三维人体重建综述

孙庆平　　S191000913

摘　要：三维人体重建是指在虚拟场景中利用计算机创建人的虚拟模型，近年来已成为计算机视觉和计算机图形学研究热点问题。利用多样的输入方式，不同的技术路线，很多已有研究成果取得了很好的效果。其中，利用深度神经网络，从RGB 图片或 RGB-D 图片重建人体是常用方法。本文首先阐明三维人体重建研究的内容和难点，接着研究了目前效果较好的重建方法，分析各种方法的原理及其优缺点，最后总结三维人体重建的研究趋势。

关键词： 三维人体重建；深度神经网络；RGB 图片；RGB-D 图片

1、前言

三维人体重建在游戏影视、虚拟试衣、安防监控等现实场景中有重要研究价值。总的来说，重建旨在利用采集数据来推断目标人体的状态信息。图片，深度图，点云往往作为原始输入数据，以单视图序列或多视图提供更多的三维信息来合理推断人体信息。而应用场景的需求不同也决定最后输出的数据形式各异，如图，输出分为精确的人体表面拓扑结构和模板型光滑的人体表面结构两大类。不同的输入输出形式同时也影响这算法选择的多样性，三维人体动态重建从技术角度上可以分为三大类：基于多视立体的方法、基于三维模板的方法、表面动态融合方法。

三维人体动态重建的难点在于重建的精度和速度，复杂场景、多人体环境、人体遮挡、实时性要求是重建任务要突破的障碍。当前很多成果能够利用显卡的计算能力实时地取得较高的重建精度，但在应用场景和硬件条件等问题上还是存在诸多限制。三维信息的完整性和更加复杂的神经网络结构成为研究的重点。

2、三维人体重建算法

2.1 基于多视立体的方法

基于多视立体的方法是通过多视点的图像，每一帧独立重建，通过传统的立体匹配，帧与帧之间的结果没有相关性。重建没有人体先验的假设，所以支持动物重建以及人与很多物体的交互，重建结果的拓扑结构是任意的。这类方法优点是支持单时刻多视点三维重建，支持任意拓扑结果，最新研究成果重建质量较高。但是实时性差，难以保证时间的连续性。

[1]开创了虚拟现实的 Free Viewpoint Video (FVV) 研究，使用一个圆顶摄像机来计算多基线立体，并通过三角合并深度图生成新的视图。[2]计算密集的深度图，将它们分割成纹理层，然后扭曲它们以渲染虚拟视点。[3]计算立体点的三维 Delaunay 三角剖分，并通过标记 Delaunay 四面体为空或被占用来重建曲面。

[4]使用同步和校准的摄像头捕捉一个或多个执行者。对于每个输入图像，通过减去背景来提取前景轮廓。参照论文[5]所述，研究的目标是估计骨骼结构（姿势），包括躯干的整体刚性转换和骨骼的关节角度，以及不能由骨骼驱动的变形表示的非关节表面变形（形状）。

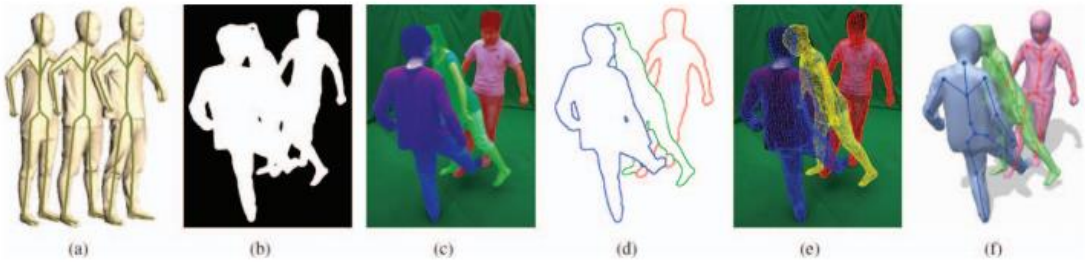

图 1　[4]中基于多视角图片分割的人体重建方法

重建的过程如图 1 所示。从上一帧中所有人的估计姿势和形状开始，该算法基于捕获的多视图图像和前景轮廓估计当前帧中的姿势和形状（图 1b）。由于未知姿态和形状参数空间对于多人来说变得非常大，所以该方法将跟踪问题分解为多视图二维分割问题（图 1c 和 1d）以及三维姿态和形状估计问题（图 1E 和 1F）。分割通过为每个前景像素分配一个标签来分隔图像域中的人。然后，根据标记的像素，分别估计每个人的姿势和形状。

图 2 [6]中捕获采集数据的绿幕场景

[6]对视频重建提出了一个端对端的方法，如图 3 所示。为了能够获得高质量输入图片，作者使用了配备高速 RGB 和红外摄像机的绿幕场景，如图 2 所示。为了校准场景，捕获以下信息：背景图像以帮助进行背景分割；校准对象的图像以计算相机参数；以及颜色校准图像以规范化相机之间的像素响应以实现一致的纹理。然后对图像进行预处理，以纠正偏差并分割背景。首先处理来自立体对和轮廓数据的图像，以生成密集的深度图。然后使用多模多视图立体算法合并深度图，并通过局部拟合人体表面来细化得到的点云。使用这个点云，可以使用轮廓约束的泊松曲面重建创建一个封闭的网格结构。这个人体网格结构可能会很粗糙，通过使用拓扑降噪和孤点去除算法来清理人体分封闭结构以外的物体。根据观察角度和表面特性，作者通过混合可见摄像头的颜色来检测和计算网格中的表面颜色。作者处理视频时自动选择合适的关键帧，并在帧序列上跟踪它们，以生成时间一致的子序列。然后确定哪些区域在感知上很重要，并自适应地决定和展开网格，保留重要的几何和纹理细节。作者利用网格跟踪来计算一个时间相关的网格展开和纹理图谱。

