

分类号: TP311.5

单位代码: 10335

密 级: 无

学 号: 21121098

浙江大学

硕士学位论文



中文论文题目: 基于 Kinect 的动态人体三维重建

英文论文题目: Kinect Based Human Performance
Reconstruction

申请人姓名: 郑家祥

指导教师: 刘新国 教授

专业学位类别: 计算机应用技术

专业学位领域: 计算机图形学

所在学院: 计算机科学与技术学院

论文提交日期

基于 Kinect 的动态人体三维重建



论文作者签名: _____

指导教师签名: _____

论文评阅人 1: 秦绪佳 教授 浙江工业大学

评阅人 2: 冯结青 教授 浙江大学

评阅人 3: 万华根 副教授 浙江大学

评阅人 4: _____

评阅人 5: _____

答辩委员会主席: 缪永伟 教授 浙江工业大学

委员 1: 冯结青 教授 浙江大学

委员 2: 金小刚 教授 浙江大学

委员 3: 刘新国 教授 浙江大学

委员 4: _____

委员 5: _____

答辩日期: 2014 年 3 月 7 日

浙江大学研究生学位论文独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得 浙江大学 或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签名：

签字日期： 年 月 日

学位论文版权使用授权书

本学位论文作者完全了解 浙江大学 有权保留并向国家有关部门或机构送交本论文的复印件和磁盘，允许论文被查阅和借阅。本人授权浙江大学可以将学位论文的全部或部分内容编入有关数据库进行检索和传播，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

（保密的学位论文在解密后适用本授权书）

学位论文作者签名：

导师签名：

签字日期： 年 月 日

签字日期： 年 月 日

摘要

准确而可靠的三维动态几何数字化在影视制作、计算机动画等图形学相关领域中是一个重要的组成部分。不同于静态几何的重建，动态几何帧与帧之间缺少准确的对应信息，此外几何体的自遮挡等问题也不可避免，这使得动态重建问题变得十分具有挑战性。本文从这些问题出发，研究动态人体重建的整体框架与基本算法，在不需要显式模板的情况下从 Kinect 组成的系统中自动重建出拓扑完备的三维动态模型。

我们的硬件系统包含两对相对朝向的 Kinect 相机，利用 Kinect 相机能够直接获得三维信息的特殊性，对相对朝向的相机我们采用了与传统方式不同的定标方法，改进棋盘定标处理相对朝向相机的不足。首先，我们采用了一个人体统计模型 SCAPE 生成隐式的模板。同时针对姿态训练数据的不足，我们设计了姿态传递的方法将其它数据库中的姿态模型迁移到训练数据库中，使得训练数据库中的姿态更加丰富多样。此外，运动重建过程中很重要的一个步骤是运动跟踪，我们采用了一个分段刚体预对齐的方法，并使用层次式的方法求解，使得整个跟踪过程更加地鲁棒和稳定。通过非刚体变形的方法，我们最终将整个运动过程以拓扑完备的模型方式表现出来。

我们采用了多个实验来验证算法的有效性，同时使用不同的姿态和体态运动数据对整个的系统进行测试。事实上，尽管 Kinect 相机输出的原始数据是充满噪声的，但重建结果表明，我们的算法是可以重建出比较理想的结果的。

关键词： 三维重建，SCAPE，运动重建，运动跟踪

Abstract

Accurate and reliable 3D digitization of dynamic geometry is a critical component for film making, computer animation and related CG areas. Different to static geometry reconstruction, dynamic reconstruction is challenging for researchers due to the lack of accurate correspondences between adjacent frames of dynamic geometry and the inevitable inherent occlusions. This thesis aims to investigate the basic algorithms and whole framework based on these problems, and try to use Kinect cameras to reconstruct deformable topologically complete geometry automatically without an explicit template.

We employed a statistical model called SCAPE for estimating our implicit template's initial shape and pose. Due to the insufficient training data, we devised a pose transfer method to expand the pose database to make it more various. Our setup contains two pairs of cameras with opposite directions. Since the traditional chessboard based calibration method is not applicable, utilizing the property of Kinect camera, we developed an easy calibration method for opposite direction cameras. Furthermore, we utilized the articulated ICP method followed by a hierarchical solution, which makes the tracking stage more robust and stable. Finally we employed a non-rigid deformation method to reconstruct the whole motion.

We conducted several experiments to evaluate the effectiveness of our method and utilized the algorithm to reconstruct different body shapes and poses. Actually, although the output of Kinect is full of noises, our experiments show the algorithm is robust to reconstruct considerable result.

Keywords: 3D reconstruction, SCAPE, motion reconstruction, motion tracking

目 录

摘要	i
Abstract	ii
第 1 章 绪论	1
1.1 课题背景	1
1.2 动态几何数据获取技术	1
1.2.1 主动式数据获取技术	2
1.2.2 被动式数据获取技术	3
1.3 动态几何重建研究现状	4
1.3.1 基于模板的运动重建	4
1.3.2 无模板的运动重建	5
1.4 本文工作及组织结构	6
第 2 章 几何对齐方法	8
2.1 几何对齐	8
2.2 刚体对齐方法	9
2.3 非刚体对齐方法	13
2.4 本章小结	15
第 3 章 SCAPE 模型及姿态数据库扩充	16
3.1 SCAPE 介绍	16
3.1.1 模型的训练	18
3.2 姿态数据库的扩充	20
3.3 本章小结	23
第 4 章 人体运动重建系统	24
4.1 定标方法	25
4.1.1 初始外参计算	25

4.1.2 外参的全局优化 26

4.2 模板生成 28

4.3 运动跟踪算法 30

4.4 运动重建 32

4.5 实验结果 33

4.5.1 对比实验 33

4.5.2 更多结果 35

第 5 章 总结与展望 41

5.1 总结 41

5.2 展望 41

参考文献 42

发表文章目录 46

致谢 47

图目录

图 1.1	Kinect 及 Kinect Fusion 重建的结果	2
图 1.2	主动式的 Photometric 方法重建结果	3
图 1.3	被动式的重建结果 ^[1]	4
图 1.4	运动重建流程 ^[2]	5
图 1.5	基于模板的运动重建	5
图 1.6	无模板的运动重建	6
图 2.1	几何对齐示意图	9
图 2.2	刚体对齐	10
图 2.3	ICP 算法框架图	11
图 2.4	ICP 算法采样方法	11
图 2.5	形变图非刚体变形示例 ^[3]	14
图 3.1	SCAPE 的应用	17
图 3.2	SCAPE 中隐式包含的骨骼系统 ^[4]	18
图 3.3	SCAPE 中变形的分解 ^[4]	18
图 3.4	训练数据对学习结果的影响	21
图 3.5	变形结果	22
图 3.6	姿态迁移结果	23
图 4.1	硬件系统	24
图 4.2	用于朝向对立的相机定标的棋盘格	26
图 4.3	利用第一帧数据生成隐式模板	29
图 4.4	层次式关节对齐方法示意图	31
图 4.5	数据库扩充前后真实学习结果对比	34
图 4.6	关节 ICP 算法对比结果	34

图 4.7 遮挡处理结果 35

图 4.8 仿真实验结果 36

图 4.9 重建结果 1 37

图 4.10 重建结果 2 38

图 4.11 重建结果 3..... 39

图 4.12 重建结果 4 40

表目录

表 4.1 定标反投影误差对比结果 28

表 4.2 实验环境 33

第 1 章 绪论

1.1 课题背景

随着近年来计算机动画、计算机视觉的交叉发展，运动物体的几何跟踪变得尤为重要。而对于动态几何准确可靠的三维数字化在现代 CG 动画的创建中亦成为一个不可或缺的组成部分。其应用领域非常广泛，包含机器人学、生物医学、交互游戏式的教育和影视制作等。

传统意义上的动作捕捉主要是靠一些绑定在动画角色身上外部设备进行，如电影《阿凡达》中角色的逼真表情与动作就是使用了很多高昂设备进行捕捉的结果。其好处在于它能够产生非常逼真的效果，但是在使用过程中由于需要在身体上附着各种高精度设备，导致运动过程有诸多不便，无法自由地表现其丰富的运动。于是，无标记的运动重建就成了一个越来越值得关注和研究的问题了。此外，运动捕捉设备造价高昂，限制了它只能用于影视制作等领域，但在大众领域，它却无法得到广泛使用。本文正是在这样的背景下进行的，主要集中于探讨使用低廉的 Kinect 相机对运动的人体模型进行三维重建。

运动重建与计算机动画不同，以人体运动为例，计算机动画关注的是如何通过捕捉到的人体各关节的运动信息虚拟地驱动一个角色模型进行运动。其输入一般用各关节相对角度的变化来表示。而与此不同的是，运动重建旨在通过三维重建的方法将整个的运动过程以网格模型的形式展现出来。由于不同帧之间得到的数据之间本身没有内部对应，运动过程中造成的不可避免的自遮挡等问题，导致运动重建的困难性。

1.2 动态几何数据获取技术

动态几何重建的整个过程中，非常重要的第一步即数据的获取。因为要获取的是一个动态数据，因而需要保证数据的获取是一个实时的技术。一类方法是可以预先拍摄出完整的多视图运动图像，然后离线进行每一帧的重建；另一类是通过实时地输出物体的三维几何结构。由于数据获取方法有很多，按其有无标记可以将这些方法划分为基于标记和无标

记的方法，本节将探究几种无标记的常见数据获取技术。

无标记的数据获取技术可以分为主动式的方法和被动式的方法，具体而言，主动式的数据获取方法通过控制场景中光照的变化，使用光度立体视觉的方法或三角测量技术等重建出每一帧的模型，被动式的方法则适用于光照条件变化没有那么丰富的情况，它基于纹理匹配的深度立体视觉方法。与主动式的方法相比，被动式的方法重建出来的结果依赖于其纹理质量及光照条件，对于纹理不够丰富的模型，这个方法就不如主动式的方法鲁棒了；此外，被动式的方法因为需要计算全局的纹理匹配，导致它的计算代价非常地大。

1.2.1 主动式数据获取技术

因为无需标记和纹理信息进行匹配，主动式的方法在最近的动态重建相关工作中使用得更为广泛。主动式的数据获取方法有很多种，虽然都是基于控制光线获得物体的三维数据，但有些方法是利用光线的反射，通过计算时间差来计算物体离相机的距离，从而得到物体的深度；有些方法则是通过发射不同的光带特征（也称为结构光）的方法，利用光学三角测量技术计算物体的深度变化。显然，对于利用光线反射的方法，当拍摄运动数据的时候，需要在短到可以忽略的时间内计算出物体的距离等信息。



图 1.1 Kinect 及 Kinect Fusion 重建的结果

光学三角测量技术通过发射光线，并计算两个相机（或一个相机一个光源发射器）之间的对应点采样几何体表面，从而将物体表面投射出更丰富的纹理，通常这个纹理是特殊设计的，具有一定的特征，有的也称为结构光。通过不同的特征，可以非常容易地匹配出像素点并计算其深度。以近年来广泛使用的 Kinect 为例，它包含一个 RGB 相机和一个红外相机，通过红外相机发射红外线，其内置感应设备可以计算出红外线从发出到经过物体反射后所需的时间，从而可以计算出物体到相机的深度，从而也即得到物体单视图的部分模型。基于 Kinect 的重建中，最著名的是 Kinect Fusion^[5]，它通过构造一个大体素 ($10\text{m} \times 10\text{m} \times 10\text{m}$)，将需要重建的物体及场景嵌入在该体素中，同时利用 GPU 加速，实时地重建出物体的完整三维几何。如图 1.1 所示，左图为 Kinect 相机，右图是使用 Kinect

Fusion 算法恢复出来的法向图及三维几何模型。但是，这是经过一定处理后得到的几何数据，原始的 Kinect 返回数据是包含很大的噪声，如图1.1右边第一个图所示。

而光度立体视觉（Photometric）的方法假定物体表面是 Lambertian 表面，使用不同方向的光照条件，该方法可以通过求解一个简单的线性方程求解物体表面的法向，从而积分求出物体的三维形状。如图1.2所示，即为使用光度立体视觉重建出来的模型。有关这个方法的综述，可以参考 Barsky 等人^[6]的工作。这个方法能够高精度的重建高频细节，但对于低频信息，则不是很准确，基于此，Nehab 等人^[7]结合法向信息与使用其它方法重建出来的粗糙模型，重建出一个更加准确的模型。基于这个方法，通过设计更复杂的光照条件可以重建出非常高的精度的模型。

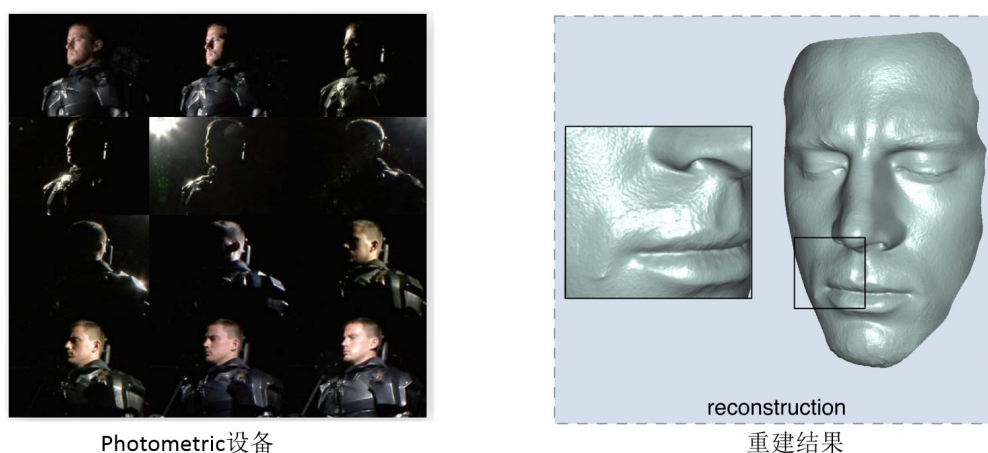
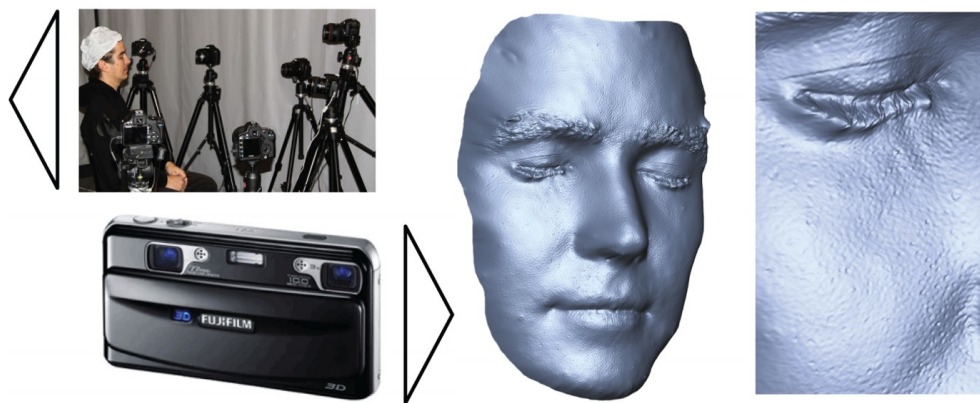


图 1.2 主动式的 Photometric 方法重建结果

1.2.2 被动式数据获取技术

被动式的方法使用光学三角测量技术计算多视图下各图像间的纹理对应信息，从而间接地通过获得几何体的深度信息，达到重建的目的。与主动式的方法相比，它没有复杂的光照特征提高其匹配结果，因此基于图像的深度估计难点往往也是在其立体匹配上。一般的方法是通过估计相机的内外参数后，使用对极几何对图像进行变换，使得匹配结果的搜索空间降低到一维的窗口进行搜索。通常而言，由于纹理的丰富性和没有明显特征辅助匹配，匹配在被动式的方法中往往更加困难。经过几十年的发展，学者提出使用一个全局优化框架，使得被动式的立体匹配技术有了很大的进步。其中一些方法通过将视差图建模成一个 Markov 随机场，从而使用诸如 Graph Cut 或 Belief Propagation 等方法求解最优的连通表面。

前面也提到，被动式的方法重建出来的精度往往不如主动式方法好。在某些特定的

图 1.3 被动式的重建结果^[1]

场景，如人脸^[1]等，被动式的方法重建出来的结果是可以达到结构光扫描的精度的。从图1.3可以看到，被动式的重建结果已经达到了毛孔级别。由于被动式的方法拍摄的每一帧都是在一个时间点就可以完成的，因此可以直接先重建出每一帧的结果，然后利用动态几何跟踪重建的方法恢复出最后的结果。这里不再赘述。

1.3 动态几何重建研究现状

近年来，许多研究工作集中于研究包括静态场景^[5;8]与动态模型（如人脸表情^[9;10]、人体运动）等。在接下来的内容中，我们主要集中于探讨人体重建。广义而言，运动重建可以简单地划分为基于模板的方法和无模板的重建方法。

1.3.1 基于模板的运动重建

无论是通过单视图或多视图获取一个几何体的三维结构信息，都不可避免地会出现运动遮挡及自遮挡问题，模板可以看作是解决这些遮挡带来的破洞的一个先验知识。其基本流程大致如下图1.4。

通过单视图我们只能得到模型的一部分信息，Li 等人^[2]为了从单视图获得的运动几何得到一个完整的运动模型，提出使用变形图的方法驱动一个精细的光滑模板发生变形。类似于图1.4的流程，首先进行一个大尺度的变形将模板变形到与扫描数据大体一致，之后再进行细节增强的一个过程。由于单视图恢复整个运动模型问题的本身就是一个欠约束的问题，因此，大部分基于模板的重建方法都是集中于对多视图得到的数据（如 Light Stage 系统）进行重建。许多学者^[11-14]使用了蒙皮的 LBS 方法驱动人体模板发生变形，达到大尺度的姿态形变目的，而后采用拉普拉斯等变形算法将细节不断地增加到模板模型上。类似

