

一种基于三维骨架化的人体姿势估计方法*

刘宏志^{1,2}, 吴中海²⁺, 李超², 王洋¹

¹(北京大学 信息科学技术学院,北京 100871)

²(北京大学 软件与微电子学院,北京 102600)

A 3D 2-Subiteration Thinning Algorithm for Human Pose Estimation

LIU Hong-Zhi^{1,2}, WU Zhong-Hai²⁺, LI Cao², WANG Yang¹

¹(School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China)

²(School of Software and Microelectronics, Peking University, Beijing 102600, China)

+ Corresponding author: E-mail: wuzh@pku.edu.cn

Liu HZ, Wu ZH, Li C, Wang Y. A 3D 2-subiteration thinning algorithm for human pose estimation. *Journal of Software*, 2010,21(Suppl.):158-164. <http://www.jos.org.cn/1000-9825/10017.htm>

Abstract: Markerless human body model reconstruction has been a hot research topic for decades, and it has many applications in human computer interaction (HCI), medical rehabilitation, sport training, and virtual human synthesis for games, cartoons and digital films. This paper presents a markerless 3D human body skeleton model reconstruction method. It takes input from multi-view video, and computes the visual hull of human body through shape-from-silhouette. It then extracts the 3D skeleton by a new hybrid 2-subiteration thinning algorithm, and registers the skeleton to a 3D human body skeleton model by a heuristic method. The experiment results show that although the intermediate results of the shape-from-silhouettes process and 3D thinning process contain much noise, the model registration method, which bases on a set of simple prior rules, can still get accurate 3D human body skeleton model.

Key words: 3D reconstruction; 3D skeletonization; pose estimation; multi-view vision; heuristic method

摘要: 提出了一种基于三维 2 步子迭代混合细化算法和启发式搜索的无标记三维人体姿势估计方法.其主要思想是首先通过一种新的 2 步子迭代混合细化算法,对由 Shape-from-Silhouette(SFS)算法从多幅二维图像中重建出的人体三维模型进行骨架化,然后通过一种启发式的方法进行模型匹配.实验结果表明,该方法不仅思想简单,而且具有较好的鲁棒性,能够很好的去除各种噪声的影响.

关键词: 三维重建;三维骨架化;姿态估计;多视角视觉;启发式方法

基于视觉的人体建模(Vision-based human body modeling)一直是计算机视觉和计算机图形学两个领域所共同研究的一个热点问题.它不仅可以用于对人体行为的分析,例如:通过识别人的手势或是动作来完成人机间的交互、通过分析重建的人体的三维人体运动数据辅助体育训练或是病人康复治疗;还可用于游戏、动画以及数字电影中虚拟人的设计^[1].

人体模型可以分为多层表示,比较常见的是 3 层表示法:骨架模型、肌肉模型和皮肤模型,其中骨架模型确定了人体的基本姿态,肌肉模型确定人体各部位的形变,而皮肤模型则确定了人的显示外观.有时会额外再加一层——服饰模型,用以表示人的头发、衣饰等装饰物品^[1].本文将主要讨论有关基于视觉的无标记三维人体姿势

* Received 2010-07-20; Accepted 2010-11-03

估计——骨架模型自动重建的问题。

目前已经有很多相关的研究成果.Villa Uriol^[2]采用切片法来提取人体骨架模型:在重建出人体的体模型(volumetric model)之后,对体模型分别从 x,y,z 三个方向进行切片操作,然后对每个切片单独进行分析,提取其中的 Blobs,并用椭圆表示每个 Blob,再通过对所有切片中椭圆的中心点进行分析,以获得整个人体的三维骨架.该算法只能粗略估算出胸和盆骨两个关节点,以及头、左右手和左右脚 5 个端点,而无法获得其它的中间关节点.胡晓雁等人^[3]也是利用这种切片来定位标准人体模型中各个关节点的位置.Ménier等人^[4]则是先采用文献[5]中的方法来提取出三维骨架,然后通过最大化后验概率(MAP)来进行模型匹配.由于他们采用了 EM 算法来进行模型匹配,该方法的运算量较大.Chu 和 Ramakant^[6]首先提取各参考图像中二维中心轴,然后通过多视角几何(multiple view geometry)计算出三维的中心轴.该方法无法避免自遮挡问题,并且由于骨架并不具有投影不变性,即三维骨架并不一定会投影到二维骨架上,所以该算法的结果也不一定准确.Vlasic等人^[7]和 Gall等人^[8]则都是采用基于能量最小的优化算法搜索各关节点在重建后的三维人体模型中的位置,这种方法虽然准确高,但是其需要手工调整^[7]或是计算量很大^[8].