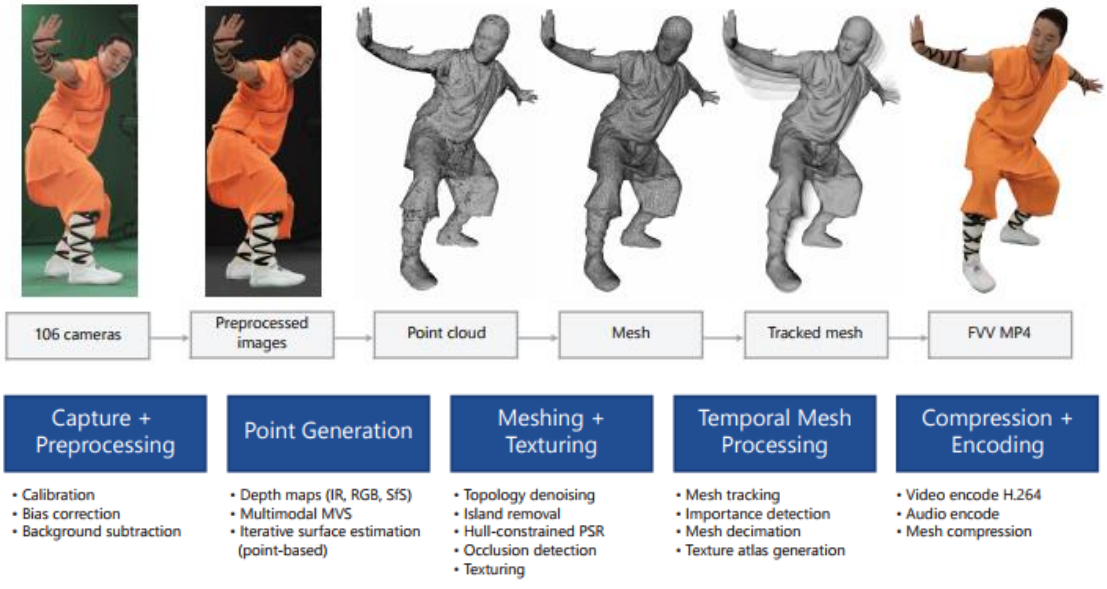

图 3 [6]中基于高分辨率视频流的重建方法

2.2 基于三维模板的方法

基于三维模板的方法可以分为基于骨架嵌入和非刚性形变的方法和基于统计模板的方法。骨架嵌入的方法可以看做非刚性形变特例的情况。通过对人体进行先验，人体的状态总的来说是有骨架来确定的，我们可以通过骨架来驱动人体的形变。这样处理可以减低形变的参数空间，几十个参数就可以确定一个人体的姿势。非刚性形变则是通过对对象的表面进行离散关键点的采样，传统的方法处理人体一般会定义一个人体表面模型，模型结构为大量定点和面的集合。通过均匀的离散采样，我们将表面定点的变量降低到几百个，通过几百个定点来驱动密集定点的形变。非刚性形变的方法和骨架嵌入是类似的思想，通过减低表面变量的个数，确定一个信息确定下的模型。这些方法一般会利用时域的信息，通过帧与帧联系确定更加精确的结果。

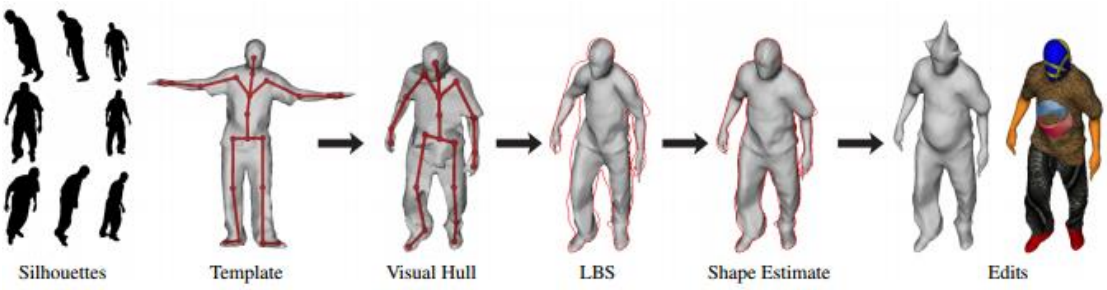

图 4 [7]中基于骨骼嵌入的重建方法

[7]多视角重建通过使用多个校准摄像头从多个角度记录，提供一组同步的高清晰度剪影视频。如图 4 所示，使用一个模板网格，该模板网格装配有一个与执行者的物理尺寸相匹配的骨架：骨架位于模板网格内，每个顶点都被指定一个权重，该权重用于用线性混合蒙皮（LBS）变形模板。最后结果输出一系列关节状态和顶点位置，以表示多视图序列每帧中人体的姿势和形状。因为该方法只依赖于轮廓，所以它完全不受颜色噪声、照明和颜色校准问题的影响。然而，这种依赖导致了两个限制。首先，法不能精确地复制远离轮廓的表面：它必须依靠模板来插入几何信息。这对于没有关节的物体尤其有问题，如长围巾、脸或飘逸的头发。以不牺牲稳健性的方式使用颜色信息将改进其重建。其次，对轮廓中的错误很敏感，并且在视觉外壳有噪音时会产生不正确的几何图形。

[8]提出了一种非刚性变形任意无标记形状物体实时重建的方案。他们的系统使用一个独立的立体摄像机单元，以 30 赫兹的频率生成时空相干的 3D 模型。非刚性重建过程分为三个部分：第一，刚性配准将模板与输入数据大致对齐；第二，通过最小化拟合能量进行非刚性表面拟合，优化过程结合了密集的几何和光度模型数据约束，以及 as-rigid-as-possible (ARAP)正则化器几个方面的约束。优

化求解过程则是利用预处理共轭梯度法（PCG），基于 GPU 的高斯-牛顿解算器将能量最小化；第三，在最精细的模板层次上进行细节集成：利用线性最小二乘，针对每个顶点模型法向位移，将模型到数据约束下的人体表面变形能量最小化。

