

硕 士 研 究 生 读 书 报 告



题目 单目RGB视频流中做人体重建

作者姓名 赵中源

作者学号 22051049

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 2020 年 12 月

**摘要**

论文首次提出了根据单眼视频实时渲染体积性能捕获的方法，消除了对于昂贵的多相机系统和繁琐的个性化模型预设置的需求。该系统借助了像素一致隐式函数（PIFu）重建了完全纹理化的3D人体，尽管PIFu实现了高分辨率的重建，但它的计算消耗不足以支持该方法实时化，因此论文提出了一种新颖的分层表面定位算法和一种无需显式提取表面网格的直接渲染方法，通过从粗到精的方式剔除出不必要的区域，最终该模型成功地将重建速度从基准提高了两个数量级而又不影响质量。此外，论文还引入了一种在线难点挖掘（OHEM）技术，根据当前的重建精度自适应地更新训练数据的采样率，从而有效减轻了重建失真。

**关键词**：3D重建 实时 深度学习

**1引言**

尽管最近在虚拟和增强现实以及3D显示方面取得了进步，但允许进行更沉浸和引人入胜的交互，但是使用单台摄像机进行视频会议仍然是长距离面对面交流的最常见方法。原因很简单：方便。尽管存在可以获取可从任意角度渲染的特定外观的高保真数字表示的技术，但是现有的捕获和流传输此数据的方法需要繁琐的捕获技术，例如大量的经过校准的相机或深度传感器，以及安装和部署这些系统的专业知识。另一方面，视频会议仅需要单个摄像机，例如在普通消费类设备上找到的摄像机。笔记本电脑和智能手机。因此，如果我们可以从单个消费级相机上捕获一个人独特的外观和动作的完整模型，那么我们可以弥合差距，防止新手用户在虚拟环境中进行沉浸式通信。

然而由于深度模糊性、拓扑结构的变化以及严重的遮挡，从单一角度同时重建人体的几何结构和纹理是一个很大的挑战。为了应对这些挑战，学者提出了众多基于数据驱动的高容量深度神经网络方法，在人体建模方面取得了众多突破。特别的，像素一致隐式函数（PIFu）实现了穿着衣服的人体的完全纹理高分辨率重建，但是PIFu的局限性在于无法达到实时渲染的要求。

为此改论文提出了一种新的表面重建算法，以及一种无需提取曲面网格即可进行渲染的直接渲染方法。算法以从粗到精的方式逐步查询3D位置，以构造要评估的点数较少的3D占用场。此外，结合所提出的表面重建算法，隐式纹理表示可直接进行新视图合成，而无需进行几何细分或纹理映射，从而将渲染所需的时间减少了一半。同时，论文还提出了一种在线难点挖掘（OHEM）方法减少了重建失真现象。最终论文实现了在15fps条件下运行2563空间分辨率的体积性能捕获。

**2方法**

该算法的目标是在给定RGB图片流的情况下，在实时状态下获得物体完整的几何结构和纹理，并且包括看不见的区域，比如人的背部。团队采取每帧独立处理的方法，因为基于跟踪的方案很可能会积累错误从而产生偏移和不稳定现象。

对于每帧图像，首先实时地将物体从背景分割出来，分割后的图像再送到增强版的PIFu中去预测物体的连续占用场，表面确定后同样适用PIFu推测出物体纹理，而且这可以从任何视角进行渲染。算法的核心是该团队开发的一种新的加速结构使得PIFu能够在实时状态下运行，除此之外，受OHEM的启发，团队还进一步地改进了系统的鲁棒性。