本文给出了一种基于三维 2 步子迭代混合细化算法和启发式搜索的三维人体骨架模型自动重建方法,并详细介绍了其中三维骨架化和模型匹配等关键模块的设计,最后给出实验测试结果并对结果进行了分析.

1 系统框架

本文采用一种基于多视角视觉分析的方法来重建三维人体骨架模型(如图 1 所示):从多个摄像头对同一场景进行图像采集(图 2(a)),并分别对各路图像进行处理,通过简单的背景相减法提取各图像中人体的轮廓(图 2(b)),然后采用基于 Shape-from-Silhouettes 的三维重建算法^[9]重建出人体的体模型(图 2(c)),接着对所得的体模型进行三维细化操作从而获得三维骨架(图 2(d)),最后通过一个启发式搜索方法进行模型匹配从而获取人体的骨架模型(图 2(e)).

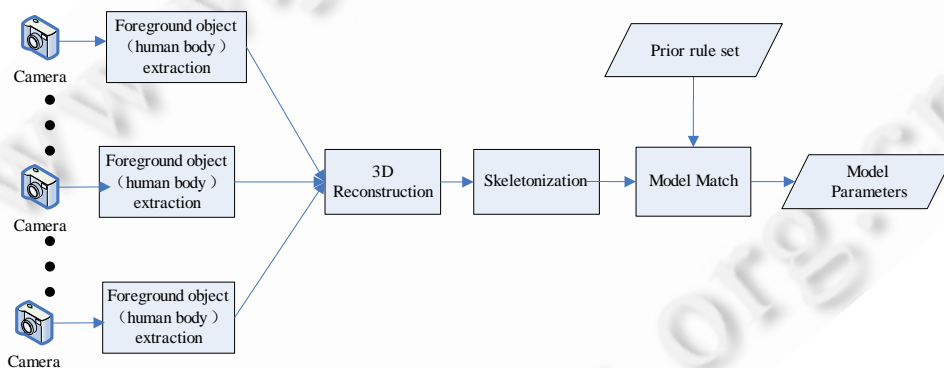


Fig.1 Pipeline of system

图 1 系统框架图

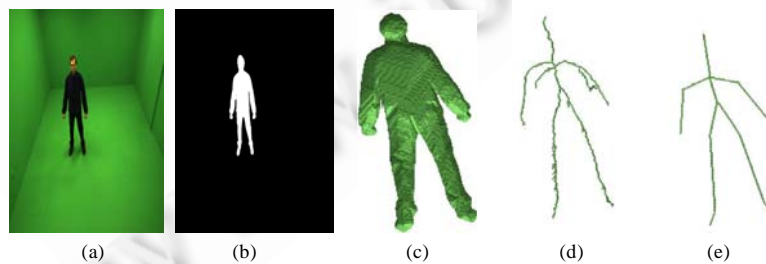


Fig.2 Dataflow of system

图 2 系统数据图

对于系统中的两个主要模块:骨架化和模型匹配,下面两节将分别进行详细讨论和说明.

2 三维骨架化

骨架化(Skeletonization)又称中轴变换(Medial Axis Transform),可以通过多种方法来实现,如:基于迭代细化的方法(Thinning)、基于距离变换的方法、基于 Voronoi 的方法和基于神经网络的方法.在这些方法中,最为常见的是迭代细化法,通过不断的细化物体,即删除表面上的体元,直到不能细化为止,即只剩下骨架上的体元.

在进行迭代细化时,最主要的一点是要保证在细化过程中不能破坏物体原有的拓扑结构.为了保证物体的拓扑结构,在细化过程中我们只能删除简单点(Simple Point).三维简单点必须满足两个条件^[10,11]:

- 1) 在该点的 26 邻域中只有一个前景(物体)连通体,
- 2) 在该点的 26 邻域中只有一个背景连通体.