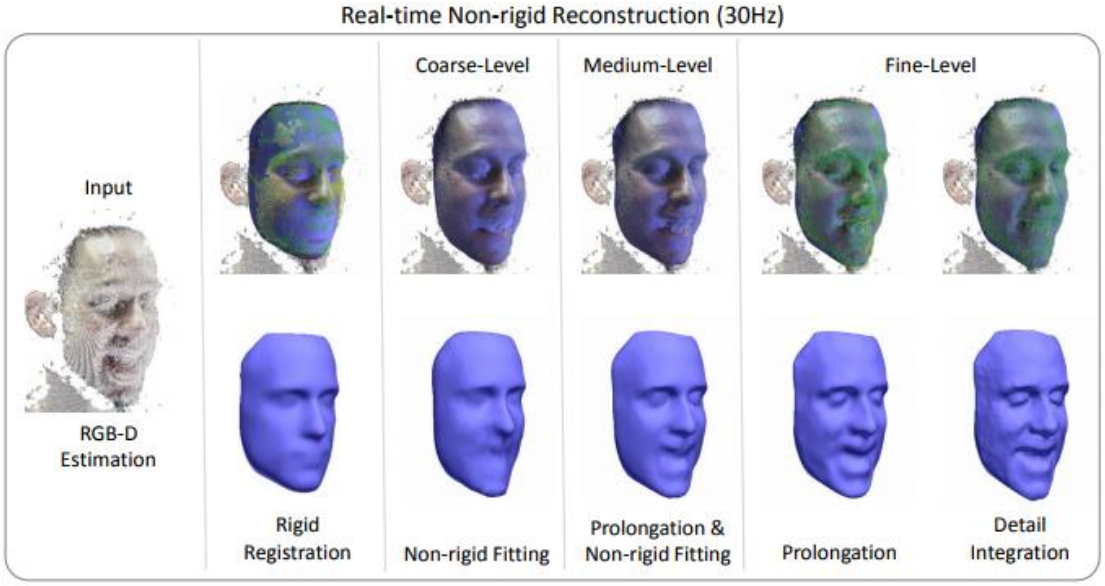

图 5 [8]中基于非刚性形变的重建方法

基于基于骨架嵌入和非刚性形变的方法在基于多视角立体方法的基础上，可以降低求解空间，重建时间可以大幅度减小。同时可以添加时间约束，获取时间连续性求解，在视频重建方面性能更趋于稳定。但是，在重建处理之前，需要在扫描模板进行骨架嵌入，计算人体表面和骨架之间的关系。所以大量的研究工作利用大量的数据集合，利用统计信息进行人体三维重建，这类方法称为基于统计模板的重建方法。

[9]把人体的重建看做是体态重建和姿势重建两个子任务，重建过程如图 6 所示。体态包括人体的高矮胖瘦等信息，姿势则是指人体的躯体的运动状态。SCAPE 子空间模型是一个参数化人体模型，该模型单独考虑了不同人体之间的形态差异和同一个人体的姿势差异。形态驱动的变形可以参数化表示为$S = \overline{U\beta + \mu}$,$\mu$是平均的人体形态，$U$是通过主成分（PCA）分析出的特征向量，$\overline{U\beta + \mu}$是向量$\overline{U\beta + \mu}$的矩阵表示。$\mu$和$U$可以从人体数据库中直接计算。参数化向量$U$代表了一个指定的人体。姿势引导的变形可以用刚性变换集合{R}和非刚性变换集合{Q}表示，人体是一个关节模型，可视为几个刚性部分（胳膊、躯干、腿等）的组合，R 定义了某个刚性部分的变换，Q 则定义了刚性部分的连接区域的变换，一般为某些刚性部分 R 的组合，R 和 Q 决定了人体的姿势。用θ表示某一个姿势，其包含的参数为刚性变换集合{R}（{Q}可看作由{R}生成）。

