

硕 士 研 究 生 读 书 报 告



题目 Capturing Detailed Deformations of Moving Human Bodies

作者姓名 陈克纯

作者学号 22251023

指导教师 李启雷

学科专业 电子信息（软件工程）

所在学院 软件学院

提交日期 二○二二 年 12月

# 摘要

这篇论文提出了一种新的方法，通过在人体采样1000多个独特的点来捕捉详细的人体运动。该方法输出高度精确的4D（时空）点坐标，自动为每个点分配唯一的标签，并可以仅从单个2D输入图像推断点的位置和唯一标签。因此，这种方法可以通过捕获点的轨迹获得输入图像的所有细节，包括因呼吸、肌肉收缩和肌肉变形引起的运动。这种方法的关键是一种新型的运动捕捉套装，它包含一种特殊的图案，带着棋盘状的角和两个字母代码。这些图案会被多摄像头系统进行一系列经过训练可识别角落并识别代码的神经网络处理。这种处理具有鲁棒性，可以适应身体的拉伸。整个系统仅依赖于标准RGB或单色传感器、全被动照明和被动套装，这使得该方法易于复制、部署和使用

**关键词**： 人体动作、运动捕捉、人体皮肤形变

# 1引言

在真实世界图像中，人体被衣服遮挡，使得精确的身体测量变得困难。然而一个小的肌肉抽搐或呼吸的速度可能包含某些至关重要的信号。针对这一点，先前的工作要么利用1）附着在运动捕捉服上或粘在皮肤上的反射标记；2）在皮肤上绘制彩色图案。但是在传统的方法中，标记强烈依赖于时间跟踪和高帧率相机，因此稳健的标记是一个难题。随着标记的数量的增加，捕捉标记的难度也会越大。而稀疏标记虽然足以拟合低维骨骼体模型，但不能捕捉由于呼吸引起的肌肉变形和人体运动细节。

为了捕捉高细节的运动，这篇论文使用了一种运动捕捉套装。套装上有一些棋盘状角，可以精确定位身体表面的点；还包含与每个角落相邻的唯一两个字母代码，使得我们能通过识别代码直接标记角落。而DFAUST方法依赖于初始几何配准和时间跟踪，这可能会导致误差累积的影响，并可能在更具挑战性的姿态或快速运动中导致不正确的局部最小值。相比之下，本论文的方法展示了更广泛的运动捕捉，包括体操练习、瑜伽体式或在地面上滚动。因为这种定位方法可以独立地处理每个图像，所以即使身体只有一小部分可见，该方法依然是成功的，不存在由于遮挡和非遮挡而使传统时间跟踪复杂化的问题。如图1所示。

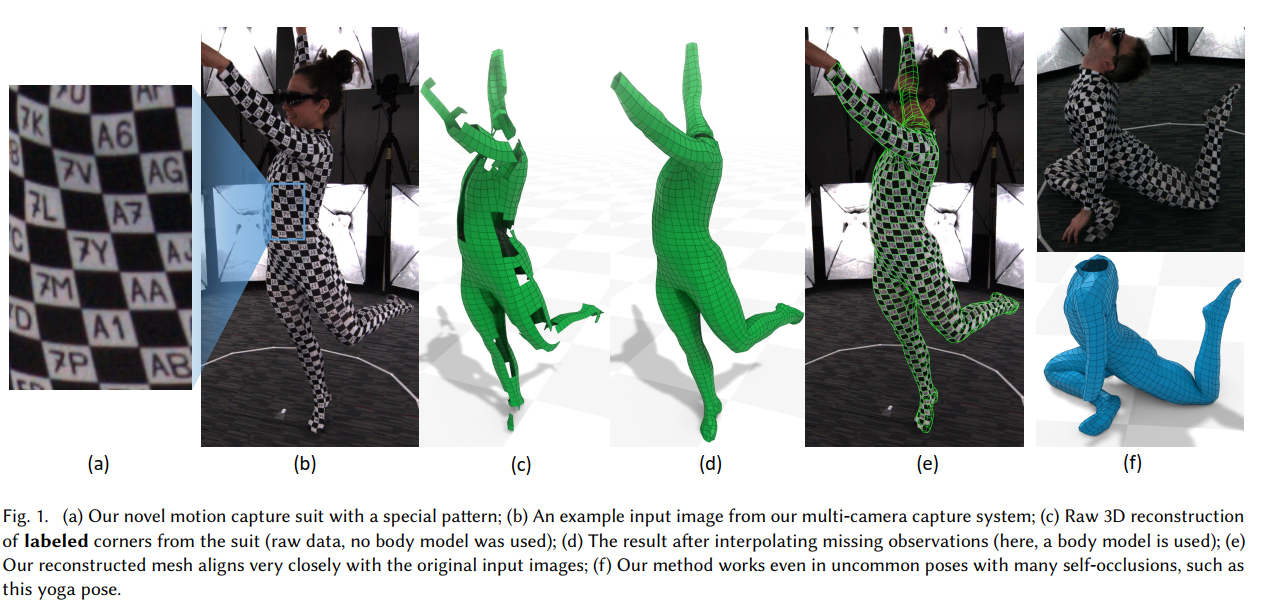


图 1 运动捕捉套装图示

# 2内容介绍

为了制作运动捕捉套装，首先购买一件紧身连体衣（自带黑白棋盘纹理）。然后用记号笔在白色的方格里写代码，如图1（a）。该方法的多相机设置包含16个标准（RGB）相机，它们排列成围绕捕获物体的圆圈，通过同步锁同步，如图2所示。