三维空间中的 26 邻域是指到当前点的几何距离不大于 $\sqrt{3}$ 的点的集合,如图 3 所示.

按照是否可以平行处理每个体元,迭代细化算法可以分为三类:串行迭代细化算法、并行迭代细化算法和混合迭代细化算法.在串行迭代细化算法中,每次迭代只删除一个体元,这类方法能够很好的保证物体的拓扑结构.在并行迭代细化算法中,由于每次迭代可以删除多个体元,理论上很难保证物体的拓扑结构,目前的大部分方法都是根据经验设计多个模板,在对每个体元进行处理时,根据模板匹配结果来决定是否删除该体元.而混合迭代细化算法则是先对各个体元进行平行标记,标记出可能会被删除的候选体元,然后再对这些候选体元进行串行删除.

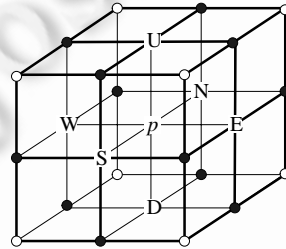


Fig.3 Neighborhood of 3D space

图 3 三维空间邻接图, p 为当前点,U,D,W,S,E,N 为其 6 邻域点

Palágyi 等人^[12]提出一种 6 步迭代混合算法,在每次迭代中,分别对上(Up)、下(Down)、前(South)、后(North)、左(West)、右(East)这 6 个方向上的边界点进行处理,如图 3 所示.在每次子迭代中,先对该方向上的所有简单点并行进行标记,然后对标记的点串行进行删除,为了保证物体的拓扑结构,在要删除一个点之前,需要再次判断其是否是简单点.

在借鉴了经典二维细化算法^[13]的核心思想的基础上,本文对 Palágyi 等人^[12]的三维 6 步迭代混合细化算法进行了改进,将上、左和前三个方向,以及下、右和后三个方向分别合并在一起进行处理,从而得到了一种 2 步迭代混合细化算法,如图 4.其中边界点(Border Point)是指其 6 邻域邻近点(几何距离为 1 的邻接点)中至少有一个点为背景点的前景点,而端点(End Point)是指其 26 邻域邻近点中只有一个为前景点的前景点.

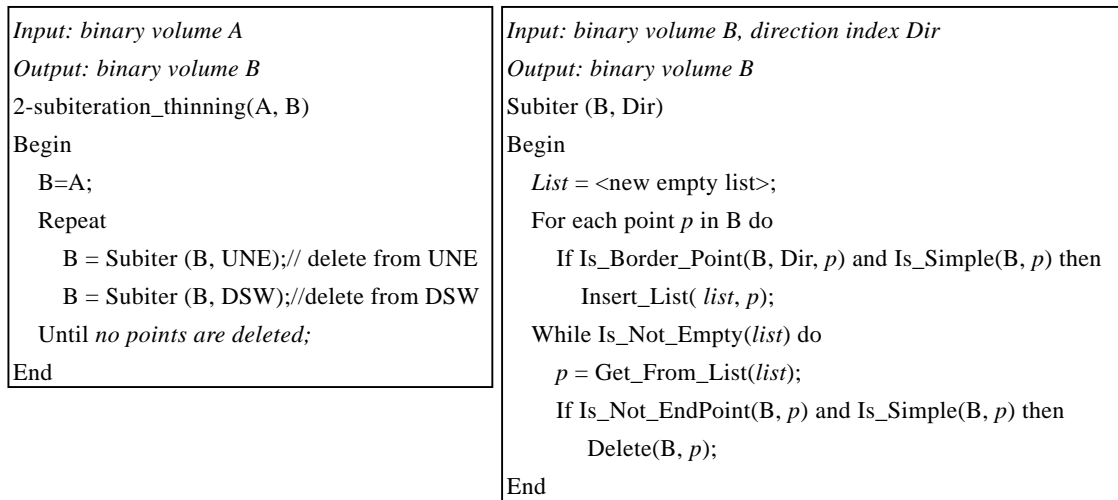


Fig.4 Algorithms of 3D thinning

图 4 三维细化算法

3 模型匹配

3.1 模型描述

我们采用如图 5 所以的人体模型,将人体分为 17 个部分,并用 18 个关键点的位置来进行表示,其中包括 13 个关节点:颈、胸、盆骨、左右臀、左右膝关节、左右踝关节、左右肩关节、左右肘关节;和 5 个端点:头、左右手以及左右脚。