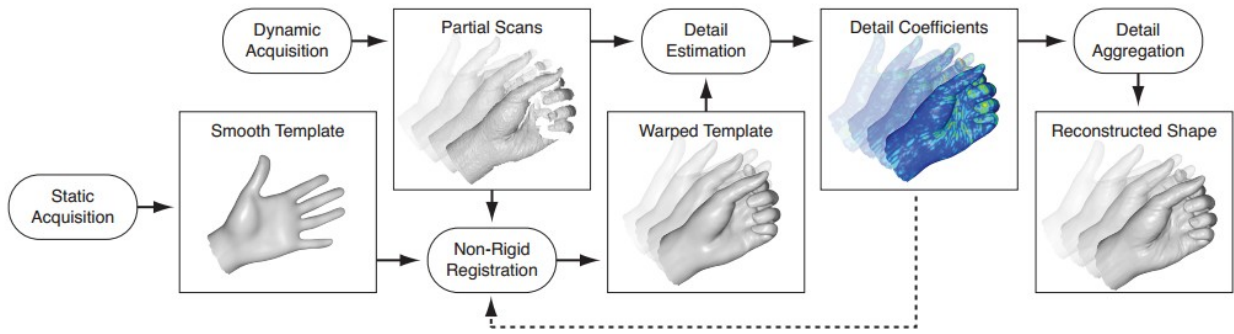


图 1.4 运动重建流程^[2]

地，Aguiar 等人^[15] 基于多视图的 Light Stage 系统提出采用基于轮廓和 SIFT 特征匹配的方式进行对应的寻找和变形。

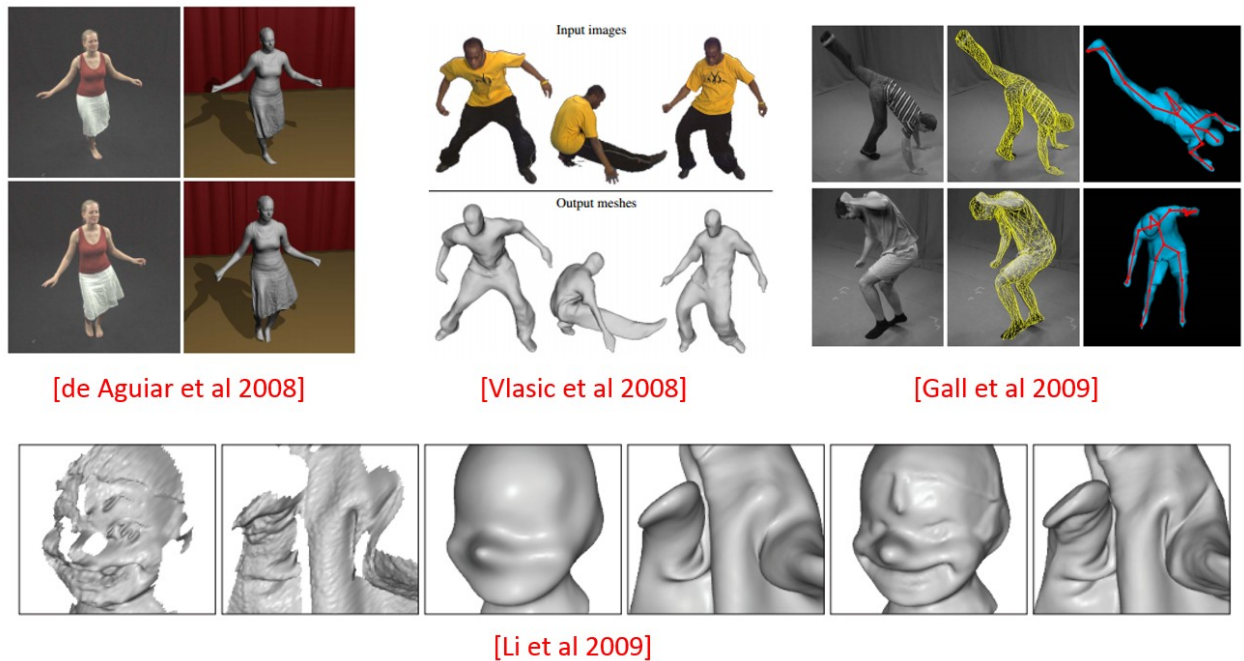


图 1.5 基于模板的运动重建

尽管模板可以给我们提供一个非常好的先验，但模板的获取本身是一个代价高昂的过程。与这些方法不同之处在于，我们提出的方法是通过采用统计模型隐式地生成一个对应的模板，达到重建的目的。

1.3.2 无模板的运动重建

前面也提到，构建模板本身是一个高昂代价的过程，针对从没有模板的几何结构中恢复出运动模型也得到了许多研究人员的关注。Vlasic 等人^[16] 在 Light Stage 系统中通过利

用光流以将运动数据对应起来，同时采用 VRIP 方法^[17] 将最终的重建结构补全成一个完备的模型。使用类似的硬件设备，Li 等人^[18] 使用各帧的图像轮廓构建出一个可视壳作为先验，并采用图变形的方法将相邻帧数据匹配对应起来，并最终使用泊松重建的方法^[19] 实现模型的补全。与这些多视图硬件获取设备不同，单视图下的运动重建也得到了诸多关注。Chang 等人^[20] 假设模型的运动过程是可以看作分段刚体的，通过这个先验假设，将整个变形的过程简化为少数的几个刚体变换估计，从而逐渐地将整个模型重建出来。这个假设显然决定了该方法也只能用于重建带关节的运动几何。Zeng 等人^[21] 则假定整个运动是一个姿态几乎不变（即近似刚体），而细节发生轻微变化的过程，这样就可以将整个过程首先用一个刚体对齐的方法将整体变换计算出来，然后采用图变形的方法将不同帧逐渐对齐并拼接成一个完整的三维模型。

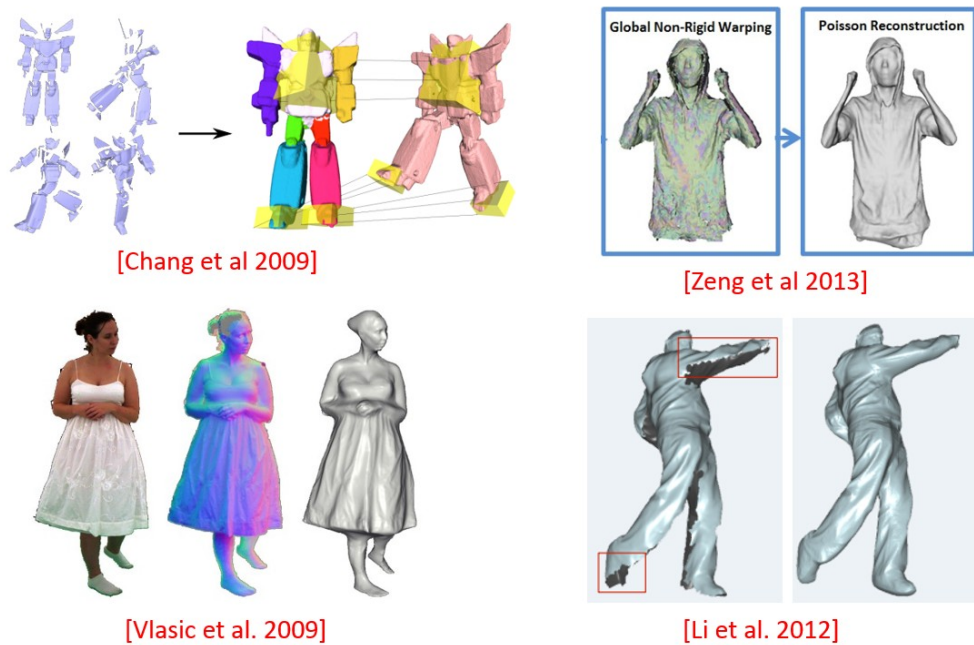


图 1.6 无模板的运动重建

1.4 本文工作及组织结构

本文的研究内容在于通过 Kinect 相机重建出动态的人体运动过程。

一方面 Kinect 相机本身噪声大，易受光照和红外线等的干扰；另一方面，人体运动具有复杂多样的特性，不可避免地会带来许多自遮挡引起的内容缺失，导致难以重建完整的模型。本文首先介绍如何将几何体进行对齐，即刚体的迭代最近邻算法和非刚体嵌入空间图变形对齐算法。接着我们详细介绍我们采用的人体姿态与体态统计模型 SCAPE，鉴于训

训练数据不足将导致学习结果的不准确，我们提出了基于变形传递的姿态数据库扩充方法。然后我们介绍我们的人体运动重建系统，在帧与帧之间的运动跟踪过程中，我们提出了使用分段刚体预对齐的方法辅助跟踪，使得整个跟踪算法更加地准确与鲁棒；最后我们介绍整个运动重建的方法，并通过实验部分展示我们算法的高效鲁棒性。

具体而言，本文的组织结构如下：

第一章为绪论部分，主要介绍本文研究的背景及意义，并对当前动态几何重建的研究现状等进行回顾。

第二章介绍几何体对齐的方法，主要是包括刚体对齐与非刚体对齐。几何对齐的重要性在三维重建中不言而喻，一方面，为了将各角度的几何信息拼接到一起，我们需要利用不同角度重叠的信息将多角度几何体拼接成一个完整的模型；另一方面，单纯地从几何的角度来看，同一个物体在不同帧中的对应要么通过几何特征（如基于曲率等局部信息），要么就需要通过对齐的方法将同一物体对应起来。这一章包括刚体对齐和非刚体对齐的方法，这些内容在后面的重建过程中需要用到。

第三章介绍 **SCAPE** 模型相关的内容。我们的重建系统中硬件设备使用 **Kinect** 相机，从它得到的只是一系列稠密点云。**SCAPE** 模型可以为我们提供人体模板作为先验知识，从而更好地便于我们处理包含遮挡及噪声的数据，并从中恢复出完整的模型。在这一章中，我们将首先介绍 **SCAPE** 模型，然后介绍我们在 **SCAPE** 数据库训练数据不足的情况下是如何解决姿态数据库扩充这个问题的。

第四章是我们的人体运动重建系统。首先我们将介绍整个系统的硬件设施，它包含四台 **Kinect** 用于获取完整人体的模型，同时由于四台相机的位置特殊性，我们提出了我们自己的定标方案并进行了全局优化，最后量化定标的结果以证明其有效性。结合前一章训练的人体统计模型，我们提出使用分段刚体对齐的方式进行运动跟踪。同时，我们详细地阐述了整个运动过程的重建，并给出相应的实验结果。

在第五章，对全文作出总结与展望。

第 2 章 几何对齐方法

在三维重建的过程中，通常无法从单视图将物体的完整三维结构恢复出来，意味着每帧只能得到单个角度的三维信息，也称为 2.5 维。为得到完整的三维模型，我们需要将各角度的重建结果拼接到一起。事实上，基于 Light Stage 的多目立体视觉或光度视觉方法通过多角度安置相机并定标达到完整重建的目的。当物体本身是刚体的时候，这和 Light Stage 下的各角度相机结果拼接是等价的，因为要进行多角度拼接，就是相当于将各个角度的相对位置求出来，而这一步在定标的过程中亦是如此。

如果物体本身是一个刚性的物体，我们要求的仅仅是一个刚体变换。但是，在实际过程中，获取三维信息的设备会包含一定的轻微畸变，而重建的物体有可能是一个易发生形变的物体（如人体或弹性物体等），往往易造成同一区域在不同时间得到的信息无法单纯地用一个刚体变换进行对齐。

鉴于几何对齐方法在三维重建领域的重要作用，我们在本章对三维几何对齐的方法进行一个简要的概述，其内容包含刚体^[22]与非刚体的对齐^[3]。

2.1 几何对齐

在介绍刚体与非刚体对齐之前，我们先在一个统一的框架下考虑几何对齐这个问题。为了叙述的方便，这里主要是针对两帧几何数据的对齐进行介绍。

对几何对齐问题形式化地描述如下：给定两帧扫描数据 $\mathcal{S}(t_1)$ 和 $\mathcal{S}(t_2)$ ，两帧数据的重叠部分 $\mathcal{S}_{t_1 \cap t_2}$ 不为空，且以 $\mathcal{S}(t_1)$ 为源数据， $\mathcal{S}(t_2)$ 为目标数据，通过某一个对齐方法 $\Phi_{t_1 \rightarrow t_2}$ 对齐后的结果为 $\mathcal{S}_{t_1 \rightarrow t_2}$ ，则我们希望 $\mathcal{S}_{t_1 \rightarrow t_2}$ 与 $\mathcal{S}(t_2)$ 在 $\mathcal{S}_{t_1 \cap t_2}$ 的区域能够尽可能地一致和重叠，这样就可以将 $\mathcal{S}(t_2)$ 中没有而 $\mathcal{S}(t_1)$ 出现的区域恢复，重复此过程即可将完整的模型逐渐构建出来。

如果物体本身是一个刚体，则对齐问题即估计出该物体的一个刚体运动，为 6 个参数，相对来说是比较简单的；如果物体本身是一个非刚体，即它会随着时间的变化而变化，则对齐变得非常复杂，因为我们事先是无法知道两帧数据的重叠部分，而且变形本身也是比

较复杂的一个过程。如图2.1所示的即非刚体对齐的过程，而图 2.2则为刚性物体的对齐。

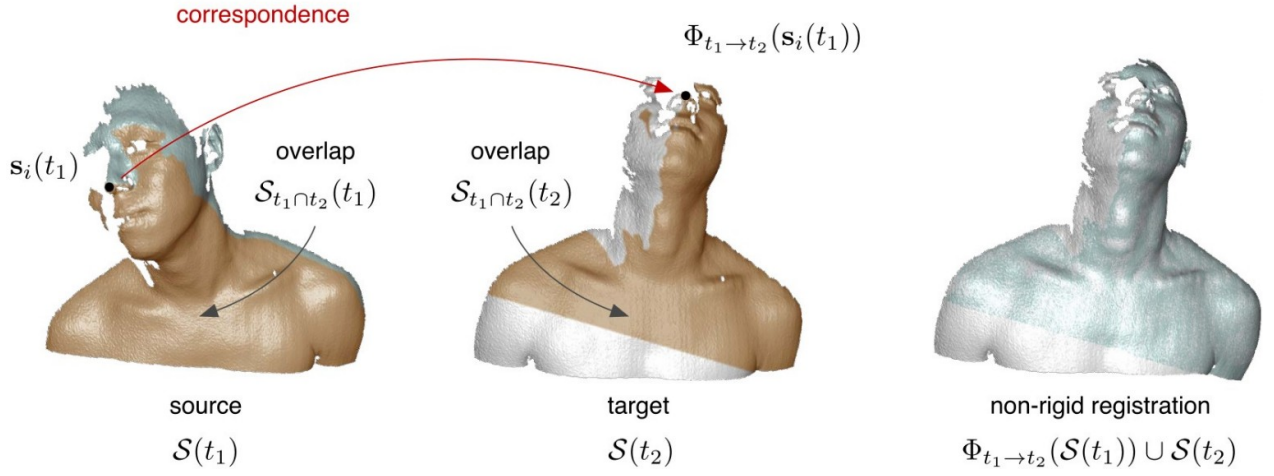


图 2.1 几何对齐示意图^[23]

统一对这个问题进行讨论，我们的目标实际上即求出一个合适的 $\Phi_{t_1 \rightarrow t_2}$ 使得满足上述的对齐要求，我们可以将这个问题描述为能量最小化的问题，能量定义为：

$$E_{\text{reg}} = E_{\text{match}} + E_{\text{prior}} \quad (2.1)$$

其中， E_{match} 用于衡量 $\mathcal{S}_{t_1 \rightarrow t_2}$ 与 $\mathcal{S}(t_2)$ 在重叠部分尽可能地一致； E_{prior} 用于控制 $\mathcal{S}(t_1)$ 所允许的变形类型，比如控制它是刚体变换、非刚体弹性变换、保持局部不变的变换等。

下面我们分别考虑刚体和非刚体两个条件下这个问题的解决方法。

2.2 刚体对齐方法

如图2.2所示， $\mathcal{S}(t_1)$ 即标记为红色的部分模型（左）， $\mathcal{S}(t_2)$ 即标记为绿色的部分模型（中），二者很大一部分区域是重叠的，但是在鼻子等部位，可以看到是存在不重叠的地方，通过应用刚体变换 $[R \ t]$ 到 $\mathcal{S}(t_1)$ ，可以看到在右图中，二者可以很好地对齐。

从前面的描述中不难发现，需要求解的变量实际上只有 6 个，即 3 个旋转分量和 3 个平移分量。因此式2.1中的 E_{prior} 即保证旋转分量为正交矩阵。假定待估计的刚体变换为 $\mathbf{T} = [R \ t]$ ，则目标即使得 $\mathbf{T} \circ \mathcal{S}(t_1)$ 与 $\mathcal{S}(t_2)$ 尽可能多地重合。式 2.1详细描述即：

$$\begin{cases} E_{\text{match}} &= \sum_{(v, v') \in C} \|\mathbf{T}v - v'\|^2 \\ E_{\text{prior}} &= R \in \text{SO}(3) \end{cases} \quad (2.2)$$

其中 C 为 $\mathcal{S}(t_1)$ 和 $\mathcal{S}(t_2)$ 的一系列对应点。

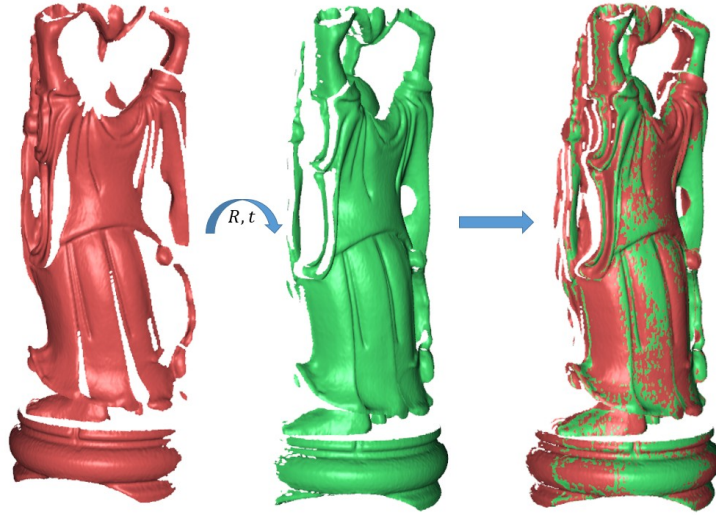


图 2.2 刚体对齐

要使得刚体变换 \mathbf{T} 在最小化上式同时保证 R 的正交性，数学上这个问题是可解的，基于 SVD 分解的方法^[24]如下：

定义关于对应点 v 和 v' 的协方差矩阵 Σ 为：

$$\Sigma = \frac{1}{\#C} \sum_{(v_i, v'_i) \in C} (v_i - \bar{v})(v'_i - \bar{v}')^T \quad (2.3)$$