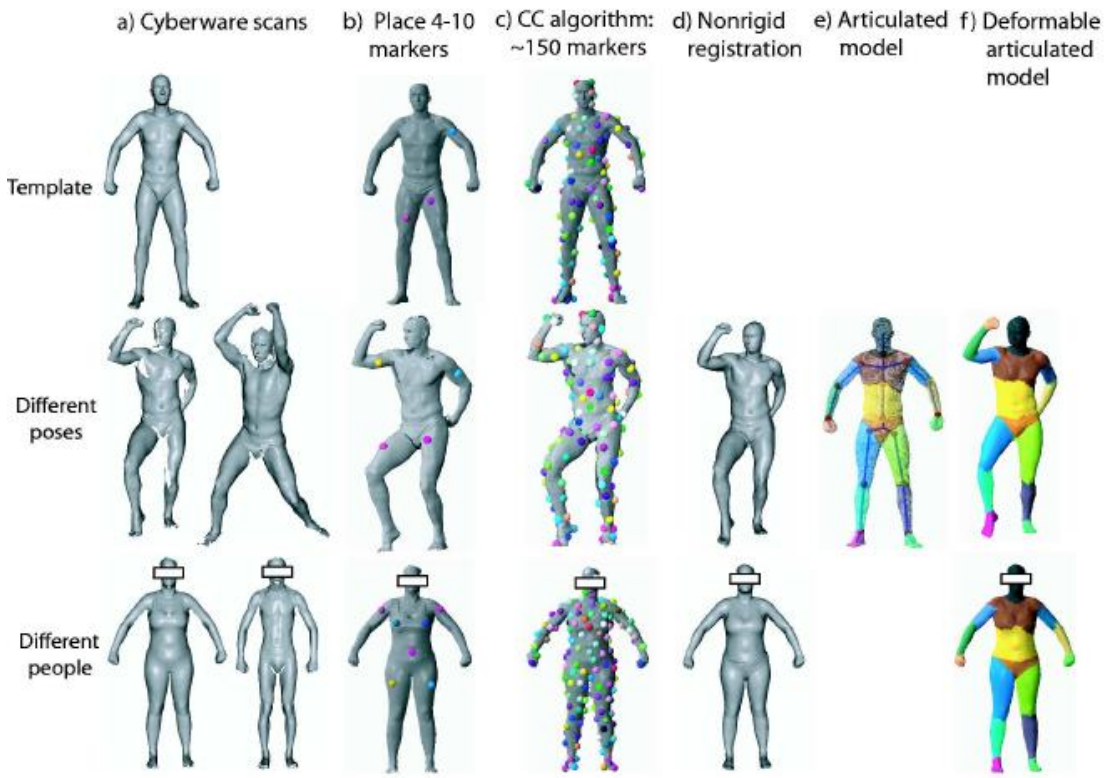

图 6　SCAPE 模型

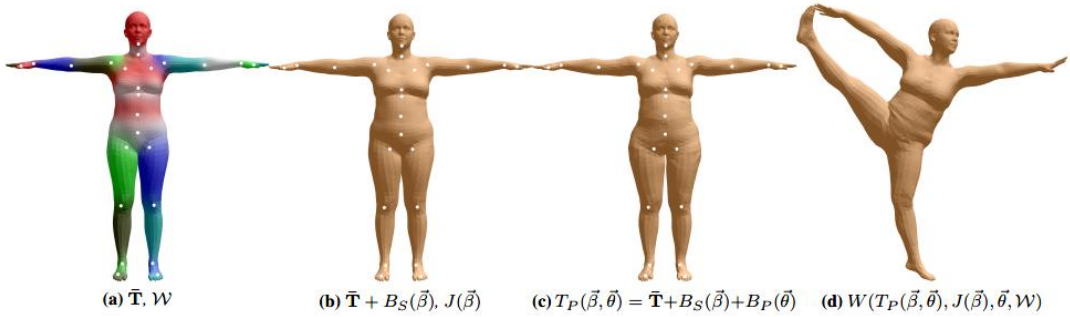

图 7　SMPL 模型

[10]在传统的 SCAPE 模型上进行的改进。在体态变形方面原理和 SCAPE 类似，但是在姿势变形上面，[10]的 SMPL 模型把线性混合蒙皮技术应用到姿势求解过程中。给定一个人体模板$\bar{T}$和蒙皮参数W进行重建。具体地，SMPL 模型把人体运动结构看做一个 23 个关节的树形结构，每个关节的旋转可以由一个3×3的旋转矩阵进行表达，每个旋转矩阵有三个变量。这样，人体的姿势完全可以由极少的 72 个参数进行表达，节省了大量重建时间。同时，如图所示，SMPL 模型还引入了体态影响姿势参数$B_P(\theta)$，考虑到在人体运动过程中，姿势的变化会对人体体态产生影响，这样对序列重建中人体软组织的模拟有很好的效果。在[54]、[55]以及[56]中，都在原本 SMPL 模型的基础上增加人手、人脸重建。

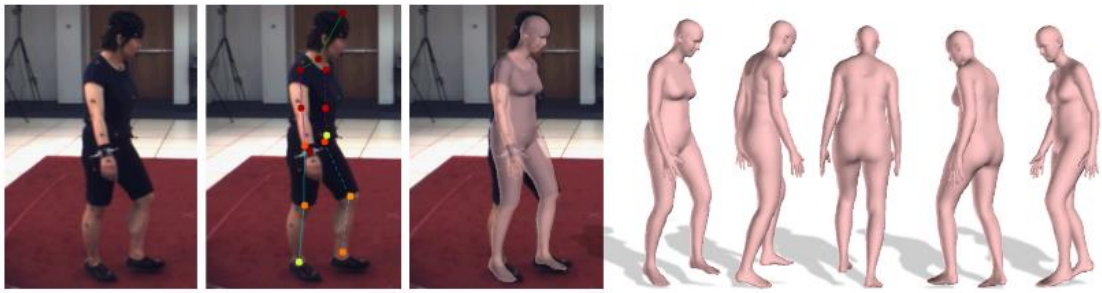

图 8　SMPLify 模型

[11]在 SMPL 模型的基础上又进行的大量的工作研究，提出了一个 SMPLify 模型。图 8 显示了 SMPLify 的重建过程。首先取一张输入图片，使用 DeepCut CNN[36]来预测二维身体关节J_{est}。对于每个二维关节，CNN 提供了一个置信值W_i。然后拟合一个三维实体模型，使模型的投影关节最小化一个鲁棒加权误差项。

[12]通过从最大的商用扫描人体数据库中重建了一个广泛使用的统计人体表达。由于预处理数千次扫描来学习模型本身就是一个挑战，因此作者着重研究扫描对齐，从而定量地生成最佳学习模型 。除了考虑顶点拟合误差，作者还加入了旋转矩阵的平滑项和人体标记约束，达到更好的重建效果。

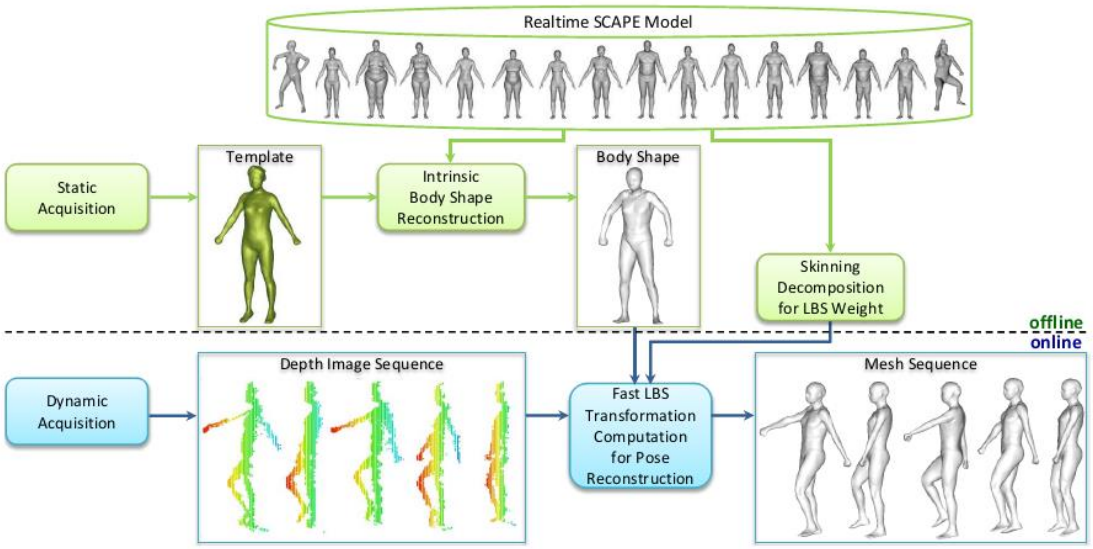

图 9 liveSCAPE 模型结构

[13]同样是在传统 SCAPE 模型基础上进行改进。首先应用深度相机对人体旋转扫描一周进行静态的人体重建，得到静态的人体模型后，根据序列的每一帧深度图，静态的模型动态地拟合深度图，得到最终结果。作者为了得到更加精确的结果，在拟合过程中设置了四个约束条件：重建的网格和深度数据的拟合程度；相邻刚性块之间应该具有共同的关节点；刚性块的运动惯性；某些刚性块应该只围绕特定的轴（该轴称为主轴）。