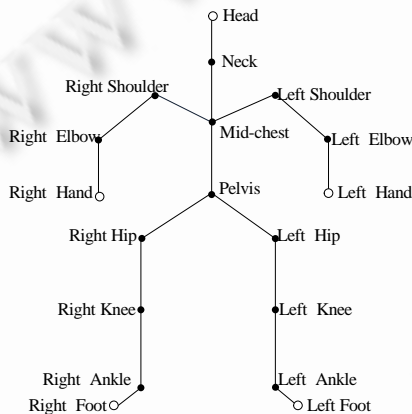


Fig.5 Skeleton model

图 5 人体骨架模型

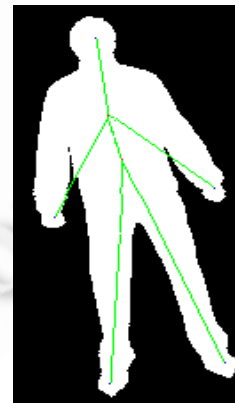


Fig.6 Intermediate result of model matching

图 6 模型匹配中间结果

3.2 模型提取

从图 2d 可以看出,由于各种噪声和衣着等因素的影响,在骨架化后的三维骨架含有很多错误的分支,有的甚至形成环路;还有一些错误的节点,例如图中的胸节点处,左右两个手臂的分支并没有连接到同一节点上。

为了去除以上影响因素,并提取出正确的三维人体骨架模型,我们首先采用一种迭代模型匹配方法(规则 1-5),逐步简化骨架模型,最后只保留胸节点和盆骨两个节点,以及头、左右脚和左右手,5 个端点,如图 6 所示。

为了提取出中间部分的关节点,包括颈关节、肩关节、肘关节、臀关节、膝关节和踝关节.先估算出中间

各骨架点的曲率,并找出各(曲率)极值点作为候选关节点(规则 6).为了避免噪声的影响,在计算曲率时,对骨架点进行重采样,以去除局部极值点.

规则 1:根据一定的阈值,去除过短的分支和相应的端点(End Point).

规则 2:根据一定的阈值,合并距离过近的邻近节点(Joint Point).

规则 3:断开环路:若环路经过两个或多个节点,去除多余的分支;若环路只经过一个节点,则创建一个新的端点.

规则 4:通过球形匹配来确定头(端点)^[14],并假设头(端点)直接与颈关节点相连.

规则 5:脚(端点)是与头(端点)距离最远的端点,并假设脚(端点)直接与盆骨节点相连.

规则 6:根据曲率来提取中间的骨架节点.

Fig.7 Prior rule set

图 7 先验规则集

4 实验结果

本文采用 Popovic 等人^[7]提供的多视角视频数据集 March 作为测试数据,该测试数据是通过 8 路已经标定好的摄像机对同一室内场景同时进行采集而得,每路视频的分辨率为 1600×1200.

4.1 参数设置

实验中主要有 3 个重要参数:三维重建中采样体的大小,以及规则 1 和规则 2 中用于滤噪的两个阈值.在实验中,两个阈值都设为几何距离不超过 8.关于采样体的大小,则根据下面的方法进行确定.

由于只需要重建场景中人体的而非整个场景的三维模型,所以在进行三维人体重建之前,需要先确定一个包含人体并且尽可能小的包围盒.先从各个参考图像中找出包括人体轮廓的矩形包围框,然后通过反投影法^[15],将各个矩形框反投到三维空间中(每一个构成一个四棱台),并求其交集,从而估算出一个包含人体长方体包围盒.

采样体是指对包围盒进行空间采样后所获得的一个离散体,其基本元素是体元(voxel).采样体的大小不仅决定了重建后三维模型的精度,同时也是影响系统运算量的一个重要因素.为了在不影响系统精度的条件下,降低系统的运算量,我们采用的是一种自适应的采样大小调整算法,根据估计的三维包围盒的大小,动态调整采样大小.