其中 $\#C$ 表示对应点的个数， \bar{v} 和 \bar{v}' 分别为 \mathcal{S} 和 \mathcal{T} 上对应点的平均值。对协方差矩阵 Σ 进行 SVD 分解，即 $\hat{\Sigma} = UDV^T$ ，从而刚体变换中旋转部分 $R = UV^T$ ，从而也就可以得到平移分量 $t = \bar{v}' - R\bar{v}$ 。具体的证明细节可以参考^[24]。

但问题在于没有任何对应的情况下，要最小化上式能量，显然需要求解的变量包含 C 和 \mathbf{T} ，而且还需要保证 \mathbf{T} 中的 R 为正交矩阵。这不是一个简单的线性问题。从点云中抽取局部特征计算 C 作为少数的约束点当然是可行的，但是计算三维几何结构的局部特征描述算子一方面是不够鲁棒，另一方面，它的计算量非常地大，即耗时很长。

事实上，刚体几何对齐的方法是计算机视觉中一个经典的问题。早在上世纪 90 年代初，就有学者^[22;25]提出了在视觉领域具有影响力的算法：迭代最近邻算法（ICP）。该算法通过不断迭代求解最近邻点作为对应，求解出一对几何体之间的刚体变换。它具有快速准确的特性，但由于是一个不断迭代求解的非线性非凸问题，它需要一个良好的迭代初始估计值。Rusinkiewicz 等人^[26]对该算法作了一个很好的综述，并对该算法每一个步骤进行优化，最终提出一个可以实时运行并高效鲁棒的 ICP 变种算法，具体的算法分析可以参考此文，接下来介绍 ICP 算法基本框架及求解思路。

ICP 采用了一个迭代的方法来求解该问题，求解框架如下图 2.3 所示：

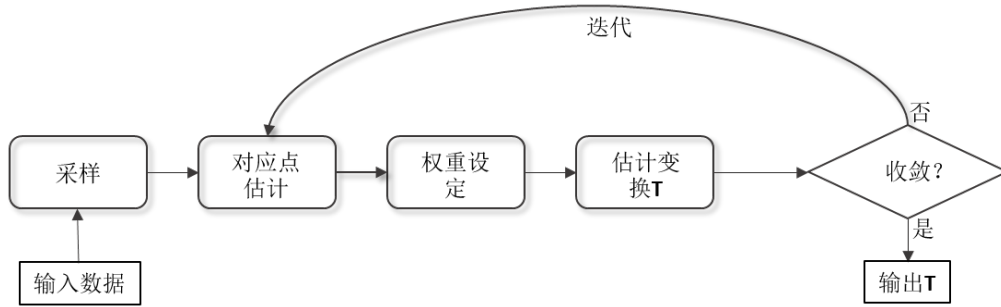


图 2.3 ICP 算法框架图

具体而言，ICP 算法首先为 $S(t_1)$ 的每一个点计算出 $S(t_2)$ 中的一个对应。对于分辨率比较高的模型，需要对点集进行随机采样，然后寻找对应点。注意到对应点如果找的不准，会对于估计结果造成重要的影响，因此还需要对找到的对应点集合进行过滤，可以将它们当作异常点直接剔除，一个策略是将那些对应点距离大于一定阈值的点对剔除。同时根据对应的可靠性分配权重，然后利用对应点结合某种评估函数采用某种方法估计出刚体变换。

下面重点分析如何进行采样、对应点估计及剔除、评估函数及其解法。一般权重设定中，对应点中权重均设为一个定值。

采样方法 一方面可以在空间位置上进行随机采样，这可以通过将几何体所处区域进行体素化，然后进行等概率地抽取一部分体素中的点；另一种是对几何体离散化的点进行采样，这样保证每次选到的都是几何体上的点；最后为了防止几何体几何结构复杂的情况，还可以使用法向平均采样，它保证均匀地在法向空间中进行采样，这样在变化剧烈的地方采样率就会比平坦区域采样率高。如图 2.4 所示。

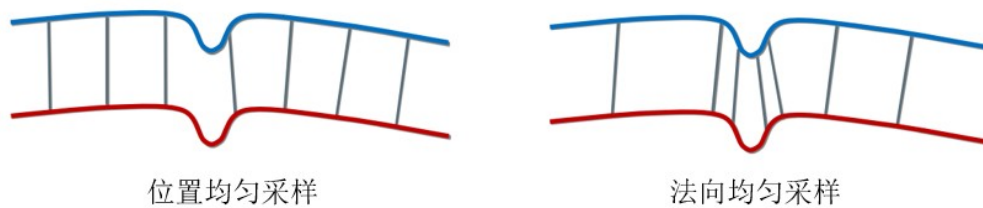


图 2.4 ICP 算法采样方法

点对估计及剔除策略 当我们从 $S(t_1)$ 中随机均匀采样出一个点集后，我们可以从 $S(t_2)$ 建立 KD 树，并从中为前一步骤中的每个采样点找到一个空间距离上最近的点。直接使用这些对应点显然不会估计出一个最优的刚体变换，因为只有那些 $S(t_1)$ 和 $S(t_2)$ 重叠的部分才是有效的对应点。由于这块重叠区域是未知的，我们当然可以将它当作一个未知数放在求解式中，即图 2.3 中的 0-1 权重情形，但这样带来的代价是方程的复杂性。事实

上, Rusinkiewicz 等人^[26]的工作表明, 通常的剔除策略显得更有效和方便。这些策略主要包括: (式中 $v \in \mathcal{S}(t_1)$, $v' \in \mathcal{S}(t_2)$)

- 当对应点的距离大于所有点对距离的中位数, 即剔除满足 $\|v - v'\|^2 > \sigma_{\text{median}}$ 的点对。
- 当 v' 落在 $\mathcal{S}(t_2)$ 的边界上的时候, 源对应点 v 通常不属于 $\mathcal{S}(t_1)$ 和 $\mathcal{S}(t_2)$ 的重叠部分上, 因此需要把满足这种条件的对应点剔除。在实际中, 通常需要结合距离策略, 因为有很多的模型本身是一个拓扑完备的, 则它并没有一个边界点。
- 处理一对对应点 (v, v') 的时候, 还需要考虑其法向是否不相容。即 $n(v)^T n(v') > \sigma_{\text{angle}}$, 当两个点法向相差太大的时候, 它们属于一对对应点的可能性很小, 这是基于 ICP 假设两个模型本身是比较接近的。
- 一方面需要从源对应点中到目标模型中找对应, 另一方面, 双向的对应往往会使得对应查找更加地鲁棒。

评估函数 事实上, 式2.2也称为点到点误差。但通常更有效的方法是结合点到面的误差, 这样可以将每个点的局部信息也考虑进来, 使得在点的切平面上也尽量保持一致, 这样做的好处是允许几何体在平坦的区域进行滑动, 对齐到更准确的位置。点到面的误差定义为:

$$E_{\text{match}} = \sum_{(v, v') \in C} \|\mathbf{T}v - v'\|^2 + \alpha \sum_{(v, v') \in C} \|n_{v'}^T (\mathbf{T}v - v')\|^2 \quad (2.4)$$

其中, $n_{v'}$ 为对应点的法向, α 为控制点到点误差与点到面误差的系数。

由于问题本身是一个非线性非凸的问题, 不少学者提出使用基于高斯牛顿迭代法、LM 迭代方法等近似逼近最优解, 下面简单介绍一种线性的近似方法^[27]。

求解方法 注意到如果 $\theta \approx 0$, 则 $\sin \theta \approx \theta$, $\cos \theta \approx 1$, 于是, 对于一个由三个旋转分量构成的旋转矩阵, 表示成欧拉角的形式即 $R = R_y(\theta_y)R_x(\theta_x)R_z(\theta_z)$, 假定 $\theta = (\theta_x, \theta_y, \theta_z)^T$ 均很小, 则真实的旋转矩阵 R 可以近似为:

$$R \approx \hat{R} = \begin{pmatrix} 1 & -\theta_z & \theta_y \\ \theta_z & 1 & -\theta_x \\ -\theta_y & \theta_x & 1 \end{pmatrix} = I + \begin{pmatrix} 0 & -\theta_z & \theta_y \\ \theta_z & 0 & -\theta_x \\ -\theta_y & \theta_x & 0 \end{pmatrix} \quad (2.5)$$

将上式代入2.4即可将问题线性化展开为:

$$E_{\text{match}} = \sum_{(v,v') \in C} \|t - (v \times \theta) + v - v'\|^2 + \alpha \sum_{(v,v') \in C} \|\theta^T(v \times n_{v'}) + n_{v'}^T t - n_{v'}^T(v' - v)\|^2 \quad (2.6)$$

这是一个可以描述成 $AX = b$ 形式的简单优化问题，为了使得在满足最小化 E_{match} ，可以使用最小二乘或 SVD 分解的方法，求出最优的 θ 后，我们需要用标准的欧拉角方法更新旋转矩阵，以此迭代，求解最优的刚体变换。

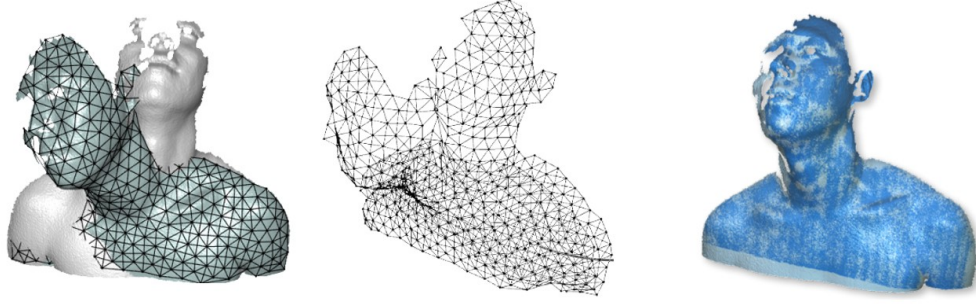
当我们使用 ICP 迭代计算出两帧几何数据的相对变换后，可以将它应用到多帧数据的拼接以形成完整的模型。但是在计算过程中由于难以避免的误差累积，可能会导致最后一帧与第一帧数据拼接的时候无法形成一个闭合回路。关于减小误差和分散误差，有诸多研究工作^[28]，这里不再赘述。

2.3 非刚体对齐方法

由于在实际的三维重建中，不可避免地会出现相机畸变或者物体本身会发生形变，因此我们需要讨论非刚体的对齐方法。针对不同的应用场景，有不同的非刚体对齐方法，Brown 等人^[29]为了解决高精度建筑重建中引入的微小非刚体畸变，使用全局优化的对齐方法处理低频区域的对齐。Huang 等人^[30]处理的是类似于人体等关节形变，因此他假定发生形变的物体局部是不会发生变化的，即整个变形过程是一个保距变换，采用的是一个保距变形方法。Brian 等人^[31]通过扩展 ICP 算法的框架，用非刚体 ICP (NICP) 的方法对源模型的每个顶点赋予一个仿射变换，使用 ICP 一样的流程但不同的方法计算所有的仿射变换，从而达到非刚体对齐的目的。后续的诸多也采用类似于 NICP 的流程，在能量优化上不断改进。下面我们介绍 Li 等人^[3]基于该流程，使用变形图模型计算仿射变换，并运用嵌入空间的方法进行变形。其好处在于它的变形比较适用于自然的弹性变形，在尽可能多地保持局部的细节的同时能够非常高效地计算出变形结果。

为了表现任意的非刚体变形细节，一般非刚体 ICP 方法对每个顶点赋予一个仿射变换，但是这样整个求解参数将是 $12N$ 个，其中 N 为模型的顶点。为降低计算的复杂度，同时处理任意的三维模型，可以使用非线性的嵌入空间的变形方法^[32]如下：

对模型所在的空间的顶点进行采样，并根据欧氏近邻或测地线距离近邻可以将顶点连接成边，这样就可以形成一个模型的简化图模型，如图2.5所示。同时，每个结点赋予一个仿射变换，代表其所在的局部空间的一个变形。由于变形图是均匀稀疏地分散在实际的三维模型表面上，可以使用变形图的仿射变换代替整个模型的变形，我们要求解的就是这

图 2.5 形变图非刚体变形示例^[3]

些简化图上每个顶点的仿射变换了。这样就可以大幅度地降低计算的成本。对于原始模型上的每个点，则可以通过将其 K -近邻的图结点变换进行加权混合，具体描述如下。

记源模型 $\mathcal{S}(t_1)$ 对应的简化的变形图为 \mathcal{G} ，其中简化图的结点记为 $v_i^{\mathcal{G}}$ ，对应的仿射变换为 $A_i = [R_i \ t_i]$ ，则对空间中任一个点 v 其变形结果 \hat{v} 为：

$$\hat{v} = \sum_{j=1}^K w_j(v) (R_i(v - v_j^{\mathcal{G}}) + v_j^{\mathcal{G}} + t_j) \quad (2.7)$$

其中 $w_j(v)$ 表示加权混合的权重，是与距离成正比的标量值，同时使用正则化方法保证每个点的所有权重和为 1，定义如下（ D 为一个距离阈值，一般选为 $K+1$ 近邻的点距离）：

$$w_j(v) = \frac{1 - \|v - v_j^{\mathcal{G}}\|/D}{\sum_{k=1}^K 1 - \|v - v_k^{\mathcal{G}}\|/D}$$

因为变形是作用于图 \mathcal{G} 上，因此式2.1可以写为：

$$\begin{aligned} E_{\text{match}} = & \sum_{(v^{\mathcal{G}}, v') \in C} \|\hat{v}_i^{\mathcal{G}} - v'\|^2 \\ & + \alpha \sum \|R'_i R_i - I\|^2 \\ & + \beta \sum_{i=1}^{N_{\mathcal{G}}} \sum_{v_j^{\mathcal{G}} \in \text{Neigh}(i)} \|R_i(v_j^{\mathcal{G}} - v_i^{\mathcal{G}}) + v_i^{\mathcal{G}} + t_i - (v_j^{\mathcal{G}} + t_j)\|^2 \end{aligned} \quad (2.8)$$

其中第一项表示为 E_{match} ，衡量每个约束点的匹配程度， C 为约束点对的集合。第二项约束仿射矩阵中第一项尽可能地是正交矩阵，以保证大部分的变形都是尽可能地平滑。第三项保证变形图中相邻结点的平滑性，不至于出现大的突变，这里 $\text{Neigh}(i)$ 即第 i 个结点的所有相邻图结点的集合。

从式中可以看出，需要求解的变量个数为 $12N_{\mathcal{G}}$ 个，这与之之前 $12N$ 相比，已经得到大大的减少。注意到式2.8是一个非线性的问题，可以用 **Guass-Newton** 方法求解。具体的求解方法可以参考 **Sumner** 等人^[32] 在 07 年的原始论文，这里不再赘述。

2.4 本章小结

在我们的系统中，几何对齐是一个非常重要的基础步骤，而在三维重建的领域，几何对齐更是一个不可或缺的过程。在本章中，我们介绍了几何数据对齐的方法，包括了几何刚体对齐与非刚体对齐。刚体几何对齐的方法利用 ICP 算法的一个框架，结合不同采样方法、不同对应选择方法及不同的衡量误差的方法，可以产生不同的结果，但总体来说，目的就是为了使几何体的对齐更加准确。非刚体对齐的方法，由于其求解问题本身难度比较大，不同的研究者依据其几何体不同选择不同变形方法进行对齐。我们这里选择的方法适用于自然变形，如人体，皮肤等弹性的变形，它保持局部特性的同时能够以非常高的效率计算出变形的结果。

第3章 SCAPE 模型及姿态数据库扩充

在运动重建过程中，不可避免地我们需要面对一些包含遮挡的不完全扫描数据，但我们的目标是希望从不完全的数据中恢复出一个完整的三维模型。一个预先给定的模板可以给我们提供一个良好的先验知识，在以往的方法中，一个显式的模板需要用如结构光扫描或多视图重建的方法事先将人体模型恢复出来，而后以这个模型作为模板先验。一方面，结构光扫描或多视图重建方法本身具有设备昂贵的缺点；另一方面，我们需要处理的是动态的人体模型，预先的静态扫描往往不易操作。在我们的重建系统中，我们使用了一个参数化人体模型（SCAPE^[4]）在动态重建的过程中隐式地生成模板，同时通过 SCAPE 中的姿态训练结果进行后续的姿态跟踪。鉴于 SCAPE 在我们系统的重要性，本章内容将首先介绍 SCAPE 及其训练方法，而后介绍如何通过姿态传递的方法对训练数据扩充，以达到更好的训练效果。

3.1 SCAPE 介绍

形状补全及人体动画方法（Shape Completion and Animation of PEople）SCAPE 是由 Angelov 等人^[4]于 2005 年 SIGGRAPH 上提出的一个基于统计方法构建的参数化人体模型，其核心在于将人体的形变分解为姿态形变（及其引起的肌肉变形）和体态形变，并分别进行训练。通过 SCAPE，我们可以在指定姿态参数与体态参数的情况下生成一个虚拟的人体模型。

近年来，基于 SCAPE 有不少的研究工作，按其输入大致可以划分为从二维图像中估计三维人体模型和从三维点云数据中提取人体模型。Weiss 等人^[33]在 ICCV2011 中通过使用 SCAPE，可以从包含噪声和遮挡的 Kinect 深度数据中直接生成对应的人体模型。考虑到 SCAPE 只适用衣着紧身的模型，Hasler 等人^[34]则在 SCAPE 的基础上提出从穿衣服的人体扫描数据中抽取其准确的人体体态模型。此外，由于 SCAPE 允许我们在指定姿态及体态参数的前提下生成新的模型，Balan 等人^[35;36]通过随机搜索的方法从多视图的二维图像中估计出三维人体模型。此后，研究人员将其方法进行改进，将人体语意信息包含到



图 3.1 SCAPE 的应用。左图是 Weiss 等人^[33]利用 Kinect 结合 SCAPE 模型生成的三维模型，右图从单张图像中生成一个准确的三维人体模型。

SCAPE 中，使用图像的轮廓信息从单视图图像中恢复出人体模型，并成功应用于虚拟试衣^[37]和图像中人体瘦身^[38]等应用。在我们的系统中，SCAPE 主要用于从三维的点云信息中估计出人体体态参数，并应用于后继的运动跟踪。