[14]利用 SMPLlify 训练 91 个关键点探测器，其中一些探测器对应于传统的身体关节，另一些则对应于身体表面的位置。然后，他们优化 SMPL 模型参数，

以与[11]相似地拟合关键点。他们还提出了一种随机森林回归方法来直接回归SMPL 参数，以准确性为代价降低时间消耗。[15]则去拟合一个刚性骨架模型来估计的二维和三维关节位置，可以在优化后恢复每个关节的三维旋转。同样，[16]直接回归固定运动树的关节旋转。还有其他相关的方法可以预测 SMPL 相关的输出： [17]使用合成数据通过卷积神经网络来获取 SMPL 参数。 [18]同样输出了人体的密集对应图。两者都是基于三维人体的 2.5D 投影。 [19]使用生成的人体模型从单个图像估计身体姿势。他们只是处理视觉上简单的图像，不评估三维姿势的准确性。 [20]以两步方法推断 SMPL 参数，首先使用合成数据学习 SMPL 到轮廓解码器，然后使用固定解码器学习图像到 SMPL 编码器网络。他们的目标是重建图像轮廓。虽然这是一个有趣的方向，但对轮廓的依赖限制了他们对正面图像的处理。

[21]应用深度学习提出了一个端到端的SMPL人体模型重建方法。如图所示，通过输入一张图片，应用 ResNet 网络进行特征提取，获取的的数据进行 3 次回归之后，输出得到人体姿势参数θ，体态参数β和相机参数。利用θ和β即可重建出 SMPL 人体模型M(θ,β)。由于目前在三维重建领域三维人体数据集很少，二维人体数据集很多，所以作者巧妙地利用二维和三维数据集，在重建过程中将求解获取的三维关节点进行投影，得到二维关节点数据，网络的损失函数主要考虑二维关节点。同时，作者应用三维数据集训练一个判别器，重建求解的结果将送入这个判别器，判断人体的姿势是否符合人体的正常生理运动状态，避免人体肢体的穿插和严重的扭曲。[22]同样利用神经网络，首先输入一张图片，通过Human2D 结构来推断人体关节点的位置和人体的掩膜，然后分开处理人体关节点和人体掩膜。推断的人体关节点信息又再次送入一个多层感知机 PosePrior 得到姿势参数，掩膜信息则送入 ShapePrior 的感知机得到体态参数。之后利用θ和β确定人体结构。值得注意的是，在得到掩膜的过程中采用的可微分的渲染方法，使得渲染过程可以直接插入到神经网络结构中。

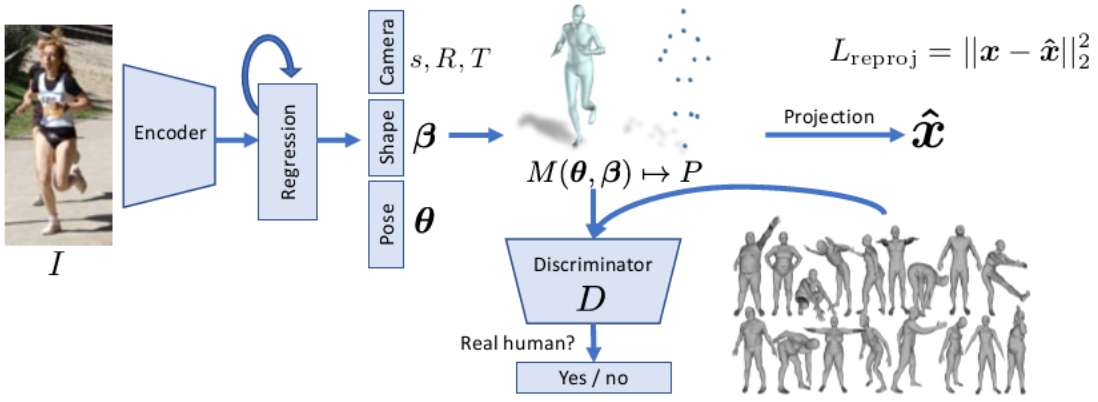

图10 [21]中基于深度学习的人体重建框架

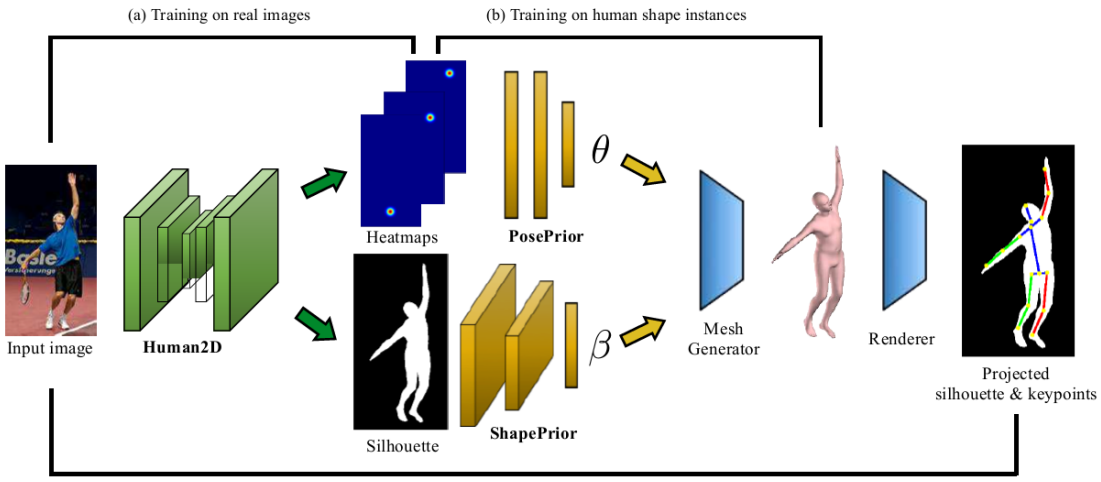

图 11 [22]中基于深度学习的人体重建框架

[43]、[45]和[50]都使用 RNN 从视频中去回归 SMPL 参数来重建三维人体模型，在[44]和[47]中在除了能够恢复三维人体体态和姿势还可以恢复表面细节。同样的在[46]里面为了使重建的模型能够包含更多的细节，从而采用符号向量场的技术，而[49]中使用自注意机制所以两者可以包含更多的细节,。在[48]里面作者提出了一种非监督的方法，这样即使没有大量的三维人体数据集使用非监督方法也能够得到很好的结果。