其基本思想为三向同密度采样,采样密度由最长边决定.根据观察,包围人体的三维盒通常并不是一个正方体,而是长方体,而且其垂直地面的边(包围盒的高)一般是最长的边.在实验中,我们将采样体的高固定为 256,对于采样体的长和宽,我们采用自适应算法进行调整,具体值根据包围盒的长(和宽)与高的比例来确定.

4.2 实验结果及分析

图 8 中的三行图像分别是第 1 帧、第 100 帧和第 220 帧时刻的实验结果.第 1 列为三维重建所得的可视化外壳,第 2 列为骨架化所得的三维骨架,第 3 列和第 4 列分别为从两个不同角度观察到的人体骨架模型.

从重建的可视化外壳以及骨架化后的骨架可以看出,由于噪声等因素的影响,所得的三维骨架中存在很多错误的分支以及一些环路,但这些都可以通过我们所设计的简单规则去除掉.从最后所得的三维人体骨架模型可以看出,此方法可以较为准确的获得胸节点和盆骨这两个关键节点,以及头、左右手和左右脚这 5 个端点.但是对于颈、肩、肘、臀以及膝等中间关节点,只有在由这个关节所连接的两个部分有一定的角度时我们才能通过曲率判断获得其位置.

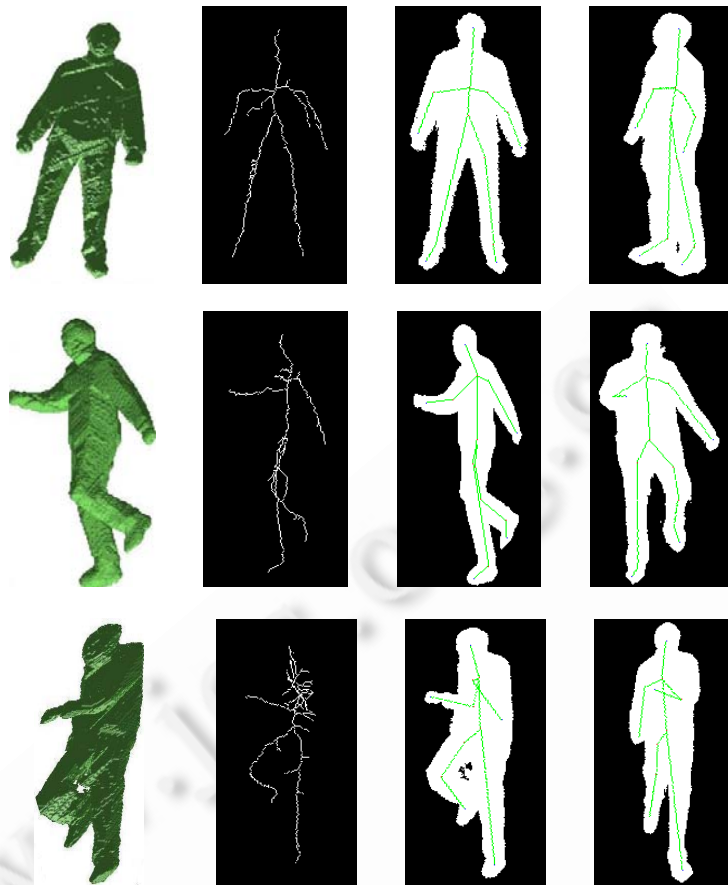


Fig.8 Experimental result

图 8 实验结果

在配有 2.3G AMD Quad-Core CPU 和 3G 内存的 PC 上,我们的姿态估计算法处理一帧图像平均需要 3.9 秒.和文献[7,8]中的方法相比(见表 1),我们的方法虽然在精度上要差一些,但是具有速度快、无需人工参与和实现简单等优点.

Table 1 Comparison of experiment results

表 1 实验结果比较

方法	速度(秒/帧)	手工调整	帧间连续性	精度
Gall 等人的方法 ^[8]	>200	×	√	高
Vlasic 等人的方法 ^[7]	7.6	√	√	高
我们的方法	3.9	×	×	中

5 结论及以后工作

本文提出了一种新的三维 2 步子迭代混合细化算法,并在此基础上提出了基于简单规则集的三维人体骨架模型重建算法.该算法具有思想简单、鲁棒性强的特点,只采用 6 条简单的先验规则便能很好的去除各种噪声的影响,并提取出人体的三维模型.