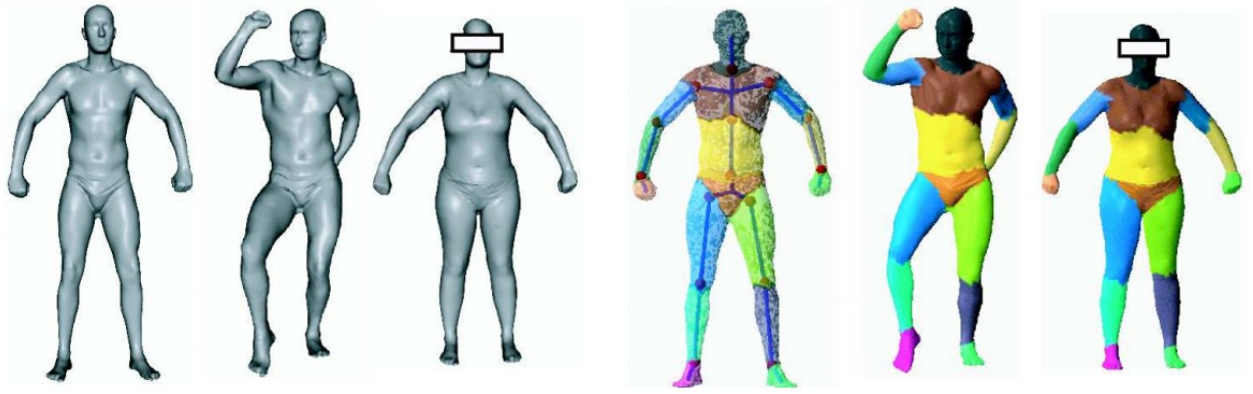
为方便后继的描述，我们预先将符号说明如下：

具体而言，SCAPE 训练数据库中包含一个标准三角网格模型（为方便记为 X ）和其它一系列训练集三角网格模型 $\mathbf{Y} = \{Y^1, \dots, Y^N\}$ （包含姿态与体态，后面以 Y 概指其中一个模型）。注意，SCAPE 要求输入的训练数据已经完全与标准网格模型一一对应起来，并且这些模型本身已经参数化，即每个网格模型点及三角面片均已一一对应，因此 X 和 Y 中所有网格顶点个数均相等，三角面片也一样。而各模型的对应可以通过模板加非刚体对齐的方法来达到要求，可以参考 SCAPE 原文和本文第二章的内容，这里不再赘述。假定 X 包含顶点 $V_X = \{x_1, \dots, x_M\}$ ，训练网格模型 Y 包含顶点 $V_Y = \{y_1, \dots, y_M\}$ ，三角面片使用 $P = \{p_1, \dots, p_P\}$ 表示。

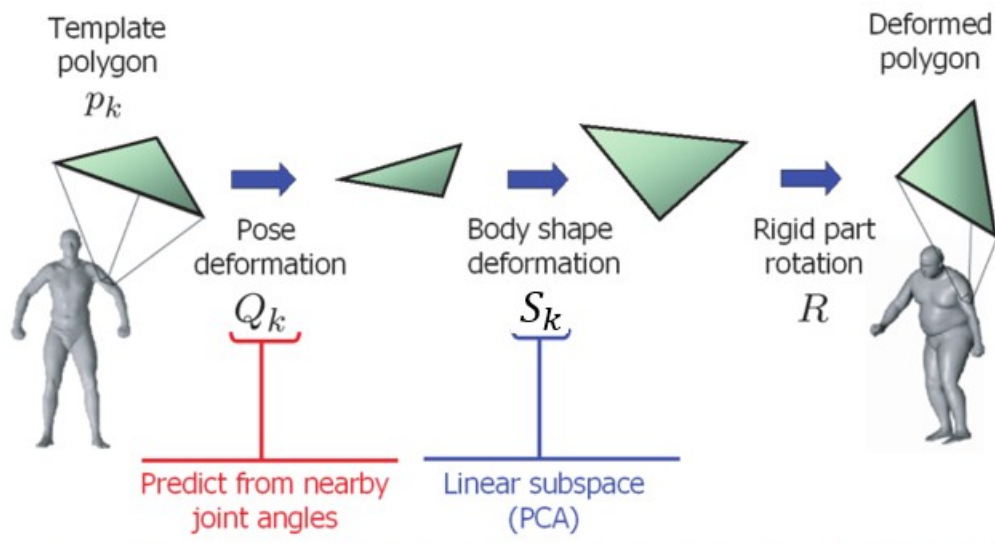
同时，注意到 SCAPE 是隐式地包含了一个骨架系统，因此网格的每个点和面片均已有一个标签，标记着它们属于身体的哪一个部分，如图 3.2 所示，各颜色标记的块即每一个部分，身体中嵌入的虚拟骨骼在计算机动画中经常用到，骨骼中的关节点可以看作唯一标识这个部分的节点。（关于如何从网格模型中如何进行身体部分的划分，可以参考计算机动画中蒙皮相关的论文^[39]）。

对于训练三角网格模型中的第 k 个三角面片，可以定义出一个局部标架为 $[y_3^k - y_1^k \ y_2^k - y_1^k \ n_y^k]$ ，其中 n_y^k 表示该面片法向。于是，从标准模型对应三角面片到当前模型三角面片，可以计算出一个局部变换：

$$T^k = [y_3^k - y_1^k \ y_2^k - y_1^k \ n_y^k] \cdot [x_3^k - x_1^k \ x_2^k - x_1^k \ n_x^k]^{-1} \quad (3.1)$$

图 3.2 SCAPE 中隐式包含的骨骼系统^[4]

SCAPE 将每个 T 分解成每个关节的旋转 R 、由 R 引起关节及肌肉变形 Q 和不同体型所引起的变形 S ，如图3.3，并对于姿态变形和体态变形进行训练，最后可以通过学习的方法生成新的姿态与体态。

图 3.3 SCAPE 中变形的分解^[4]

3.1.1 模型的训练

在描述训练方法之前，我们先考虑在给定每个三角面片的局部变换后，我们如何利用它与标准模型 X 生成对应的模型 \tilde{Y} 呢？由于变换本身全是在局部坐标下进行操作，如果直接将变换应用到每个三角面片，会导致各顶点的不连续，因为每个顶点会被多个面片共享，不同的面片变形后同一顶点可能不会再变换到同一个位置。因此，可以考虑使用

Sumner 等人^[40]在梯度空间进行操作的方法作如下变形:

$$\arg \min_Y \sum_k \sum_{j=2,3} \|T^k(x_j^k - x_1^k) - (y_j^k - y_1^k)\|^2 \quad (3.2)$$

其中 T 作用于标准三角网格模型, $y \in Y$ 为待求网格各顶点。

对于训练数据, 网格模型的每个部分的绝对旋转 R^l 是很容易计算的。如果把每个部分看作一个刚体的变化, 由于模型本身是参数化的模型, 分离出每个部分的一系列对应顶点(标准模型和当前训练模型)后, 可以使用前一章提到的 SVD 分解式 2.3 计算出旋转分量, 即该部分的绝对旋转。

给定每个关节的绝对旋转 R 后, 我们需要进行 Q 的训练, 但事实上 Q 本身是未知的, 因此在训练之前需要进行对 Q 的计算。当我们使用姿态训练数据的时候, 如果标准模型与姿态训练数据是同一个人体的话, 体型是不参与到计算的过程中的, 即此时我们可以认为 S 为单位阵。为了保证属于相同关节的相邻三角面片对应的 Q 是尽可能一致的(即保证光滑性), 利用训练数据 Y 可以使用下式进行 Q 的估计:

$$\begin{aligned} \arg \min_{Q_1, \dots, Q_P} \sum_k \sum_{j=2,3} \|R_{l(k)} Q_k (x_j^k - x_1^k) - (y_j^k - y_1^k)\|^2 \\ + w \sum_{(k_1, k_2) \in \text{Neigh}} I(l(k_1) = l(k_2)) \|Q_{k_1} - Q_{k_2}\|^2 \end{aligned} \quad (3.3)$$

其中 w 用于平衡约束项和光滑项的系数, $I(\text{cond})$ 为示性函数, 当条件为真时为 1, 假为 0。注意到这是一个线性的问题, 可以用稀疏矩阵分解的方法进行最小二乘计算即可求出结果。

一旦我们得到每个姿态训练模型对应的 Q 矩阵, 我们可以用 R 进行训练。由于同一个姿态整体旋转一定角度后, 姿态其实是没有发生变化的, 但对于 R , 其实是发生了变化, 因此为了避免这个影响, 可以使用每个关节的相对旋转作为输入。因为这里 Q 实际上是针对每个三角面片的, 它对应的是每个部分(part), 因此利用每个部分关联的关节相对旋转作为输入。为方便表示, 使用旋转矩阵的指数表示^[41]将旋转矩阵线性化为一个旋转角向量的表示方法, 加上一个偏差项, 形成一个输入向量 $[\Delta r_1 \ \Delta r_2 \ 1]^T$, 这里为方便假定所有输入向量长度均为 7。

于是, 对 Q 的训练即归结为一个简单的线性回归问题。输出结果为 $9N \times P$ 个元素, 构成 Q 矩阵, 而输入为 $7N \times P$ 个元素, 待拟合的回归系数记为 a_k , 则生成一个新的 Q_k 即 $Q_k = \Omega_{a_k}(\Delta r_{l[k]}), \Omega$ 表示的是一个简单的线性组合。

对于不同的体型模型, 当估计出姿态对应的 Q 和 R 后, 同样地, 我们需要为每个模

型先计算 S ，类似于式3.3:

$$\begin{aligned} \arg \min_{S_1, \dots, S_P} \quad & \sum_k \sum_{j=2,3} \|R_{l(k)} S_k Q_k (x_j^k - x_1^k) - (y_j^k - y_1^k)\|^2 \\ & + w \sum_{(k_1, k_2) \in \text{Neigh}} \|Q_{k_1} - Q_{k_2}\|^2 \end{aligned} \quad (3.4)$$

各参数意义同式3.3。

一旦计算出了每个体型模型对应的 S 以后，我们可以用 PCA 的方法对 S 进行主成分的提取，这样就可以得到决定体型变化的最重要的前几个分量。为计算的方便，可以将每个训练模型对应的 S 全部向量化为 $9 \times P$ 的向量。在我们的实验中，前 10 个特征分量已经决定了 95% 以上的信息，假定特征向量表示为 U_1, \dots, U_{10} ，平均值向量为 μ ，这样新的体型可以用体型系数 β 生成：

$$S = \mathfrak{S}_{U, \mu}(\beta) = U\beta + \mu$$

当然，在生成 S 的时候，注意需要将 $9 \times P$ 维的向量形式转化为矩阵形式。

训练完上面的姿态与体态形变后，给定姿态参数 θ 与体型参数 β 新的模型的生成类似于式3.2，只不过这里的 T 发生了变化，新模型的合成即相当于最小化能量 E_Y 定义如下：

$$E_Y = \sum_{k=1}^P \sum_{j=\{2,3\}} \|R_{l(k)} S_{U, \mu}^k Q_k(\Delta\theta)(x_j^k - x_1^k) - (y_j^k - y_1^k)\|^2 \quad (3.5)$$

3.2 姿态数据库的扩充

在我们实验中，使用了 Hasler^[42] 提供的人体数据库，原因在于它包含的体型比较丰富，而 Angelov^[4] 提供的数据库中不包含体型数据，只包含了 71 个姿态模型。在训练的过程中，我们发现使用的数据库中体型数据足够，有 105 个体态模型，但姿态模型只有 34 个。从图3.4中左图可以看出，如果直接使用原始的姿态数据进行训练，势必会导致学习不出来的结果。

考虑到 Hasler^[42] 给出的数据库中人体姿态过少，可能会造成我们训练的样本不足以张成整个人体姿态空间，而事实上在我们的实验中训练的结果也确实不如人意，如图3.4所示。因此，要么对训练的方法本身进行改进，如采用更复杂的非线性训练方法（结合非线性核函数的 SVM 方法等）；要么使用更多的训练样本，因为毕竟 34 个姿态确实不足以覆盖完整的人体姿态空间。考虑到复杂的方法可能会存在不稳定性，同时线性的方法足够简单，我们决定采用增大训练样本这一个途径来提高训练的结果。

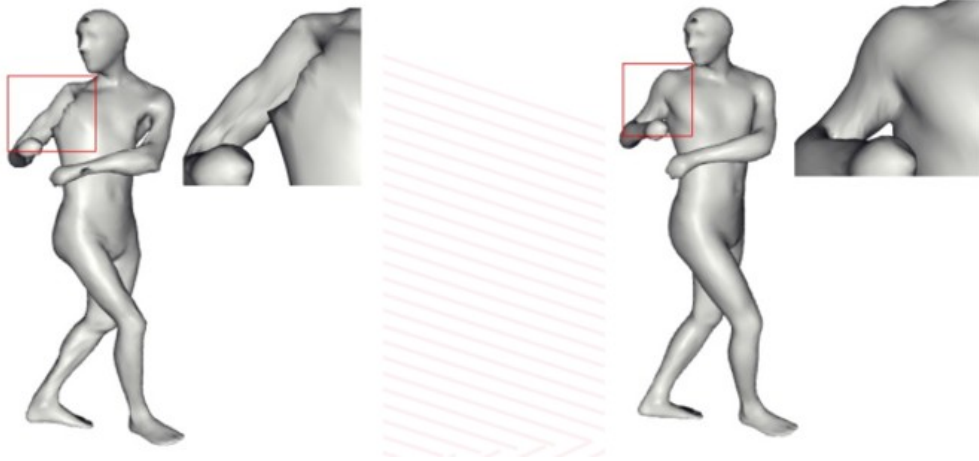


图 3.4 训练数据对学习结果的影响。左图表示使用 Hasler 原始数据库的情况下训练后学习出来的结果，右图是使用扩充后的数据库进行训练学习出的结果。

前面也提到，SCAPE 是一个参数化的人体模型，它要求输入的各训练模型与标准模型完全对应，即意味着我们无法将任意的一个人体姿态模型直接添加到训练的样本中。注意到 Anguelov^[4] 数据库中姿态比较丰富，因此可以考虑将其中的姿态迁移到 Hasler^[42] 数据库中，这样就能使得总的姿态训练模型达到 100 个以上。这里，我们借鉴 Sumner 等人^[40] 提出的形变传递方法，将两个数据库中的模板先通过变形方法使得目标模板中每一个三角面片都能与源模板的三角面片对应起来，从而将源数据库中的运动信息迁移到目标三角网格中。事实上，该方法针对更多数据库的合并依然有效，为方便叙述，我们假定待合并的数据库为 2 个。

将两个数据库分别记为 \mathcal{D} 和 \mathcal{D}' ，其标准模型分别记为 X 和 X' 。我们考虑将 \mathcal{D}' 中的姿态迁移到 \mathcal{D} 中。为了将 \mathcal{D} 和 \mathcal{D}' 对应起来，我们需要先建立起标准模型 X 与 X' 的对应关系。这可以通过将 X 变形到 X' ，然后通过最近邻查找的方法将每个 \mathcal{D} 中的三角面片与 \mathcal{D}' 对应起来。具体的变形方法我们在这里采用 Sumner 等人提出的方法进行：

$$E = w_1 E_{\text{smooth}} + w_2 E_{\text{identity}} + w_3 E_{\text{constraints}} \quad (3.6)$$

其中，各项能量分别用于控制平滑项、正则项、约束项。

首先，变形后的结果 $X_{X'}$ 显然和 X 是能够对应的，所以对于每个三角面片，可以用式 3.1 的方法将其局部变换确定出来，显然 X 是已知，而 $X_{X'}$ 为待求项，从 3.1 可以看出，对局部变换的线性操作即相当于是对各顶点的一个线性组合。于是前面的能量式关于各三角面片的局部变换 T_k 可以分别写成：

$$\begin{cases} E_{\text{smooth}} &= \sum_{p=1}^P \sum_{q \in \text{Neigh}(p)} \|T_p - T_q\|_F^2 \\ E_{\text{identity}} &= \sum_{p=1}^P \|T_p - I_3\|_F^2 \\ E_{\text{constraints}} &= \sum_{(v,v') \in C} \|v - v'\|_F^2 \end{cases} \quad (3.7)$$

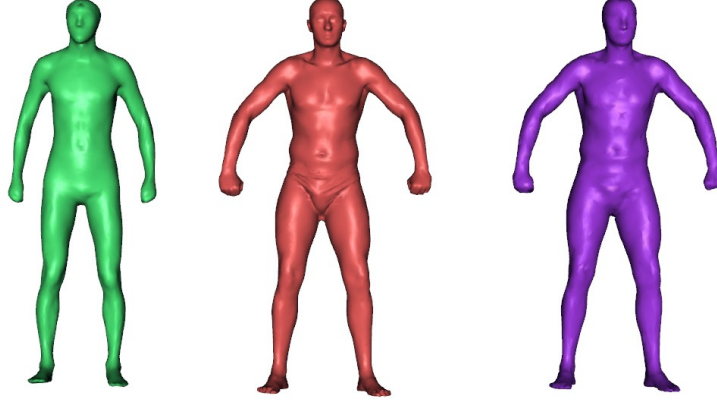


图 3.5 变形结果。以绿色模型为源模型，红色为目标模型，紫色为变形后的结果。

其中 C 表示约束点，即 X 与 X' 上一系列的对应点。这样， E 就是关于未知变量 V （网格各顶点）的一个线性问题。可以直接用稀疏矩阵分解的方法进行求解。但是注意到问题的关键在于 C ，一开始我们是没有任何对应信息的，因此，可以考虑先手动地选择几十个初始对应引导方程进行变形，然后再不断调整各项系数 w 并采用最近邻的方法寻找更多的对应点。这样，当变形后的结果 X_X' 与 X' 形状一致后（如图 3.5），我们就可以直接采用最近邻的方法建立起三角面片的对应了。