2.3 表面动态融合方法

表面动态融合和方法通过深度深度相机进行人体动态融合。针对动态的对象表面不断的形变进行动态融合。比如，应用单个深度相机，获取到人体的正面深度信息，而要获取人体完整的信息则需要当人体在转动的时候捕捉到背面或其他角度的深度信息，多个视角的信息动态融合以此得到完整的人体结构。

近来，人们提出了许多解决自由形式捕获的方法：线性变分变形[23]、变形图[24]、子空间变形[25]、关节变形[26][27]和[28]、4d 时空表面[29]和[30]、不可压缩流[31]、动画制图[32]、准刚性运动[33]和定向字段[34]。

[35]提出了一种分层节点图结构和一个近似的直接 GPU 计算器，以实现实时捕获非刚性场景。 [36]提出了一种实时管线，利用动态场景的阴影信息来改善非刚性配准，同时利用精确的时间对应来估计人体表面。[37]使用 SIFT 功能改进跟踪， [38]提出了规范化的惩罚约束。然而，这两种方法都没有表现出用自然动作捕捉全身性能。Fusion4D[39]使用 8 个深度摄像头设置装备，以实时捕捉具有挑战性动作的动态场景。BodyFusion[40]利用骨骼先验进行人体重建，但无法处理具有挑战性的快速运动，也无法推断人体内部形状。

[41]中的主要挑战是采用 SMPL 模型，最初不完整的外表面导致很难去进行

模型拟合。解决方案是不断更新的形状和姿态，加入更多的几何融合。依次执行关节运动跟踪、几何融合和体态姿态优化。作者采用三维网格化体素的方法处理人体结构，第一帧用户用 A-Pose 进行初始化，在已知初始值的情况下把求解直接转化为线性的。这篇文章在输出形式上也比较创新，不仅输出了一个光滑的人体内表面，还输出了一个模拟人体服装的外表面。这样的输出结构可以应对很多应用场景。对于每一帧只需要 32 毫秒就可以重建出两个表面，在实时性和准确定方面有很好的性能。但是动态重建也有很多局限：衣服太厚的重建结果较差，可能时与设定的阈值有关；不能处理外表面的几何分离。

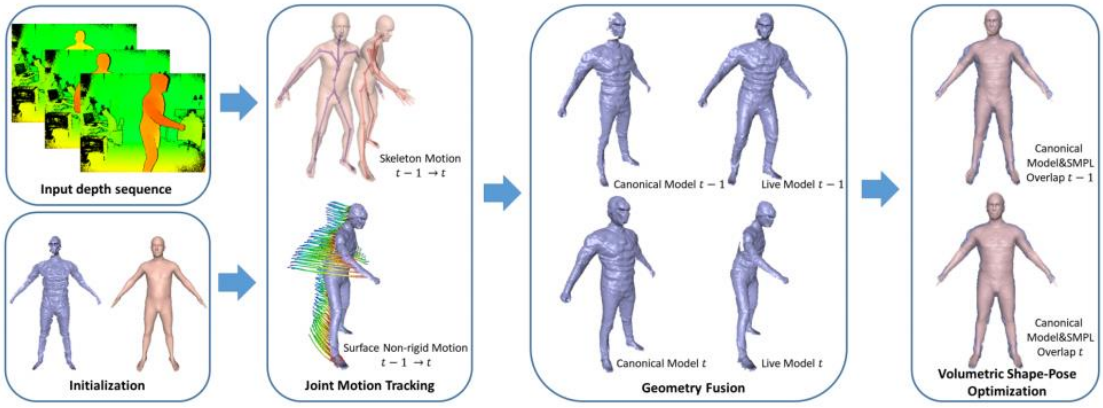

图 12 [41]中动态融合重建流程

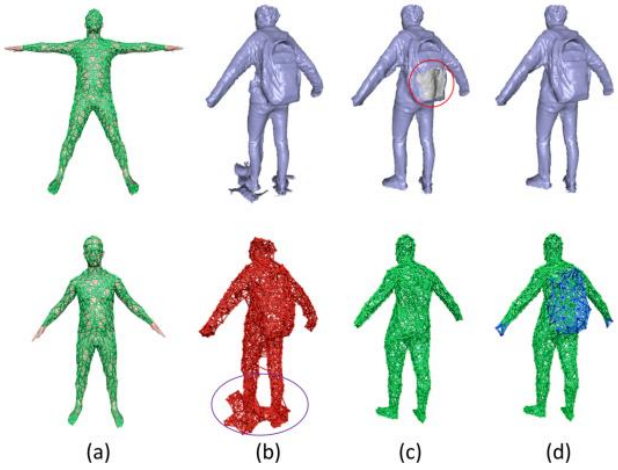

图 12 [41]中输出结构

[42]在[41]的基础上，加上了 8 个人体的惯性测量单元 IMU（指在人体上放置几个稀疏的标记），使得姿势估计更加准确，如图 13 所示。.在混合运动跟踪中，由于 IMU 和附加骨骼之间缺乏真实的对应关系，导致跟踪性能不稳定，所以采用每帧都对 IMU 进行标定。作者还对 IMU 的数量的影响也做了对比实验，最终取了 8 个。