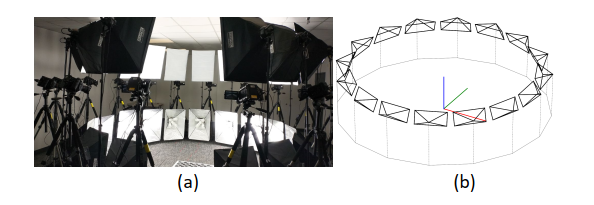


图 2 多相机设置图

校准的相机生成图像序列，这些图像将由管道处理。首先以亚像素精度检测输入图像中的棋盘状角。接下来，需要通过识别相邻的两个字母代码来唯一地标记检测到的角。这种方法在设计上是局部的，即，管道的每个阶段都与输入图像的小补丁一起工作。这样的方法有如下几个优点：a）能够提取人体的可靠几何信息，甚至可以从一小块衣服中提取关键的对应关系。这使得对人体的遮挡或局部视图非常鲁棒，例如由于放大的相机。b） 通过将套装分解为小四边形并使用单应性消除它们，可以抵消大部分投影变形和套装拉伸，从而简化学习任务。c） CNN四元分类器包括一个质量控制机制，拒绝质量可疑的白色方块，并进一步提高了我们方法的鲁棒性

角检测器的任务是检测并定位输入图像中的所有棋盘状角。这项任务非常重要，因为背景中有类似角落的特征，西装随着皮肤伸展，并且有明显的照明变化。通常在每个输入图像中检测到数百个角。下一步是读取代码并将它们链接到角落，这将为每个角落提供一个独特的标签。此时生成的四边形是“候选四边形”，因为它们可能（但不保证）包含正确的两个字母代码。使用单应变换将每个候选四边形的四个角变换为标准化的正方形，以简化后续处理。

该论文训练了两个四元分类器RejectorNet和RecogNet。RejectorNet是一个二元分类器，用于预测候选四边形是否有效，即四个角是否位于正确的位置，以及它们相对于直立代码方向的顺序是否正确。有效的四边形被传递给RecogNet读取代码。 RecogNet是一个具有两个头部的多类分类器，每个头部对应两个字母代码的每个字符。

此时，有效四边形的两个字母代码已被识别，包括其直立方向。下一步是对每个角落进行唯一标记。上述标签一致性检查仅在西装中存在两个相邻的双字母代码且在图像中可见时有效。该方法的一个关键特征是，所有网络都只在小图像块上训练。这使得我们的训练模型可以推广到不同的服装、捕捉环境、相机配置和不在训练集合中的身体姿势，因为局部基准标记显示出的可变性比完整人类姿势的图像小得多。这与基于深度学习的方法非常不同，这些方法执行全局姿势预测，将身体作为一个整体来寻找。网络的培训不需要大型培训集。作者自己准备了训练数据，不使用任何外部注释服务或现有数据集。数据集包含28张手动注释的图像，其中24张是从三位演员穿着手绘西装的照片中随机选择的，还拍摄了一位穿着印花西装的不同（第四位）演员，并对其四幅图像进行了注释，以评估对印刷字符的处理方法，而不是手绘。印花西装的数据仅用于测试CNN的性能。在角点标注中，以亚像素精度手动标注西装上所有棋盘式角点。在四边形标注中，手动将上一步骤中标注的角连接到四边形中。具体来说，该方法创建了对应于套装中带有两个字母代码的有效白色正方形的四边形，注释器还记下了每个带注释的四边形的代码。该方法通过几何和颜色空间运算对训练数据进行数据增强，还通过渲染人体纹理模型生成综合训练数据。最后通过一种插值（修复）缺失角的方法，不依赖以前的统计体型模型，而是使用该方法捕捉给演员的示例动作，使用这些数据创建更精确的精细化身体模型。

总结，这篇论文提出了一种在移动人体表面捕获1000多个唯一标记点的方法。这项技术是通过新型运动捕捉套装实现的，该套装带有棋盘式的角和两个字母代码，能够对每个角进行独特的标记。结果是使用现成组件构建的多摄像头系统获得的，其成本仅为全身3DMD设置的一小部分，同时展示了比DFAUST数据集更广泛的运动种类，包括体操、瑜伽姿势和在地上滚动。用于重建标记的3D点的方法不依赖于时间相关性，这使得它对去遮挡非常鲁棒，并且还需要并行处理。

# 3 创新点与局限性分析

3.1 创新点

本文提出的方法具有以下几个创新点。

（1）能够精确地识别人体的各种姿势并进行捕捉，包括各种具有挑战性的动作，如翻滚、体操和瑜伽。

（2）能够独立且并行地处理每一个2D图像，仅使用局部图像块实现标记定位和唯一标记，不使用时间跟踪，因此对遮挡具有鲁棒性。

（3）能自动推断相应的标记标签，不受任何类型的建模或归纳偏见的影响。

（4）易于复制、部署和使用。

3.2 局限性

（1）特殊的动作捕捉服是必须的。

（2）当前的相机分辨率（4000×2160）无法从皮肤获得足够的细节。

（3）忽略了脸部和手部的运动捕捉。

（4）无法进行实时的数据处理。