给定两个标准网格模型 X 与 X' 的对应点 C 后，我们就可以将 \mathcal{D}' 中的姿态迁移到 \mathcal{D} 中了。因为对 \mathcal{D}' 中的每一个姿态模型，相对于 X' 每个三角面片都有一个局部变换 T'_k ，因此可以为 X 赋予一个与 T'_k 尽可能一致的局部变换，然后与此同时保持 X 本身的局部结构不变。这可以通过最小化如下能量变形得到：

$$E_{\mathbf{t}} = \sum_{i=1}^P \|T_i - T'_i\|^2 + \gamma \sum_{i=1}^P \sum_{j \in \text{Neigh}(i)} \|T_i - T_j\|^2 \quad (3.8)$$

其中 T'_i 为 \mathcal{D}' 中与 X 的第 i 个三角面片对应的三角形的局部变换，后一项中 j 为所有与 i 面片近邻的三角面片。 γ 用于平衡两项能量式。

另外还需要注意的一点是，在传递姿态的时候，这里我们假定的前提是 X 和 X' 具有相似的姿态。如果两个数据库中姿态比较多，是比较容易找到相似姿态作为参考模型的，但是如果两个数据库中标准姿态相差很大，我们需要先作一个预处理过程。即先把其中一

个姿态模型用蒙皮或变形的方法变到具有与另一个模型相似姿态的模型上去，然后再使用上述的能量式进行传递变形。这里就不再赘述。

图3.6是我们进行姿态变形的结果，注意到右图绿色模型实际上是不存在于最初的 Hasler 数据库中。通过将 Anguelov 数据库的姿态进行迁移，我们可以看到姿态很好地传递到了 Hasler 数据库的标准模型上，同时，局部的模型是得到了保持的。

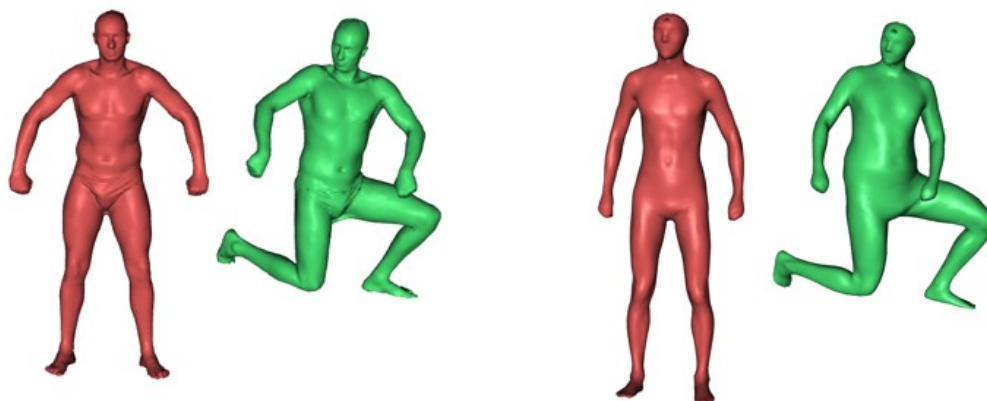


图 3.6 姿态迁移结果。红色模型为 \mathcal{D} (右) 和 \mathcal{D}' (左) 中的标准模型，绿色左图表示 \mathcal{D} 中的一个原始姿态，绿色右图表示将左图姿态前移后的结果。

3.3 本章小结

在本章内容中，我们详细介绍了 SCAPE 模型的一些相关工作，随后介绍了其训练方法。在后一部分，我们在原始姿态训练数据不足的情况下采用变形的方法将其它数据库中的姿态迁移到我们的训练数据库中，并取得了比较良好的效果。在下一章的实验结果部分，我们将采用真实数据对 SCAPE 模型数据库的扩充进行评估，以证明我们的方法的确是有效。

第4章 人体运动重建系统

我们的硬件系统如图4.1所示。为了在数据的精度与采集空间范围之间作一个权衡，我们的系统包含4台 Kinect 相机（红色矩形框中所示）。其中2台用于获取正面数据，2台用于获取背面数据。位于同一侧的相机分别用于拍摄人体上半部分及下半部分的数据。为避免同侧相机间红外线的相互干扰，我们调整单侧相机使得其视角尽可能没有重叠。同时，为获取完整的运动数据，表演者离两侧相机水平距离均为1.5米。

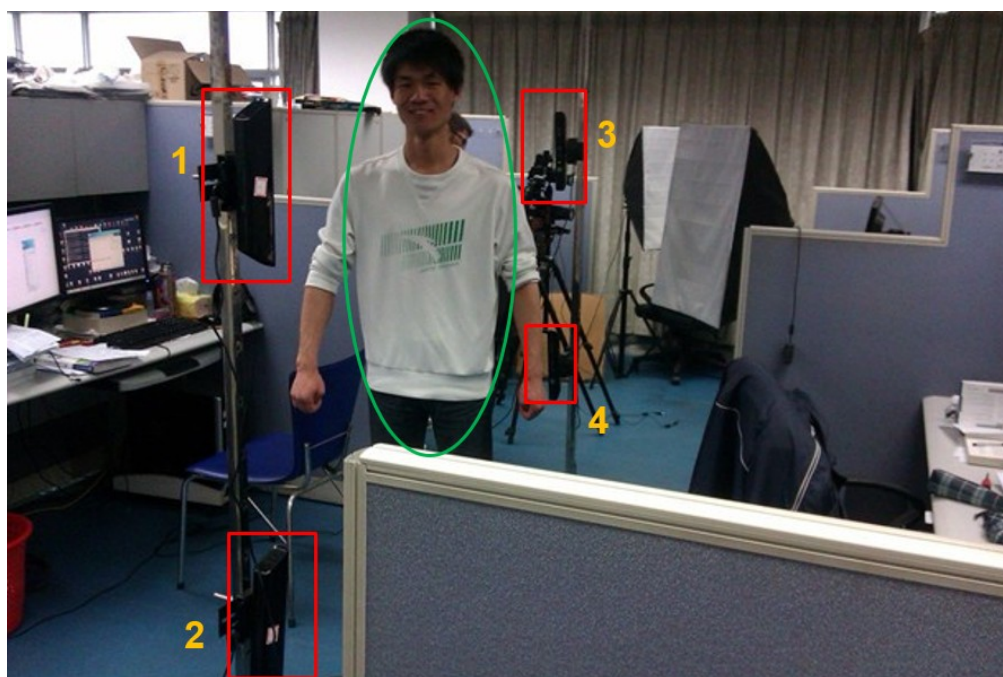


图 4.1 硬件系统包含四个相机，表演者在中间的区域进行表演

在本章接下来的部分，我们将首先介绍系统的相机定标，然后介绍使用 **SCAPE** 模型进行隐式模板的估计，并利用该模板使用 **SCAPE** 及分段刚体对齐的方法进行大尺度姿态跟踪，而后利用拉普拉斯变形算法作细节的增强，最后通过时空滤波达到完整的运动重建过程。

4.1 定标方法

为将多视图下的点云统一到全局的坐标系下，需要对系统各相机之间的相对位置进行标定。事实上，相机标定是计算机视觉中一个经典的问题，包括单个相机内参，双相机外参的计算^[43]以及多视图下各相机内外参的计算^[44;45]等等。其中，^[43]使用棋盘角点对应的方法同时可以计算出两个相机的内外参，这种方法优点在于操作简单，而且比较准确。而^[44]要解决的是大型的 LightStage 系统中数百个相机参数的标定，不同于棋盘方法，它使用 LED 激光灯进行定标，要求背景尽量为黑色，当 LED 灯亮的时候就可以很容易检测出 LED 灯对应像素的位置，从而准确地将多相机对应点找出来计算参数。

4.1.1 初始外参计算

由于我们的系统中涉及到四个相机，其中两两朝向是对立的。如果简单地使用棋盘定标方法，必然会导致棋盘在两个对立相机中一个可见一个不可见，此外，两两定标出来的结果将会导致有一个定标结果是冗余的。考虑 $V_1 \rightarrow V_2 \rightarrow V_3 \rightarrow V_4 \rightarrow V_1$ 这样一个两两定标的问题， $V_3 \rightarrow V_1$ 就可以通过 $V_3 \rightarrow V_2 \rightarrow V_1$ 和 $V_3 \rightarrow V_4 \rightarrow V_1$ 两种方式计算得到，如果所有的数据都是完全精确的，理论上说两种结果是一样的，即出现了冗余。事实上，由于数据本身的噪音及定标误差影响，两种计算出来的结果会出现不一样，因此我们希望最好 $V_3 \rightarrow V_1$ 能够综合平衡这两种方式得出的结果。考虑 LED 灯多视图定标方法，对于我们当前的设备来说，这是一种不太适合的方法，毕竟该方法本身就是设计用于像 Light Stage 这样复杂的动态光场采集系统，其中的灯光照明是可控的，因此定标时使背景全黑是可行的；但对于我们的系统，显然不可取。此外，该算法是针对数十数百个相机的系统进行标定，其本身造成的计算误差必然会不断地累积。

由于我们使用的不同于传统的 RGB 相机，Kinect 可以直接获得每个像素对应的深度值，即我们可以计算出每个像素点的准确坐标（相对于这个相机）。为了充分地利用这些信息，受棋盘定标方法和^[22]的启发，我们设计了一种实用的方法，用于我们的采集系统的定标，介绍如下：

从一幅 RGB 图像中检测出棋盘的角点是很容易的，可以使用检测算法先检测出每个图片中棋盘角点的位置 p ，由于 Kinect 的特殊性，可以将 p 对应到其三维坐标 P 。由于每两两进行棋盘标定的时候棋盘角点的对应也是隐含地可以知道，于是对于每次的两个视图下的定标图片，我们就得到了一系列的对应点集 \mathcal{P}^S 和 \mathcal{P}^T 。考虑到棋盘本身是一个平面，容易造成误差集中在某一个方向，因此可以采集多个角度下的棋盘数据，使得生成的对应点集 \mathcal{P}^S 和 \mathcal{P}^T 能够对采集区域的空间进行足够的采样。使用第二章式2.3提到的 SVD

分解方法，可以准确地计算出点集 \mathcal{P}^S 到 \mathcal{P}^T 的刚体变换，即两个视图下得到的点云的刚体变换 $\mathbf{T} = [R \ t]$ ，从而也即得到两个相机的外参。

对于同一侧相机，使用上面的方法是足够而准确的，为了避免同一侧相机红外的干扰，可以通过依次挡住两个相机的红外发射器来达到要求。但是，前面提到两个朝向相反的相机会导致棋盘图像在一个相机中不可见，为了解决这个问题，我们设计使用一张印有两面棋盘格的硬纸板对朝向相对的相机进行标定，其中纸板上两面的棋盘格是完全重叠对应的。同时，我们要求纸是足够薄的，这样我们可以忽略其厚度，意味着对应点在三维空间中为同一个真实点，从而只需要把两个棋盘角点对应起来，利用前面的方法就可以对这对相机进行定标了。为叙述方便，记两个相机对应的角点从左到右从上到下依次记为 $c_{i,j}^1$ 和 $c_{i,j}^2$ ，显然对应点不再是依次对应，点 $c_{i,j}^1$ 在图像 2 中的对应点为 $c_{i,N-j}^2$ ，其中 N 表示图像中角点特征的列数，如图 4.2 所示。

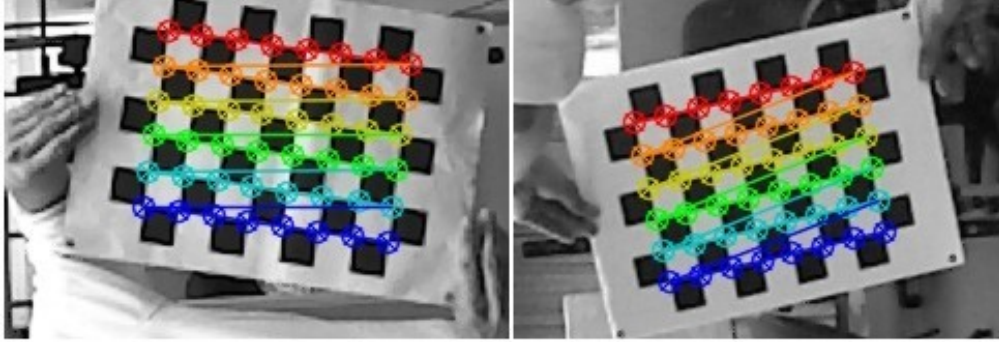


图 4.2 用于朝向对立的相机定标的棋盘格

我们记四台相机分别为 V_1, \dots, V_4 ，其中 V_1 和 V_2 为同一侧相机，并且 V_1 与 V_4 相对， V_2 与 V_3 相对。利用前面提到的定标方法，我们可以依次求出 V_i 到 V_{i+1} ($i = 4$ 时 $i + 1$ 为相机 1) 的变换 \mathbf{T}_i 。为了统一到一个全局坐标系下，可以以 V_1 为标准，这样利用矩阵相乘即可求出其它相机的外参。

4.1.2 外参的全局优化

前面小节一开始我们提到在系统中希望最好 $V_3 \rightarrow V_1$ 能够综合平衡两种传递计算方式 ($V_3 \rightarrow V_4 \rightarrow V_1$ 和 $V_3 \rightarrow V_2 \rightarrow V_1$) 得出的结果；此外，从上一小节中我们得到连续的两两相机之间的变换，但最终我们是需要把所有相机中的数据对齐到一个全局的世界坐标系下，假设以相机 1 为世界坐标系原点，则其它相机到该原点的变换需要通过一系列变换的传递求解出来，显然，在这个传递的过程中是会出现误差累积的。因此有必要对定标出来的结果作全局的优化。

Sharp 等人^[28]从理论上分析并提出一种将误差分散的方法,使得最终得出来的一系列变换满足闭合回路的性质,但是却没有考虑计算变换时两个方向传递结果的平衡。因为它只是单纯地满足了一个闭合回路的性质,而且该算法在计算过程中需要选择特殊的点和权重,在我们的实验中反而是不合理的,将导致最终接合的结果膨胀开来。因此,我们提出一种新的基于优化的方法对上一节求解出来的外参进行全局的优化。

为了叙述的通用性,我们假设系统中有 N 台相机 ($V_1 \dots V_N$) 构成一个环形系统需要进行定标。而且它们已经通过前面棋盘角点对应加 SVD 分解的方法计算出了 V_i 到 V_{i+1} 初始的相对变换,通过传递关系将它们全部变换到以 V_1 为世界坐标系下的变换,显然这样得到的变换的误差会不断地增大,记我们得到的这些全局坐标系下的变换为 $\mathbf{T}_1, \dots, \mathbf{T}_N$, 其中 \mathbf{T}_1 显然为单位阵。

显然,对于每一对前面棋盘方法得出的对应点集合 $(\mathcal{P}^S, \mathcal{P}^T)$, 则需要满足如下约束:

$$(\mathbf{T}_1 \dots \mathbf{T}_N) = \arg \min \sum_{i=1}^N w_i \sum_{(p,q) \in (\mathcal{P}^S, \mathcal{P}^T)} \|\mathbf{T}_i \hat{p} - \mathbf{T}_{i+1} \hat{q}\|^2 \quad (4.1)$$

其中, (p, q) 表示对应点集中的一点对, \hat{p} 表示齐次坐标的形式, w_i 作为每一项的权重。注意到 \mathbf{T} 是包含一个正交变换 R 和一个平移变换 t 的, 因此上述方程是非线性的。

为求解出一个满足上式的最优变换,注意到我们已经得到的初始变换实际上是已经比较接近真实值的,所以可以采用迭代的方式每次求解变换的一个增量 $\Delta \mathbf{T}$, 记初始变换为 \mathbf{T}^0 , \mathbf{T} 可以表示为 $\Delta \mathbf{T} \mathbf{T}^0$, 然后应用到初始变换 \mathbf{T}^0 , 再代入到上述式中求解 $\Delta \mathbf{T}$, 如此迭代直到最终的能量式收敛。对于 $\Delta \mathbf{T}$, 其中的正交矩阵可以分解为类似于^[27], 使用弧度来近似代替真实的旋转变换(因为当 $\theta \approx 0$ 时, 有 $\sin \theta \approx \theta$ 成立), 设每次增量对应的旋转欧拉角分别为 α 、 β 和 γ , 则

$$\Delta \mathbf{T}(\alpha, \beta, \gamma, \Delta t_x, \Delta t_y, \Delta t_z) = \begin{pmatrix} 1 & -\gamma & \beta & \Delta t_x \\ \gamma & 1 & -\alpha & \Delta t_y \\ -\beta & \alpha & 1 & \Delta t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (4.2)$$

则能量式4.1变为:

$$(\Delta \mathbf{T}_1 \dots \Delta \mathbf{T}_N) = \arg \min \sum_{i=1}^N w_i \sum_{(p,q) \in (\mathcal{P}, \mathcal{Q})} \|\Delta \mathbf{T}_i \mathbf{T}_i^0 \hat{p} - \Delta \mathbf{T}_{i+1} \mathbf{T}_{i+1}^0 \hat{q}\|^2$$

当每次更新 \mathbf{T}^0 的时候, 我们不使用式4.2更新, 而是将 $[\alpha \ \beta \ \gamma]$ 代入到真实的对应欧拉角矩阵中作为 $\Delta \mathbf{T}$, 具体公式参考^[27], 这里不再赘述。表 4.1 为改进后的方法与之前

方法的对比结果，可以看到，在平均误差一栏中， $V_2 \rightarrow V_3$ 的误差原本超过 1cm 误差，但优化后使得这个误差大大降低了。

表 4.1 定标反投影误差对比结果，每一格中左为原始反投影结果，右为优化后结果。单位为米。

	$V_1 \rightarrow V_2$	$V_2 \rightarrow V_3$	$V_3 \rightarrow V_4$	$V_4 \rightarrow V_1$
最大误差	0.051708/0.050945	0.038821/0.024134	0.030385/0.031345	0.037101/0.039082
平均误差	0.006900/0.006935	0.013710/0.004874	0.006409/0.006541	0.006807/0.007230

4.2 模板生成

完成定标后，我们就可以把每帧的各角度数据拼接到一起，形成在全局坐标系下统一的点云。接下来的处理中，我们把所有帧的点云数据集合记为 $\mathcal{D} = \{D^1, \dots, D^n\}$ ， n 表示帧数，而 D^t 指第 t 帧的完整点云数据。