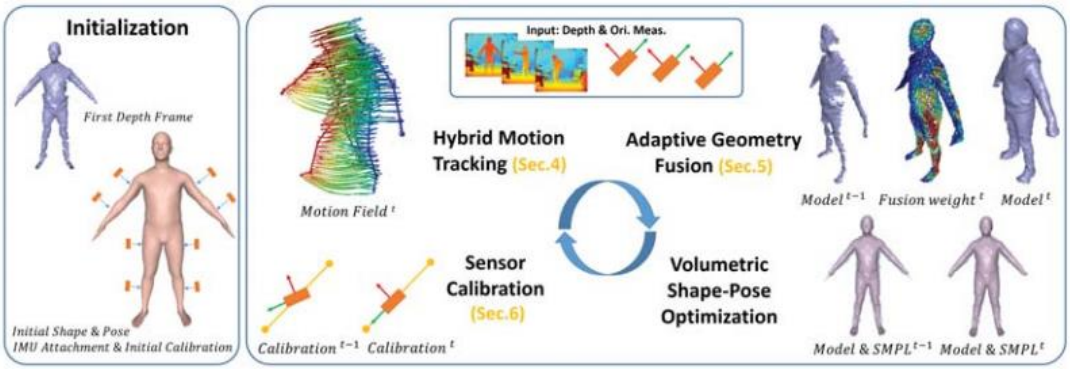

图 13 [42]中基于 IMU 的动态融合方法

3、三维人体重建的研究趋势

目前基于多视立体的方法去重建三维人体重建技术已经十分成熟，而且已经在工业中得到广泛的应用。而目前研究热点主要是利用统计模板和深度学习技术。以往的统计模板如 SMPL、SCAPE 本身就缺少很多细节，虽然后面的 SMPL-H 和 SMPL-X 在此基础上加入人手以及人脸的细节，但是仍然无法去重建人体表面的衣物。所以如何利用统计模板去能够重建人体表面的衣服是一个研究热点。虽然现在有部分人[44]、[46]、[49]提供了 GCN 以及符号向量场的技术能够重建人体表面衣物，但这些方法能力有限。

目前三维领域深度学习技术都只是在二维领域技术上更改的，所以得到的效果并不理想，虽然人们分别提出了针对网格、点云、体素的神经网络模型，但目前这些模型仍然无法得到很好的效果。学术界也出现[46]、[53]这种新的网络模型，所以如何针对三维数据特点设计一个网络模型，在未来也是一个研究热点。

目前三维人体领域缺乏 3D 数据集，所以在未来使用非监督学习技术也是研究热点。

参考文献

[1] Kanade T, Rander P, Narayanan P J. Virtualized reality: Constructing virtual worlds from real scenes[J]. IEEE multimedia, 1997, 4(1): 34-47.

[2] Zitnick C L, Kang S B, Uyttendaele M, et al. High-quality video view interpolation using a layered representation[C]//ACM transactions on graphics (TOG). ACM, 2004, 23(3): 600-608.

[3] Labatut P, Pons J P, Keriven R. Efficient multi-view reconstruction of large-scale scenes using interest points, delaunay triangulation and graph cuts[C]//Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. IEEE, 2007: 1-8.

[4] Liu Y , Gall J , Stoll C , et al. Markerless Motion Capture of Multiple Characters Using Multiview Image Segmentation[J]. IEEE Transactions on Software Engineering, 2013.

[5] Vlasic D , Baran I , Matusik W , et al. Articulated mesh animation from multi-view silhouettes[J]. ACM Transactions on Graphics, 2008, 27(3):1.

[6] Collet A , Chuang M , Sweeney P , et al. High-Quality Streamable Free-Viewpoint Video[J]. ACM Transactions on Graphics, 2015, 34(4):1-13.

[7] Vlasic D , Baran I , Matusik W , et al. Articulated mesh animation from multi-view silhouettes[J]. ACM Transactions on Graphics, 2008, 27(3):1.

[8] ZollhFer M , Theobalt C , Stamminger M , et al. Real-time non-rigid reconstruction using an RGB-D camera[J]. ACM Transactions on Graphics, 2014, 33(4):1-12.

[9] Anguelov D , Srinivasan P , Koller D , et al. SCAPE: shape completion and animation of people[C]// Acm Siggraph. ACM, 2005.

[10] Loper M , Mahmood N , Romero J , et al. SMPL: A Skinned Multi-Person Linear Model[J]. Acm Transactions on Graphics, 2015, 34(6):248.

[11] Bogo F , Kanazawa A , Lassner C , et al. Keep it SMPL: Automatic Estimation of 3D Human Pose and Shape from a Single Image[J]. 2016.

[12] Pishchulin L, Wuhrer S, Helten T, et al. Building statistical shape spaces for 3d human modeling[J]. Pattern Recognition, 2017, 67: 276-286.

[13] Chen Y , Cheng Z Q , Lai C , et al. Realtime Reconstruction of an Animating Human Body from a Single Depth Camera[J]. IEEE Transactions on Visualization & Computer Graphics, 2016, 22(8):2000-2011.

[14] Lassner C, Romero J, Kiefel M, et al. Unite the people: Closing the loop between 3d and 2d human representations[C]//IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2017, 2: 3.

[15] Mehta D, Sridhar S, Sotnychenko O, et al. Vnect: Real-time 3d human pose estimation with a single rgb camera[J]. ACM Transactions on Graphics (TOG), 2017, 36(4): 44.

[16] Zhou X, Sun X, Zhang W, et al. Deep kinematic pose regression[C]//European Conference on Computer Vision. Springer, Cham, 2016: 186-201.

[17] Varol G, Romero J, Martin X, et al. Learning from synthetic humans[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017). IEEE, 2017: 4627-4635.

[18] Güler R A, Trigeorgis G, Antonakos E, et al. DenseReg: Fully Convolutional Dense Shape Regression In-the-Wild[C]//CVPR. 2017, 2: 5.