本文在重建人体骨架模型时只考虑了三维空间信息,重建的是静态的骨架模型,即在每一时刻,只考虑当前时刻从各视角获得采集图像,没有考虑到视频数据在时间上连续性.下一步我们打算利用时间上的连续性,对此方法进行进一步的改进,并将其应用到人体运动捕捉系统中.

References:

- [1] Xia SH, Wang ZQ. Study on synthesis of virtual human. Science in China (Series F: Information Sciences), 2009,39(5):483-498.
- [2] Uriol V, Cruz M. Video-Based avatar reconstruction and motion capture [Ph.D. Thesis]. Irvine: University of California, 2005.
- [3] Hu XY, Liang XH, Zhao QP. Match the virtual human model and motion data automatically. Journal of Software, 2006,17(10): 2181-2191 (in Chinese with English abstract)
- [4] Ménier C, Boyer E, Raffin B. 3D Skeleton-Based Body Pose Recovery. In: Proc. of the 3rd Int'l Symp. on 3D Data Processing, Visualization and Transmission, (3DPVT 2006). 2006. 389-396.
- [5] Attali D, Montanvert A. Modeling noise for a better simplification of skeletons. In: Proc. of the ICIP. Lausanne (Switzerland). 1996. 13-16.
- [6] Chu CW, Ramakant N. Real-Time 3D Body Pose Tracking from Multiple 2D Images. Articulated Motion and Deformable Objects, LNCS. 2008. 42-52.
- [7] Vlasic D, Baran I, Matusik W, Popovic J. Articulated mesh animation from multi-view silhouettes. ACM Trans. on Graph, 2008, 27(3):97:1-97:9.
- [8] Gall J, Stoll C, de Aguiar E, Theobalt C, Rosenhahn B, Seidel H. Motion capture using joint skeleton tracking and surface estimation. In: Proc. of the CVPR 2009. 2009. 1746-1753.
- [9] Laurentini A. The visual hull concept for silhouette-based image understanding, IEEE Trans. on PAMI, 1994,16(2):150-162.
- [10] Saha PK, Chaudhuri BB. Detection of 3-D simple points for topology preserving transformations with application to thinning. IEEE Trans. on PAMI, 1994,16(10):1028-1032.
- [11] Teng QZ, Kang X, Tang T, He XH. Parallel 3D skeletonization algorithm based on sort ascending rechecking. Optics and Precision Engineering, 2009,17(10):2528-2534(in Chinese with English abstract).
- [12] Palágyi K, Sorantin E, Balogh E, Kuba A, Halmai CS, Erdöhelyi B, Hausegger K. A sequential 3D thinning algorithm and its medical applications. In: Proc. of the 17th Int'l Conf. on Information Processing in Medical Imaging (IPMI 2001). LNCS, 2001. 409-415.
- [13] Zhang TY, Suen CY. A fast parallel algorithm for thinning digital patterns. Comm. ACM. 1984,27(3):236-239.
- [14] Michoud B, Guillou E, Briceno H, Bouakaz S. Real-Time marker-free motion capture from multiple cameras. In: Proc. on 11th Int'l Conf. on Computer Vision ICCV 2007. Rio de Janeiro, 2007. 1-7.
- [15] Hartley R, Zisserman A. Multiple View Geometry in Computer Vision. Cambridge University Press, 2003.

附中文参考文献:

- [3] 胡晓雁,梁晓辉,赵沁平.自动匹配虚拟人模型与运动数据.软件学报,2006,17(10):2181-2191.
- [11] 滕奇志,康瑕,唐棠,何小海.基于升序复核的并行三维图像骨架化算法.光学精密工程,2009,17(10):2528-2534.



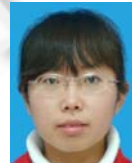
刘宏志(1982-),男,湖北武汉人,博士生,主要研究领域为三维重建,图像分析.



李超(1988-),男,硕士生,主要研究领域为三维重建,真实感图像绘制.



吴中海(1968-),男,教授,博士生导师,主要研究领域为数字艺术技术,嵌入式软件与系统,情境感知服务.



王洋(1984-),女,博士生,主要研究领域为数字艺术技术,情境感知服务.