由于在人体运动重建过程中，模板可以为大量包含噪声和遮挡的数据提供一个很好先验知识，因此需要生成一个与扫描数据相对应的模板。但从第一章中，我们也看到，要预先生成一个精细的静态模板是不太容易的。幸运的是，前面提到 **SCAPE** 模型通过参数化人体姿态与体态变形，可以利用姿态和体态参数自动生成一个新的模型。一个很自然的想法即通过给定我们的点云，生成对应的 **SCAPE** 人体网格模型，并以此作为隐式的模板。首先，由式 3.5 知，要利用该式生成一个模型，需要预先知道矩阵姿态变形和体态变形矩阵，即意味着需要事先知道二个变形参数。但关键在于我们没有办法事先得到这两个参数，我们的输入只有稠密的无对应点云数据，而且只能利用这个输入生成姿态参数 θ ，体态参数 β ，并用它们生成新的满足 **SCAPE** 性质的模型 Y ，注意到 Y 是在梯度空间中生成的，可能与约束点云不完全对应，因此还需要利用点约束（记为 C ）求出新的模型 \hat{Y} 使得它既尽可能保持 **SCAPE** 性质又满足点的约束。具体而言，我们需要解决一个如下式的问题：

$$\arg \min_{\hat{Y}, \beta, \theta} \sum_{y, y' \in C} \|y - y'\|^2 + E_Y \quad (4.3)$$

其中， C 为约束点， E_Y 的定义即式 3.5。

注意到式 4.3 包含三个未知参数 \hat{Y} 、 θ 和 β ，而且该问题是一个非线性非凸的问题，无法用常规的线性解法求解，这里我们采用 Anguelov 等人^[4]提出的类 EM 算法，迭代地固定两个变量求解另一个变量，具体的数值解法可以参考 Anguelov 的博士论文^[46]，这里不再赘述。在我们的实验中，首先固定 θ 为标准姿态，即相对旋转角全为 0， $\beta = 0$ 表示初

始化为平均体型。然后依次反复迭代求解 \hat{Y} 和两个参数。如图 4.3 所示，给定输入数据后，姿态和体型都很好地与原始数据相匹配。

利用上式，我们以第一帧数据 D^1 作为约束输入生成一个体型与姿态符合我们扫描数据的模板模型。但是，注意到式中要求约束点 C 作为已知数据，但我们实际上是没有这个对应信息的。因此，首先，我们需要手工地在 **SCAPE** 标准模型 X 与 D^1 上标注十几个对应点（关节上标注），迭代生成初始的模板后，我们需要在这个模板上使用最近邻的方法从输入数据中寻找更多的稠密对应，使得生成的结果与原始数据更加地匹配。因此我们这里采用的策略与 **ICP** 算法有点类似，即先生成一个模型后，寻找对应点，然后再利用这些对应点生成更准确的模型，以此迭代三四次基本上可以达到比较合理的结果。我们记对应为 C ，并记生成的 **SCAPE** 模型为 Y^1 。显然直接使用 Y^1 当作我们的隐式模板是可行的，但注意到后面我们需要重建出扫描数据的细节，而 **SCAPE** 模型本身一般是分辨率非常低的，因此我们需要对生成的 Y^1 进行上采样，增加其分辨率，同时与 Li 等人^[2] 方法类似，我们需要先将 Y^1 进行平滑，使得它不牵扯到模型的细节内容，但又保持模型的体型姿态等大尺度的变化。

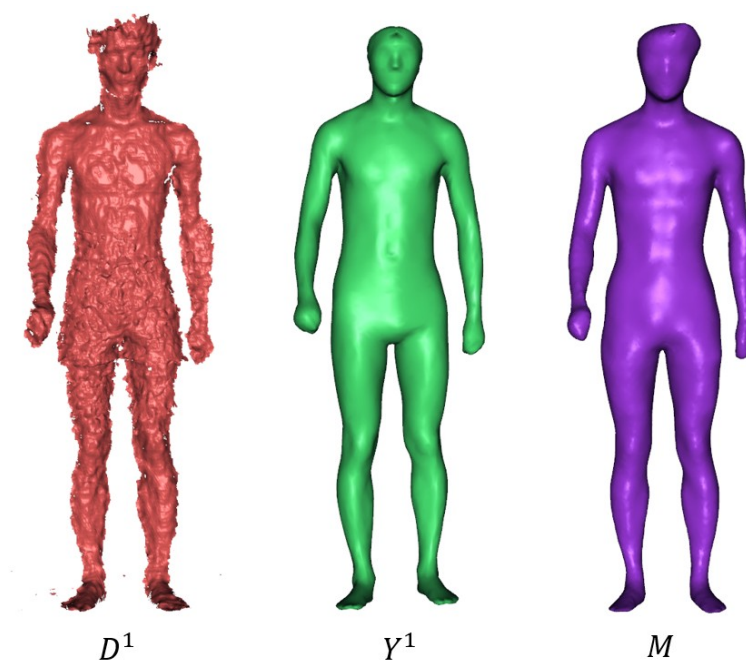


图 4.3 利用第一帧数据生成隐式模板。左图红色为输入数据，即扫描过程中第一帧数据；中间绿色为生成的 **SCAPE** 模型，即正文中的 Y^1 ；右紫色图为生成的模板 M 。

在我们实验中，由于要处理的往往是穿衣服的模型，这时，我们还需要把一些大尺度衣服变形加入到模板中，因为在这种情况下， Y^1 与 D^1 只是一个整体的相似，我们需要把上采样后的结果作非刚体变形，这个变形过程即后面的人体重建细节增强，后面将会提到。

然后再进行平滑，去除高频细节，如图4.3所示。

我们把得到的上采样并作平滑后的模板记为 M 。在后续的姿态跟踪过程中，我们只是进行姿态参数 θ 的估计，因为一个人的体态在其运动过程中是不会更改的，因此我们保持固定 β 。

4.3 运动跟踪算法

如前所述，我们的输入是由四台 Kinect 深度数据转换拼接而成的运动几何的稠密点云，而我们的目标是重新估计出一个拓扑完备的运动人体模型，使这个运动人体模型在每一帧都可以与稠密点云尽量接近。前一节我们已经得到一个与第一帧对齐的模板模型，接下来我们需要利用 M 和 SCAPE 模型对整个运动过程进行跟踪重建。从式 4.3 可以看到，它包含了对于姿态的估计，因此我们只需要给定准确的模板与输入信息的对应，就可以利用最小化式 4.3 达到姿态跟踪的目的。

对于 Kinect 点云数据，我们没有办法预先知道相邻帧几何体的对应关系，因此几何跟踪算法的关键在于对应点的估计。事实上，对于运动几何的重建有很大一部分工作是在解决准确对应点建立这个问题。有一些学者^[15;47]提出使用 RGB 信息上的特征点对应或光流跟踪等；也有一部分学者^[2]在几何信息上进行对应建立，考虑到相邻帧的动作不会很大，他们提出使用最近邻的方法先建立一个粗糙的对应，而后不断地迭代优化的方法将相邻帧的对应准确地对应起来。我们的目标也是希望能够在纯粹的几何信息中建立一个准确的对应，但是我们的实验表明，最近邻的对应是不够的，如图4.6所示，由于不准确的对应导致式 4.3 的误差不断地累积，从而导致出现不自然的变形，因此我们尝试通过另一个思路解决这个问题。我们将模型的整体姿态调整到与点云真实姿态一致，通过这个预处理后再采用最近邻找到准确的对应。

注意到 SCAPE 模型的训练过程中，我们使用了关节的旋转分量作为训练的参数，我们可以通过 SCAPE 本身已经包含了一个隐含的骨架信息解决这个问题，如图 3.2。考虑将 SCAPE 模型按其关节进行划分，将身体的每一个部分的运动先看作是一个整体上刚体的变化，这样就可以先把姿态的变化进行简化，问题就降低为估计每个关节的一个刚体变换 T ，总共只需要估计十几个刚体变换参数。总结成一个能量式即：

$$\arg \min \quad w_1 \sum_{i=1}^M \|T_{[i]}v_i - v'_i\|^2 + w_2 \sum_{(i,j) \in Q} \|T_i J_{ij} - T_j J_{ji}\|^2 \quad (4.4)$$

式中 w 为平衡系数， T_i 为各部分对应的刚体变换，而 J_{ij} 表示属于第 i 个部分的连接 i - j 两个部分的关节点，而 J_{ji} 表示属于第 j 个部分的连接 i - j 两个部分的关节点，前一项作为约

束项控制对应点应该变换到与目标点尽可能接近，后一项则用于控制两个变换仍然能够将两个相邻关节 J_{ij} 和 J_{ji} 粘在一起。

事实上，这个方法在以前学者的研究工作^[48]中也有所涉及，称为关节 ICP 方法。传统的关节 ICP 方法用于解决多视图的运动捕捉系统中人体运动的跟踪问题，因为不需要任何的标记或绑定设备，所以是一个全自动化的跟踪方法。传统的方法中，通过构建多视图下的可视锥，然后跟踪这个三维可视壳计算出每一帧的关节变化情况。对于式4.4，传统的方法使用全局优化的方法解决这个非线性非凸的问题，Corazza 等人^[49]使用了 Levenberg-Marquardt 最小化方法来逼近其解，而 Ganapathi 等人^[50]将这个工作推广到实时的跟踪问题中，并使用概率图模型的方法重新形式化这个问题。我们这里，不同于传统的求解关节 ICP 的方法，使用一个层次式的求解方法，描述如下图4.4：

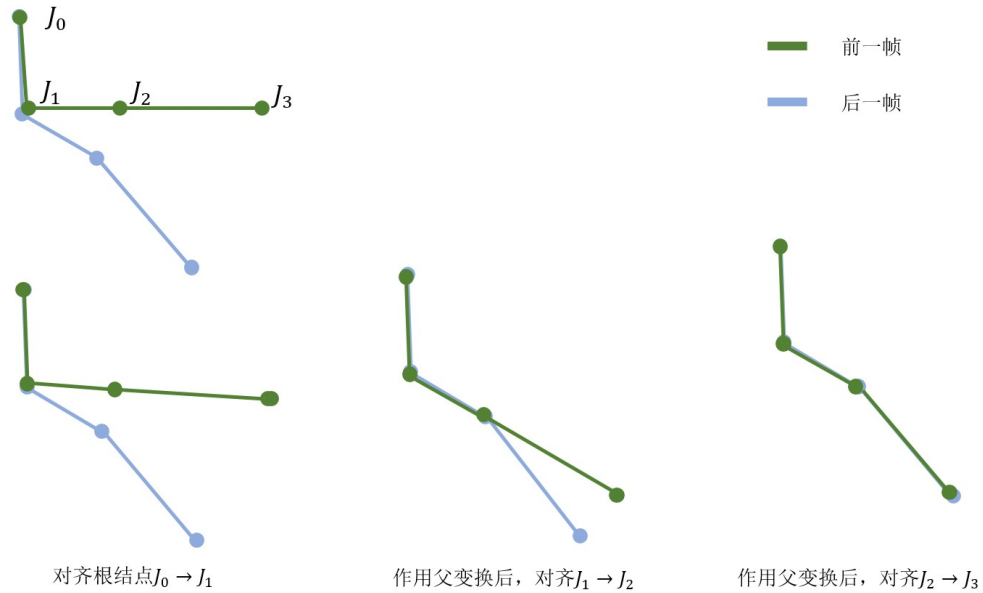


图 4.4 层次式关节对齐方法示意图。绿色作为前一帧的结果，蓝色作为下一帧的输入数据，我们需要将绿色与蓝色对齐。传统方法对这个问题使用一个全局的估计，我们使用分层的方法，先估计出父结点的变换，然后将父结点的变换应用到所有的子结点后，再估计子结点的变换。

从根结点开始，假定初始变换为单位阵，可以通过下式进行求解：

$$\arg \min_{R,t} \sum \|R(v-c) + c + t - v'\|^2 + w\|t\|^2 \quad (4.5)$$

其中， (v, v') 表示的是一个关节对应的所有相关顶点对。当这个关节对应的变换求解出来后，可以将它视为父变换 R_p ，将它应用到该关节相关联的所有子孙关节，然后继续利用上式依次求解其子关节。

当给定父关节的变换 $T_p = [R_p \ t_p]$ 和当前节点变换 $T = [R \ t]$ 后，我们需要利用上式分别求解其子关节各变换，同时我们可以将其子关节对应的变换更新为

$T = [RR_p \ R(t_p - c) + c + t]$, 其中 c 为当前结点对应的部分中心。重复递归执行这个过程, 直到所有的结点变换都估计完成。记这个过程对齐的模型为 Y_A^t 。

虽然 Y_A^t 与输入点云对齐, 但这只是对每个部分的对齐, 在关节点位置, 会出现严重的变形, 因此我们并没有直接使用这个估计的结果, 而是使用它与原始点云作最近邻点的估计, 找出一系列准确的对应。考虑到关节点附近的不准确, 我们只对非关节点附近的点作最近邻的查找。通过这种方式, 我们可以得到一个比较准确的对应点, 使用式 4.3 我们就可以估计出相对应的 SCAPE 模型 Y^t 了。实验结果见下一章的结果部分。

4.4 运动重建

为从包含噪声和遮挡的 Kinect 数据中重建出一个拓扑完整的三维运动几何模型, 当我们通过前面描述的方法达到跟踪的目的之后, 还需要重建出完整的几何模型。本节介绍我们的运动重建方法。

对于每一帧, 当我们跟踪并估计出对应的 SCAPE 模型 Y^t , 我们需要将 Y^t 的姿态映射到我们的模板 M 上, 进行一个大尺度的变形。事实上, 这与前一章数据库扩充中姿态传递非常相似, 我们可以通过将 Y^t 的姿态传递到 M 达到大尺度姿态变形的目的。另一方面, Y^1 与 M 在上一节中已经建立好对应关系, 我们可以直接使用这个方法进行变形。

但是, 注意到式 3.8 是在梯度空间中进行的, 当每个顶点坐标在优化过程中, 会发生一定的偏移, 造成形变后的模板模型可能会与点云之间的对齐出现一些问题。因此, 我们改写式 3.8 如下:

$$\arg \min_{\mathbf{v}} E_t + \gamma \sum_{(\mathbf{v}_i, \mathbf{v}'_i) \in C^t} \|\mathbf{v}_i - \mathbf{v}'_i\|^2 \quad (4.6)$$

其中 γ 用于控制形变传递与顶点约束的结果。

当我们完成姿态的变形后, 我们需要把扫描出来的细节加到大尺度变形后的模板上, 这与 Li 等人的工作^[2]有些类似, 只不过他们使用的沿着每个顶点的法向计算准确的细节偏移, 这要求输入模型具有光滑准确的特性, 显然我们的数据本身就包含大量的噪声, 导致这个方法的不可取。事实上, 我们使用的是一个渐近式地拉普拉斯能量优化解决方案。其能量定义如下:

$$\arg \min_{\mathbf{v}} w_1 \|L\mathbf{v} - \delta\|^2 + w_2 \sum_{(\mathbf{v}_i, \mathbf{v}'_i) \in C} \|\mathbf{v}_i - \mathbf{v}'_i\|^2 \quad (4.7)$$

其中 L 表示正切拉普拉斯矩阵, 而 δ 表示微分坐标, C 表示的是使用最近邻对应方法得到的约束点对。

对于其中的参数,我们初始化平滑项系数 $w_1 = 1.0$, 正则项系数 $w_2 = 0.01$, 然后每次迭代的过程中保持平滑项系数不变, 正则系数更新为 $w_2 \leftarrow 5w_2$ 。

当上面这一步完成之后, 注意到重建结果在相邻帧的很多区域是会有一些变形跳变, 我们需要进行一个时序上的平滑, 保证重建结果在时序上的连续性。我们使用双边滤波的方法对相邻帧的结果进行平滑, 记第 t 帧结果第 i 个顶点为 \mathbf{v}_i^t , 则新的顶点坐标更新为:

$$\mathbf{v}_i^t = \mathbf{v}_i^t + \left(\frac{\mathbf{v}_i^{t-1} + \mathbf{v}_i^{t+1} - 2\mathbf{v}_i^t}{4} \right) \exp\left(\frac{-\|\mathbf{v}_i^{t+1} - \mathbf{v}_i^t\|^2 - \|\mathbf{v}_i^{t-1} - \mathbf{v}_i^t\|^2}{\sigma^2}\right) \quad (4.8)$$

其中 σ 在这里设置为 $\sigma = 0.05$ 。

4.5 实验结果

表 4.2 实验环境

操作系统	Windows 7
编程语言	Microsoft Visual Studio 2010, C++
CPU	2.8GHz
内存	4GB RAM

我们的实验环境如表4.2所示, 不考虑 SCAPE 模型的离线训练耗时, 整个重建过程中, 算法的时间瓶颈主要是在变形这一步, 大致需要 20 秒一帧。具体而言, 姿态的传递大致需要不到 5 秒的时间, 最终的细节变形需要十几秒一帧。下面将从多个方面评估整个重建系统。

4.5.1 对比实验

数据库扩充 首先, 我们通过真实数据表明 SCAPE 姿态数据库扩充的必要性, 如图 4.5所示。如果仅仅使用原始的姿态数据库训练结果进行学习, 当姿态与数据库中的姿态相差太远的时候, 是难以学习出好的结果的。即使整体的姿态是正确的, 但对于关节处, 还是会出现比较大的不自然变形结果。相反, 使用了扩充后的数据库进行训练后, 学习重建出的结果是比较地合理和平滑的。由此, 也可以看出我们使用扩充数据库的必要性。

关节 ICP 算法 其次, 我们验证关节 ICP 算法的有效性和必要性。如图 4.6所示, 如果没有关节 ICP 的预对齐, 错误的对应会导致左膝部位和脚处出现不自然的形变。尽管

