Computer Vision. Amsterdam, The Netherlands: [s.n.], 2016: 561-578.
- [17] KOLOTOUROS N, PAVLAKOS G, BLACK M J, et al. Learning to reconstruct 3D human pose and shape via model-fitting in the loop [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, South Korea: [s.n.], 2019: 2252-2261.
- [18] GUAN P, WEISS A, BALAN A O, et al. Estimating human shape and pose from a single image [C]//Proceedings of the IEEE International Conference on Computer Vision. Kyoto, Japan: [s.n.], 2009: 1381-1388.
- [19] JOOH, NEVEROVA N, VEDALDI A. Exemplar fine-tuning for 3D human model fitting towards in-the-wild 3D human pose estimation [C]//Proceedings of the 2021 International Conference on 3D Vision. Virtual Event, 2021. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers, 2021: 42-52.
- [20] LI Zhihao, LIU Jianzhuang, ZHANG Zhensong, et al. Cliff: Carrying location information in full frames into human pose and shape estimation [C]//Proceedings of the European Conference on Computer Vision. Tel Aviv, Israel: [s.n.], 2022: 590-606.
- [21] YANG Fengyuan, GU Kerui, YAO A. KITRO: Refining human mesh by 2D clues and kinematic-tree rotation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: [s.n.], 2024: 1052-1061.
- [22] BOGIN B, VARELA-SILVA M I. Leg length, body proportion, and health: A review with a note on beauty [J]. International journal of environmental research and public health, 2010, 7(3): 1047-1075.
- [23] WANG Nan, MA Jie, JIN Dan, et al. A special golden curve in human upper limbs' length proportion: A functional partition which is different from anatomy [J]. BioMed Research International, 2017: Article No.4158561.

(编辑:刘彦东)