[19] Kulkarni T D, Kohli P, Tenenbaum J B, et al. Picture: A probabilistic programming language for scene perception[C]//Proceedings of the ieee conference on computer vision and pattern

recognition. 2015: 4390-4399.
[20] Tan J, Budvytis I, Cipolla R. Indirect deep structured learning for 3D human body shape and pose prediction[C]//BMVC. 2017, 3(5): 6.
[21] Kanazawa A, Black M J, Jacobs D W, et al. End-to-end recovery of human shape and pose[C]//The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018.
[22] Pavlakos G, Zhu L, Zhou X, et al. Learning to Estimate 3D Human Pose and Shape from a Single Color Image[J]. arXiv preprint arXiv:1805.04092, 2018.
[23] Liao M, Zhang Q, Wang H, et al. Modeling deformable objects from a single depth camera[C]//Computer Vision, 2009 IEEE 12th International Conference on. IEEE, 2009: 167-174.
[24] Li H, Sumner R W, Pauly M. Global correspondence optimization for non-rigid registration of depth scans[C]//Computer graphics forum. Oxford, UK: Blackwell Publishing Ltd, 2008, 27(5): 1421-1430.
[25] Wand M, Adams B, Ovsjanikov M, et al. Efficient reconstruction of nonrigid shape and motion from real-time 3D scanner data[J]. ACM Transactions on Graphics (TOG), 2009, 28(2): 15.
[26] Chang W, Zwicker M. Range scan registration using reduced deformable models[C]//Computer Graphics Forum. Oxford, UK: Blackwell Publishing Ltd, 2009, 28(2): 447-456.
[27] Chang W, Zwicker M. Global registration of dynamic range scans for articulated model reconstruction[J]. ACM Transactions on Graphics (TOG), 2011, 30(3): 26.
[28] Pekelny Y, Gotsman C. Articulated object reconstruction and markerless motion capture from depth video[C]//Computer Graphics Forum. Oxford, UK: Blackwell Publishing Ltd, 2008, 27(2): 399-408.
[29] Mitra N J, Flöry S, Ovsjanikov M, et al. Dynamic geometry registration[C]//Symposium on geometry processing. 2007: 173-182.
[30] Süßmuth J, Winter M, Greiner G. Reconstructing animated meshes from time-varying point clouds[C]//Computer Graphics Forum. Oxford, UK: Blackwell Publishing Ltd, 2008, 27(5): 1469-1476.
[31] Sharf A, Alcantara D A, Lewiner T, et al. Space-time surface reconstruction using incompressible flow[J]. ACM Transactions on Graphics (TOG), 2008, 27(5): 110.
[32] Tevs A, Berner A, Wand M, et al. Animation cartography—intrinsic reconstruction of shape and motion[J]. ACM Transactions on Graphics (TOG), 2012, 31(2): 12.
[33] Li H, Vouga E, Gudym A, et al. 3D self-portraits[J]. ACM Transactions on Graphics (TOG), 2013, 32(6): 187.
[34] Dou M, Fuchs H, Frahm J M. Scanning and tracking dynamic objects with commodity depth cameras[C]//Mixed and Augmented Reality (ISMAR), 2013 IEEE International Symposium on. IEEE, 2013: 99-106.
[35] Newcombe R A, Fox D, Seitz S M. Dynamicfusion: Reconstruction and tracking of non-rigid scenes in real-time[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 343-352.
[36] Liu Y, Liu Y, Liu Y, et al. Real-Time Geometry, Albedo, and Motion Reconstruction Using a Single RGB-D Camera[J]. Acm Transactions on Graphics, 2017, 36(3):32.
[37] Innmann M, Zollhöfer M, Nießner M, et al. VolumeDeform: Real-Time Volumetric Non-rigid Reconstruction[J]. 2016.

[38] Slavcheva M , Baust M , Cremers D , et al. [IEEE 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - Honolulu, HI (2017.7.21-2017.7.26)] 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

[39] Dou M , Khamis S , Degtyarev Y , et al. Fusion4D: Real-time Performance Capture of Challenging Scenes[J]. Acm Transactions on Graphics, 2016, 35(4):114.

[40] Yu T , Guo K , Xu F , et al. BodyFusion: Real-Time Capture of Human Motion and Surface Geometry Using a Single Depth Camera[C]// 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, 2017.

[41] Yu T, Zheng Z, Guo K, et al. DoubleFusion: Real-time Capture of Human Performances with Inner Body Shapes from a Single Depth Sensor[J]. arXiv preprint arXiv:1804.06023, 2018.

[42] Zheng Z, Yu T, Li H, et al. HybridFusion: real-time performance capture using a single depth sensor and sparse IMUs[C]//European Conference on Computer Vision (ECCV). 2018.

[43] Zhang J Y, Felsen P, Kanazawa A, et al. Predicting 3d human dynamics from video[C]//Proceedings of the IEEE International Conference on Computer Vision. 2019: 7114-7123.

[44] Kolotouros N, Pavlakos G, Daniilidis K. Convolutional Mesh Regression for Single-Image Human Shape Reconstruction[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 4501-4510.

[45] Kanazawa A, Zhang J Y, Felsen P, et al. Learning 3d human dynamics from video[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 5614-5623.

[46] Saito S, Huang Z, Natsume R, et al. PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization[J]. arXiv preprint arXiv:1905.05172, 2019.

[47] Zhu H, Zuo X, Wang S, et al. Detailed Human Shape Estimation from a Single Image by Hierarchical Mesh Deformation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 4491-4500.

[48] Rüegg N, Lassner C, Black M, et al. Chained Representation Cycling: Learning to Estimate 3D Human Pose and Shape by Cycling Between Representations[J].//AAAI 2020

[49] Sun Y, Ye Y, Liu W, et al. Human mesh recovery from monocular images via a skeleton-disentangled representation[C]//Proceedings of the IEEE International Conference on Computer Vision. 2019: 5349-5358.

[50] Kocabas M, Athanasiou N, Black M J. VIBE: Video Inference for Human Body Pose and Shape Estimation[J]. arXiv preprint arXiv:1912.05656, 2019.

[51] Zhang Y, Hassan M, Neumann H, et al. Generating 3D People in Scenes without People[J]. arXiv preprint arXiv:1912.02923, 2019.

[52] Varol G, Laptev I, Schmid C, et al. Synthetic Humans for Action Recognition from Unseen Viewpoints[J]. arXiv preprint arXiv:1912.04070, 2019.

[53] Genova K, Cole F, Sud A, et al. Deep Structured Implicit Functions[J]. arXiv preprint arXiv:1912.06126, 2019.

[54] Joo H, Simon T, Sheikh Y. Total capture: A 3d deformation model for tracking faces, hands, and bodies[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 8320-8329.

[55] Romero J, Tzionas D, Black M J. Embodied hands: Modeling and capturing hands and bodies together[J]. ACM Transactions on Graphics (TOG), 2017, 36(6): 245.

[56] Pavlakos G, Choutas V, Ghorbani N, et al. Expressive body capture: 3d hands, face, and body from a single image[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 10975-10985.