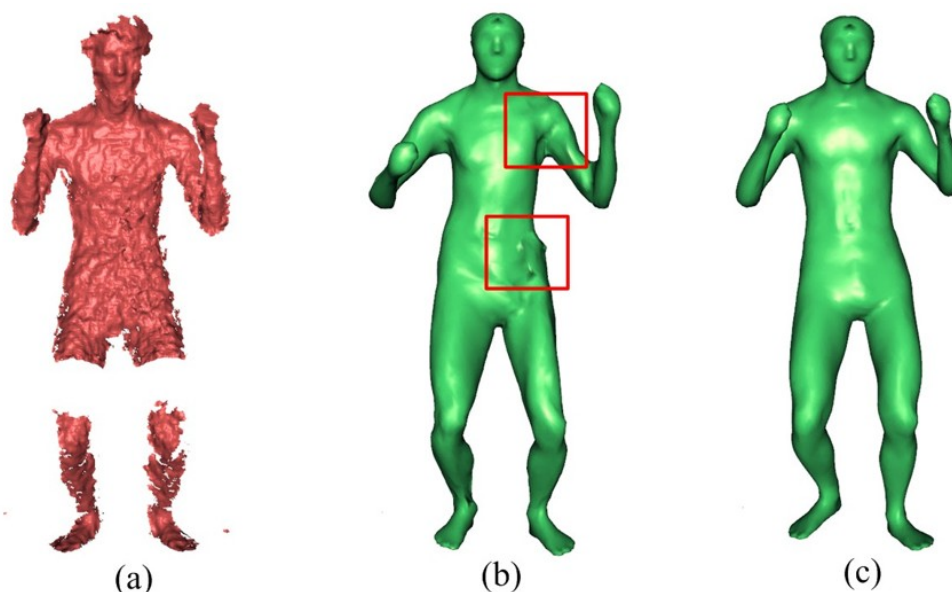


图 4.5 数据库扩充前后真实学习结果对比, (a) 为真实的扫描数据, (b) 为扩充前学习的结果, (c) 为扩充后学习的结果。

SCAPE 模型包含了一个关节的变形, 但由于对应点中大部分点都出现了错误的对应, 导致出现类似于膝盖部位这种“平坦的”形变。

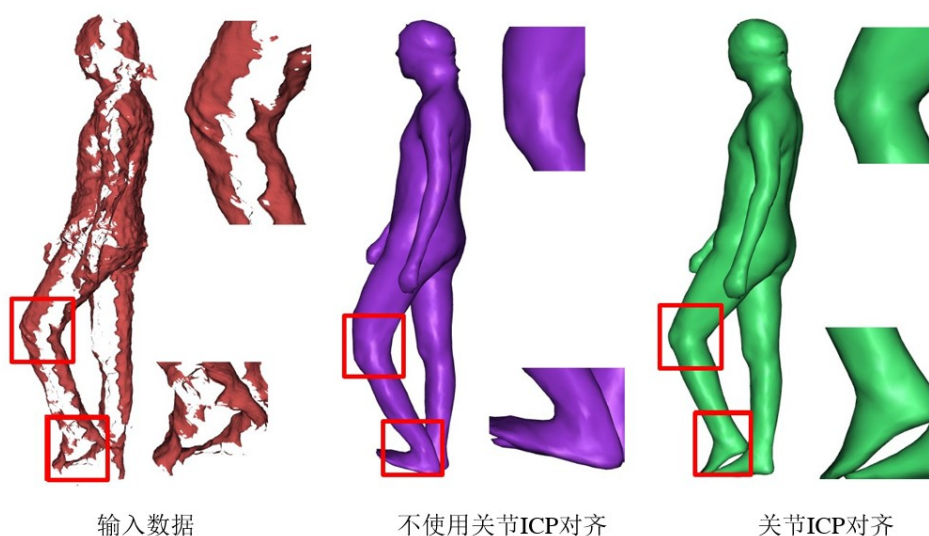


图 4.6 关节 ICP 算法对比结果。左为输入数据, 中间是不使用关节 ICP 而直接使用最近邻对应的结果, 最右为使用关节 ICP 作预对齐后再使用最近邻对应的结果。

遮挡情况的处理 图4.7为我们处理遮挡的实验。在实验的过程中, 经常会出现一个关节对应的部分全部都被遮挡导致不可见, 我们的算法是可以处理这种情况的。因为使用了层次式的解决方案, 只要这个被遮挡关节对应的父关节是可见的, 那么当前的关节就会一直被父关节的变换所制约, 当该关节重新出现的时候, 显然我们就可以利用最近邻的方法

渐渐地将该关节对齐并恢复到准确的姿态。同时在这个图中，我们可以看到在膝盖部位有些数据是没有的，我们的方法可以很好地处理这种情况，因为膝盖相连的两个关节大部分数据是存在的，这就足够用算法去准确地估计整体的姿态了。

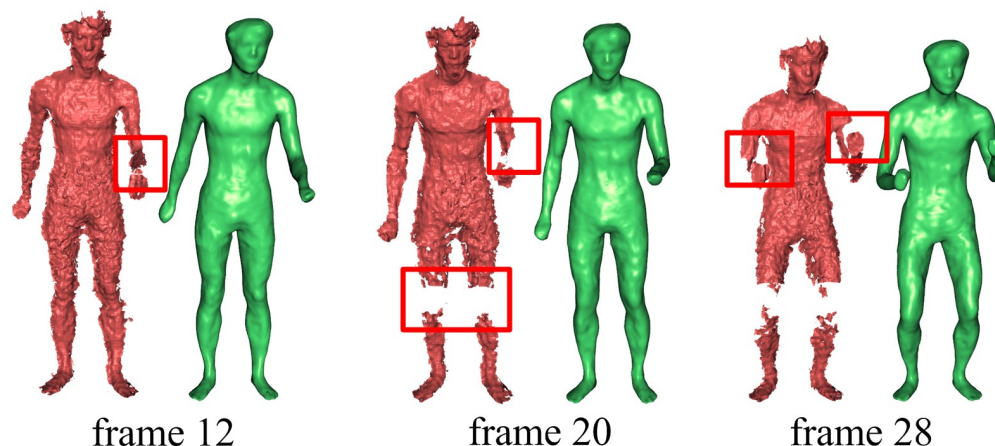


图 4.7 遮挡处理结果，如红色矩形所示，手肘附近由于遮挡导致逐渐消失，而后又重新出现。我们的方法对这种情况是鲁棒的。

仿真结果 为衡量整个重建结果的精度及准确性，我们使用 Vlasic 等人^[11] 开放的一组运动数据作为真实值，为模拟我们使用 Kinect 获取数据的环境，我们对将原始数据进行上采样同时将那些视角方向垂直的面片全部删除，并添加一些随机高斯噪声。使用第一帧的数据生成模板后，我们利用这个模板和处理后的输入数据进行跟踪并重建，结果如图 4.8 所示，在包含遮挡和破洞的地方，平均误差距离不到 5mm，而整个重建结果与真实数据对比的平均误差不超过 2mm。

4.5.2 更多结果

下面我们展示更多的一些实验结果。我们对不同的运动姿态进行重建，以表明我们的系统对姿态是比较鲁棒的。此外考虑到我们使用的隐式模板，我们采集了三位表演者的运动数据，分别进行重建。从后面的结果中，我们可以看到重建的模板与真实的体型是十分一致的。我们使用本章所述的硬件系统进行数据的获取后，进行跟踪重建。从图中，可以看到尽管 Kinect 数据质量不是很好，我们的细节重建结果仍能把其中的一些诸如衣服褶皱的细节表现出来。图 4.9 的姿态与初始姿态相关不是很大，但图 4.10 则与初始姿态相差很大，可以看出，在细节的地方仍是比较满意的。图 4.11 所示的姿态在 SCAPE 空间中实际上是不存在的，但可以看到我们还是能够比较自然的将这个姿态重建出来。图 4.12 是一组遮挡比较严重的的数据，但通过关节 ICP 的跟踪，系统还是比较好地将这个姿态重建了出来。

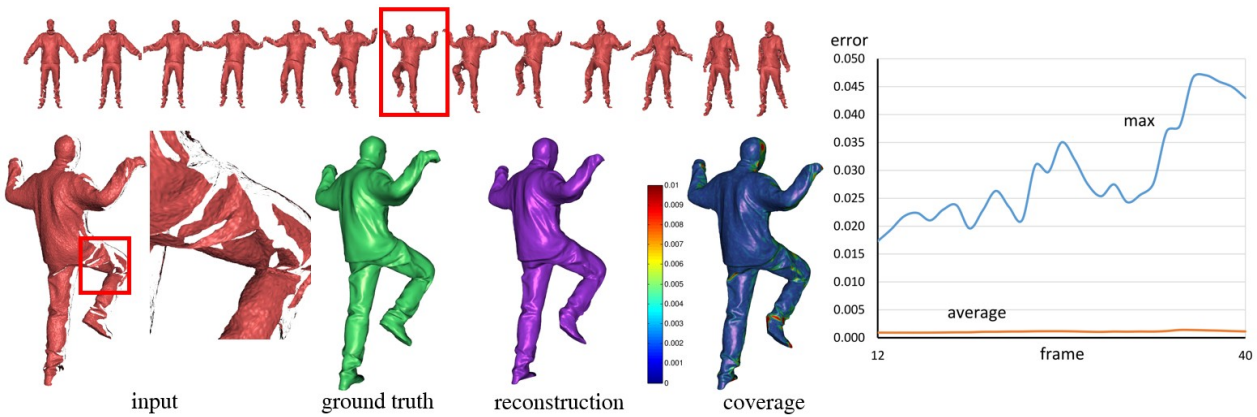


图 4.8 仿真实验结果。最顶上一行为输入的仿真运动序列数据，通过局部放大，我们可以看到处理后的结果已经包含遮挡和噪声，足以代表真实的数据，从第一帧开始进行跟踪并重建的结果如紫色的模型所示，从绿色的真实模型相对比结果可视化可知，重建结果与真实误差相差非常小。右面的图显示最大误差与平均误差随帧数的变化结果，随着姿态的复杂和误差的不可避免的积累，最大误差有增加的趋势，但平均误差始终保持在一个稳定的水平。

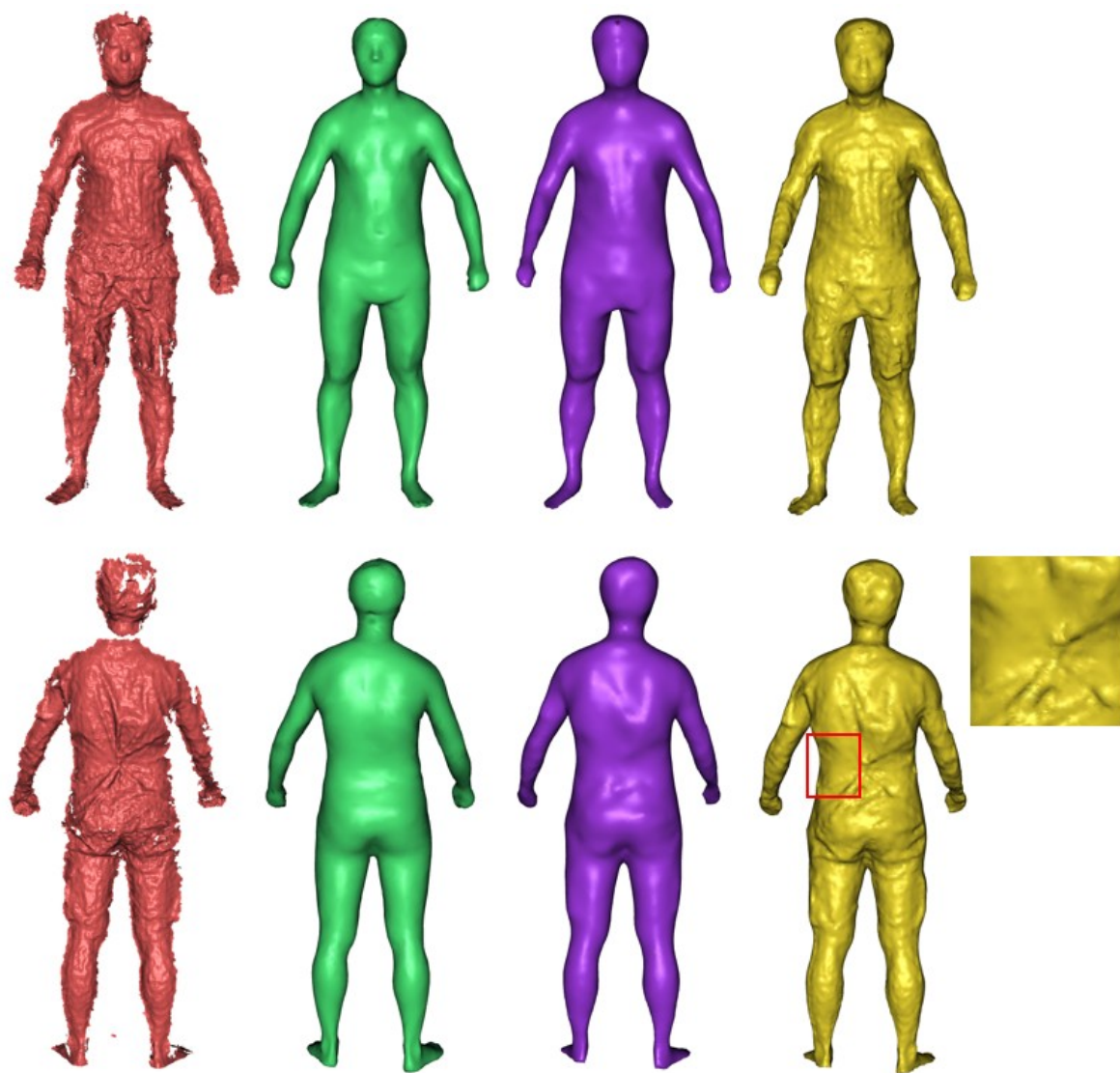


图 4.9 重建结果 1。分前后两个视图，各模型依次为输入数据、估计的 SCAPE 模型、姿态变化后的模板模型、最终的重建结果。

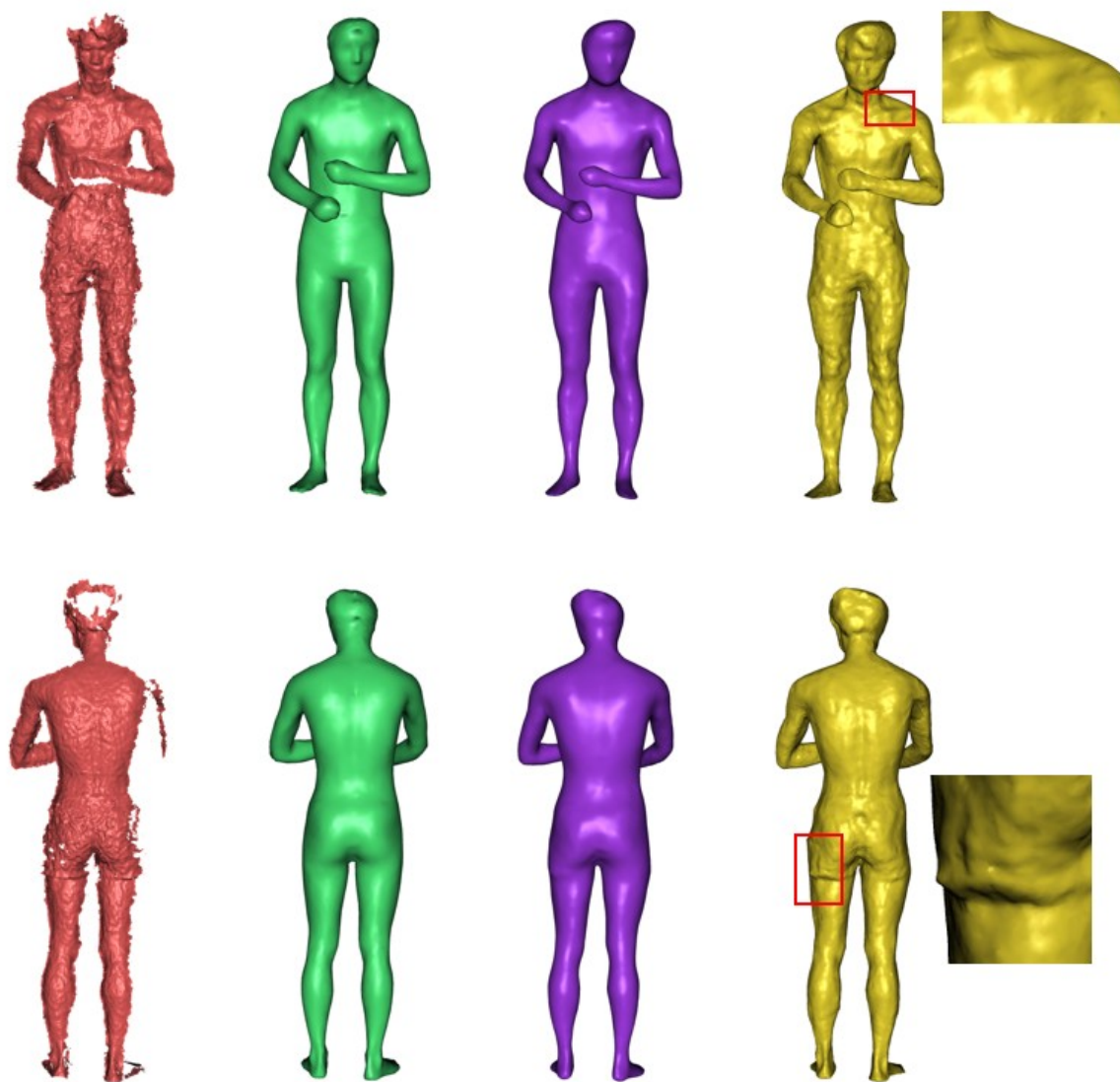


图 4.10 重建结果 2。各模型意义同图4.9。

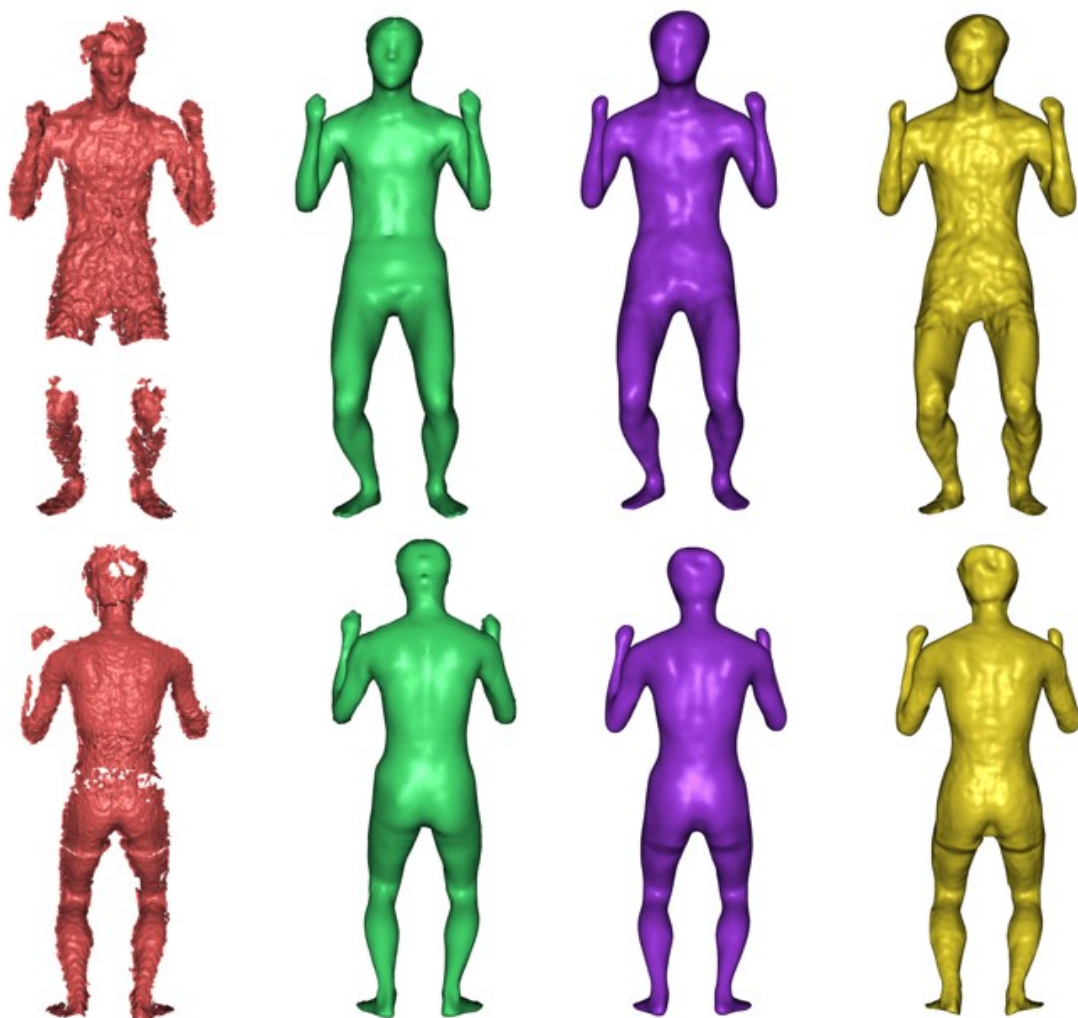


图 4.11 重建结果 3。各模型意义同图4.9。



图 4.12 重建结果 4。各模型意义同图4.9。

第 5 章 总结与展望

5.1 总结

运动物体的几何重建已经成为计算机视觉中的一个重要问题，它广泛应用于三维重建、人机交互等领域。

本文着重研究了使用以 **Kinect** 为代表的低廉设备对运动的人体进行重建。在前人提出的重建框架下，我们尝试在不需要显式地重建模板的情况下对人体运动进行建模，另一方面模板能够给我们提供良好的先验知识，因此我们采用统计模型学出一个隐式生成的模板。同时，我们扩充统计模型的训练数据库，以增强其学习的能力。在动态重建的过程中，我们使用了一个全局优化的方法对多台朝向对立的相机进行标定，同时，在进行运动跟踪的过程中，我们采用分段层次式对齐方法作为一个预处理的过程，极大地增强了算法的鲁棒性。最后，我们给出了我们的重建结果。

5.2 展望

考虑到我们这个方法本身具有很多的局限性，今后将在以下方面进行深入研究：

首先，考虑到我们系统使用的 **Kinect**，为了在保证一定的活动空间同时还保证相机的输出精度，我们的系统只能允许用户作非常小尺度的动作。因此，我们可以考虑使用更大范围的系统，如 **Light Stage** 光场等，使得允许用户有任意自由的动作。

另外，考虑到我们目前使用的 **SCAPE** 模型只是针对裸体的模型，事实上针对穿宽松衣服模型并不是非常地鲁棒，因此我们需要考虑重新审视我们的训练模型。比如重新训练一个针对穿衣服的模型，抽取其衣服的距离场与动作姿态变形分开进行训练，即训练一个穿衣服的 **SCAPE** 模型。

最后，在我们使用分段刚体预对齐之后，使用的是 **SCAPE** 的姿态生成方法，但这个跟踪生成方法计算比较耗时，尽管只要对应点比较准确的情况下生成的模型是比较地鲁棒，但我们还是需要考虑使用其它的方法对比，如基于蒙皮的方法进行深入地挖掘。

参考文献

- [1] Thabo Beeler, Bernd Bickel, Paul Beardsley, Bob Sumner, Markus Gross. High-quality single-shot capture of facial geometry[J]. ACM Transactions on Graphics (TOG), 2010. 29(4):40.
- [2] Hao Li, Bart Adams, Leonidas J. Guibas, Mark Pauly. Robust single-view geometry and motion reconstruction[J]. ACM Trans. Graph., December 2009. 28(5):175:1–175:10. ISSN 0730-0301. doi: 10.1145/1618452.1618521.
- [3] Hao Li, Robert W Sumner, Mark Pauly. Global correspondence optimization for non-rigid registration of depth scans[C]//Computer graphics forum. volume 27. Wiley Online Library, 2008:1421–1430.
- [4] Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, James Davis. Scape: shape completion and animation of people[C]//ACM Transactions on Graphics (TOG). volume 24. ACM, 2005:408–416.
- [5] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, Andrew Fitzgibbon. Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera[C]//Proceedings of the 24th annual ACM symposium on User interface software and technology. UIST '11. New York, NY, USA:ACM, 2011:559–568.
- [6] Svetlana Barsky, Maria Petrou. The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2003. 25(10):1239–1252.
- [7] Diego Nehab, Szymon Rusinkiewicz, James Davis, Ravi Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry[C]//ACM Transactions on Graphics (TOG). volume 24. ACM, 2005:536–543.
- [8] Ming Zeng, Fukai Zhao, Jiayang Zheng, Xinguo Liu. Octree-based fusion for realtime 3d reconstruction[J]. Graph. Models, May 2013. 75(3):126–136. ISSN 1524-0703.
- [9] Thibaut Weise, Sofien Bouaziz, Hao Li, Mark Pauly. Realtime performance-based facial animation[J]. ACM Trans. Graph., July 2011. 30(4):77:1–77:10. ISSN 0730-0301.
- [10] Hao Li, Jihun Yu, Yuting Ye, Chris Bregler. Realtime facial animation with on-the-fly correctives[J]. ACM Trans. Graph., July 2013. 32(4):42:1–42:10. ISSN 0730-0301.

- [11] Daniel Vlastic, Ilya Baran, Wojciech Matusik, Jovan Popović. Articulated mesh animation from multi-view silhouettes[J]. *ACM Trans. Graph.*, August 2008. 27(3):97:1–97:9. ISSN 0730-0301.
- [12] Genzhi Ye, Yebin Liu, Nils Hasler, Xiangyang Ji, Qionghai Dai, Christian Theobalt. Performance capture of interacting characters with handheld kinects[C]//*Proceedings of the 12th European conference on Computer Vision - Volume Part II. ECCV'12. Berlin, Heidelberg:Springer-Verlag. ISBN 978-3-642-33708-6, 2012:828–841.*
- [13] Juergen Gall, Carsten Stoll, Edilson De Aguiar, Christian Theobalt, Bodo Rosenhahn, H-P Seidel. Motion capture using joint skeleton tracking and surface estimation[C]//*Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009:1746–1753.*
- [14] Carsten Stoll, Juergen Gall, Edilson de Aguiar, Sebastian Thrun, Christian Theobalt. Video-based reconstruction of animatable human characters[J]. *ACM Trans. Graph.*, December 2010. 29(6):139:1–139:10. ISSN 0730-0301.
- [15] Edilson de Aguiar, Carsten Stoll, Christian Theobalt, Naveed Ahmed, Hans-Peter Seidel, Sebastian Thrun. Performance capture from sparse multi-view video[J]. *ACM Trans. Graph.*, August 2008. 27(3):98:1–98:10. ISSN 0730-0301.
- [16] Daniel Vlastic, Pieter Peers, Ilya Baran, Paul Debevec, Jovan Popović, Szymon Rusinkiewicz, Wojciech Matusik. Dynamic shape capture using multi-view photometric stereo[J]. *ACM Trans. Graph.*, 2009. 28(5):174:1–174:11.
- [17] Brian Curless, Marc Levoy. A volumetric method for building complex models from range images[C]//*Proceedings of the 23rd annual conference on Computer graphics and interactive techniques. ACM, 1996:303–312.*
- [18] Hao Li, Linjie Luo, Daniel Vlastic, Pieter Peers, Jovan Popović, Mark Pauly, Szymon Rusinkiewicz. Temporally coherent completion of dynamic shapes[J]. *ACM Trans. Graph.*, February 2012. 31(1):2:1–2:11. ISSN 0730-0301.
- [19] Michael Kazhdan, Matthew Bolitho, Hugues Hoppe. Poisson surface reconstruction[C]//*Proceedings of the Fourth Eurographics Symposium on Geometry Processing. SGP '06. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association. ISBN 3-905673-36-3, 2006:61–70.*
- [20] Will Chang, Matthias Zwicker. Global registration of dynamic range scans for articulated model reconstruction[J]. *ACM Trans. Graph.*, May 2011. 30(3):26:1–26:15. ISSN 0730-0301.
- [21] Ming Zeng, Jiayang Zheng, Xuan Cheng, Xinguo Liu. Templateless quasi-rigid shape modeling with implicit loop-closure[C]//*Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. CVPR '13. Washington, DC, USA:IEEE Computer Society, 2013:145–152.*
- [22] Paul J Besl, Neil D McKay. Method for registration of 3-d shapes[C]//*Robotics-DL tentative. International Society for Optics and Photonics, 1992:586–606.*
- [23] Hao Li. Animation reconstruction of deformable surface[D]. :ETH Zurich, 2010.

- [24] K Somani Arun, Thomas S Huang, Steven D Blostein. Least-squares fitting of two 3-d point sets[J]. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 1987. (5):698–700.
- [25] Chen Yang, Gérard Medioni. Object modelling by registration of multiple range images[J]. *Image and vision computing*, 1992. 10(3):145–155.
- [26] Szymon Rusinkiewicz, Marc Levoy. Efficient variants of the icp algorithm[C]//3-D Digital Imaging and Modeling, 2001. *Proceedings. Third International Conference on*. IEEE, 2001:145–152.
- [27] Kok-Lim Low. Linear least-squares optimization for point-to-plane icp surface registration[J]. Chapel Hill, University of North Carolina, 2004.
- [28] Gregory C Sharp, Sang W Lee, David K Wehe. Multiview registration of 3D scenes by minimizing error between coordinate frames[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004. 26(8):1037–1050. ISSN 01628828. doi:10.1109/TPAMI.2004.49.
- [29] Benedict J Brown, Szymon Rusinkiewicz. Global non-rigid alignment of 3-d scans[J]. *ACM Transactions on Graphics (TOG)*, 2007. 26(3):21.
- [30] Qi-Xing Huang, Bart Adams, Martin Wicke, Leonidas J Guibas. Non-rigid registration under isometric deformations[C]//Computer Graphics Forum. volume 27. Wiley Online Library, 2008:1449–1457.
- [31] Brian Amberg, Sami Romdhani, Thomas Vetter. Optimal step nonrigid icp algorithms for surface registration[C]//Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on. IEEE, 2007:1–8.
- [32] Robert W. Sumner, Johannes Schmid, Mark Pauly. Embedded deformation for shape manipulation[C]//ACM SIGGRAPH 2007 papers. SIGGRAPH '07. New York, NY, USA:ACM, 2007.
- [33] Alexander Weiss, David Hirshberg, Michael J Black. Home 3d body scans from noisy image and range data[C]//Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011:1951–1958.
- [34] Nils Hasler, Carsten Stoll, Bodo Rosenhahn, Thorsten Thormählen, Hans-Peter Seidel. Technical section: Estimating body shape of dressed humans[J]. *Comput. Graph.*, 2009. 33(3):211–216.
- [35] Alexandru O Balan, Leonid Sigal, Michael J Black, James E Davis, Horst W Haussecker. Detailed human shape and pose from images[C]//Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on. IEEE, 2007:1–8.
- [36] Alexandru O. Bălan, Michael J. Black. The naked truth: Estimating body shape under clothing[C]//Proceedings of the 10th European Conference on Computer Vision: Part II. ECCV '08. Berlin, Heidelberg:Springer-Verlag, 2008:15–29.
- [37] Peng Guan, Alexander Weiss, Alexandru O Balan, Michael J Black. Estimating human shape and pose from a single image[C]//Computer Vision, 2009 IEEE 12th International Conference on. IEEE, 2009:1381–1388.
- [38] Shizhe Zhou, Hongbo Fu, Ligang Liu, Daniel Cohen-Or, Xiaoguang Han. Parametric reshaping of human bodies in images[C]//ACM SIGGRAPH 2010 papers. SIGGRAPH '10. New York, NY, USA:ACM, 2010:126:1–126:10.

- [39] Ilya Baran, Jovan Popović. Automatic rigging and animation of 3d characters[C]//ACM Transactions on Graphics (TOG). volume 26. ACM, 2007:72.
- [40] Robert W Sumner, Jovan Popović. Deformation transfer for triangle meshes[J]. ACM Transactions on Graphics (TOG), 2004. 23(3):399–405.
- [41] Yi Ma. An invitation to 3-d vision: from images to geometric models[M], volume 26. springer, 2004.
- [42] Nils Hasler, Carsten Stoll, Martin Sunkel, Bodo Rosenhahn, H-P Seidel. A statistical model of human pose and body shape[C]//Computer Graphics Forum. volume 28. Wiley Online Library, 2009:337–346.
- [43] Zhengyou Zhang. A flexible new technique for camera calibration[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2000. 22(11):1330–1334.
- [44] Tomas Svoboda. A software for complete calibration of multicamera systems[C]//Electronic Imaging 2005. International Society for Optics and Photonics, 2005:115–128.
- [45] Christoph Strecha, Wolfgang von Hansen, Luc Van Gool, Pascal Fua, Ulrich Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery[C]//Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008:1–8.
- [46] Dragomir Anguelov. Learning Models of Shape from Three-dimensional Range Data[D]. Stanford, CA, USA:, 2006. AAI3197402.
- [47] Jonathan Starck, Adrian Hilton. Surface capture for performance-based animation[J]. IEEE Comput. Graph. Appl., May 2007. 27(3):21–31. ISSN 0272-1716.
- [48] Lars Mundermann, Stefano Corazza, Thomas P Andriacchi. Accurately measuring human movement using articulated icp with soft-joint constraints and a repository of articulated models[C]//Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on. IEEE, 2007:1–6.
- [49] Stefano Corazza, Lars Mundermann, Emiliano Gambaretto, Giancarlo Ferrigno, Thomas P. Andriacchi. Markerless motion capture through visual hull, articulated icp and subject specific model generation[J]. Int. J. Comput. Vision, March 2010. 87(1-2):156–169. ISSN 0920-5691.
- [50] Varun Ganapathi, Christian Plagemann, Daphne Koller, Sebastian Thrun. Real-time human pose tracking from range data[C]//Proceedings of the 12th European conference on Computer Vision - Volume Part VI. ECCV'12. Berlin, Heidelberg:Springer-Verlag. ISBN 978-3-642-33782-6, 2012:738–751.

发表文章目录

1. **Jiaxiang Zheng**, Ming Zeng, Xuan Cheng, Xinguo Liu. SCAPE-Based Human Performance Reconstruction. Computer & Graphics 2013.
2. Ming Zeng, **Jiaxiang Zheng**, Xuan Cheng, Xinguo Liu. Dynamic Human Surface Reconstruction Using a Single Kinect. CAD/Graphics 2013.
3. Ming Zeng, **Jiaxiang Zheng**, Xuan Cheng, Xinguo Liu. Templateless Quasi-Rigid Shape Modeling with Implicit Loop-Closure[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR) 2013.
4. Ming Zeng, Fukai Zhao, **Jiaxiang Zheng**, Xinguo Liu. A Memory-Efficient KinectFusion Using Octree. Computational Visual Media(CVM) 2012.

致 谢

值此学位论文完稿之际，我谨向所有在生活和学习中曾经给予过我关心和帮助的老师、同学及亲朋致以衷心的感谢！

首先，我要感谢导师刘新国教授。两年前的我初到 CAD&CG 实验室时，对于图形学和视觉一无所知，是刘老师的指导让我开始了这个领域的学习。“授人以鱼不如授人以渔”，刘老师正是这样做的，二年多的时光里，我慢慢懂得如何才是一个正确地研究态度和学习态度。此外，不仅仅在学习和研究中的指导，刘老师在生活和做人上也给我诸多启示。记得刚来实验室的时候，我曾鲁莽地给刘老师发过没有礼貌的邮件，刘老师并没有批评我，而是耐心地教导我以后应该如何给长辈邮件，这令我十分感动。

其次，我还要感谢我的师兄曾鸣博士。如果说是刘老师引导我走向图形与视觉领域的话，曾师兄则是牵着我走过这其中的各种困难。初来之时，我编程能力不是太好，师兄从算法设计到代码调试上给予我全方位的指导，令我难以忘怀。同时他对于研究和学习的态度也令让我值得学习。

感谢彭组讨论班的各位老师同学，彭老师在每次讨论中深入浅出的点评与总结总能给予我醍醐灌顶的启示；王章野和于金辉等老师的讨论也不断启发和指导着我的学习热情。此外，同组的师兄师姐和师弟们在平时的学习和生活中也给我诸多帮助，在此一并致谢，他们是：姜玻、赵福恺、曾鸣、吴子朝、康菁菁、王开盛、梁伯均、程轩、寿如阳、陈坤、蒋剑、谭歆、郭骏、郑辉。同时谢谢实验室 408 和 415 的诸多同学两年多以来的帮助，谢谢陪伴两年多的同寝室友韩祥、朱萌和吴东东，他们在我的研究生生活中带来了无数的帮助与欢乐。

最后，我要感谢我的家人，是他们长期以来的默默关怀令我对未来更加有信心，每周的电话里，哪怕只是一句轻声的问候，也令我感觉到温暖。感谢我的女朋友刘秒，正是她在我背后默默的支持，才让我在生活和学习中更加自信地面对一切。

郑家祥

于浙江大学 CAD&CG 国家重点实验室

2014 年 3 月 6 日