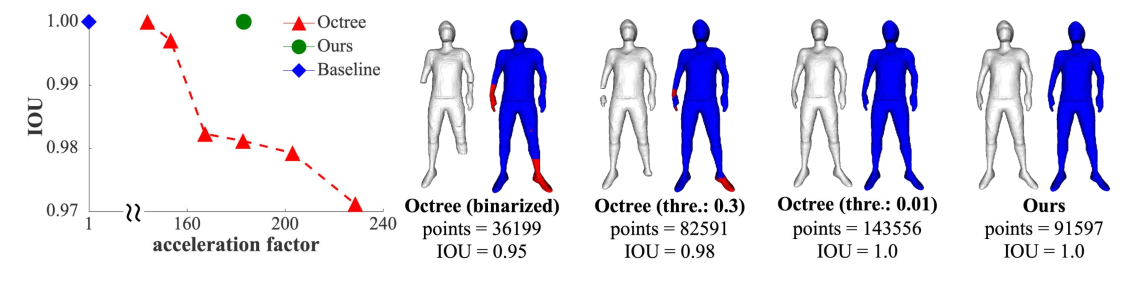
**2.1像素一致隐式函数（PIFu）**

在体积捕获中，3D几何表示为连续标量场的水平设置表面，比如有一个输入帧，我们需要确定一个在3D空间中的点是在人体内部还是外部，当目标空间已经显式离散化时，可以直接通过体素（voxels）计算得到，PIFu可以建模查询任意一3D点并预测该点在标准化设备坐标系下的占有场，值得注意的是，这种方法并不需要将3D物体离散化，使得在物体可以在任意分辨率下重建。

**2.2实时推测与渲染**

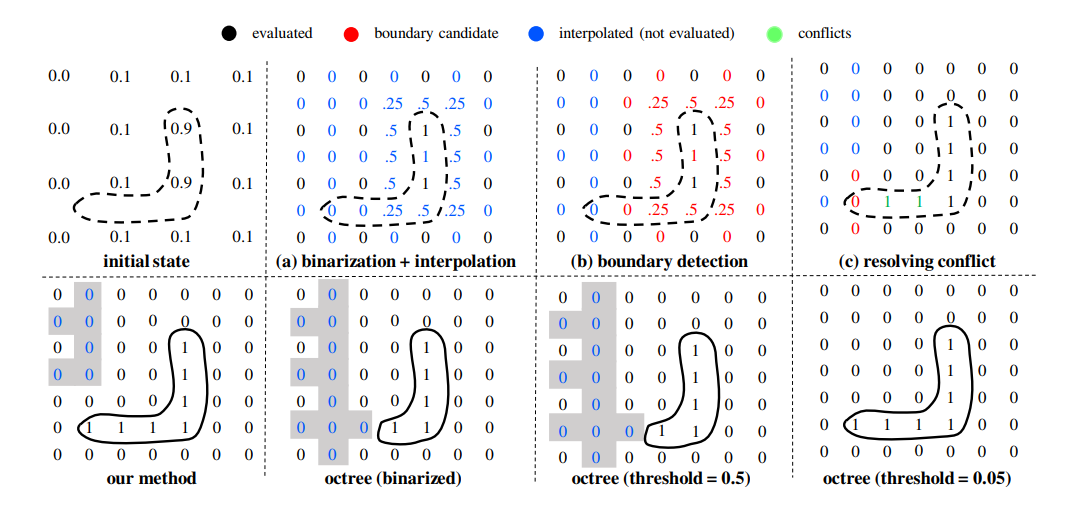
为了减少实时性能捕获所需的计算量，我们引入了两种新的加速技术。首先，我们提出了一种有效的表面定位算法，该算法保留了蛮力重构的精度，且复杂度与基于基础八叉树的重构算法相同。此外，由于我们的最终输出是从新的角度进行渲染，因此我们通过直接从PIFu生成新视图渲染来绕过显式网格重建阶段。通过结合这两种算法，我们成功实现了任意角度的实时渲染。

管线的主要瓶颈在于需要所需要的处理点数过多，因此减少所需处理的点的个数将大大提高重建效率，八叉树是一个常用的减少存储节点的方法，为了将八叉树应用在神经网络参数化的隐式表面上，最近的工作提出了一种算法：只细分那些在边界节点（在内部节点和外部节点之间的表面）的网格，但是团队发现这种方法和暴力求解相比经常产生不准确的重建结果。



**图1 表面重建方法比较。折线图表示了加速效果和准确度的权衡关系（trade-off）**

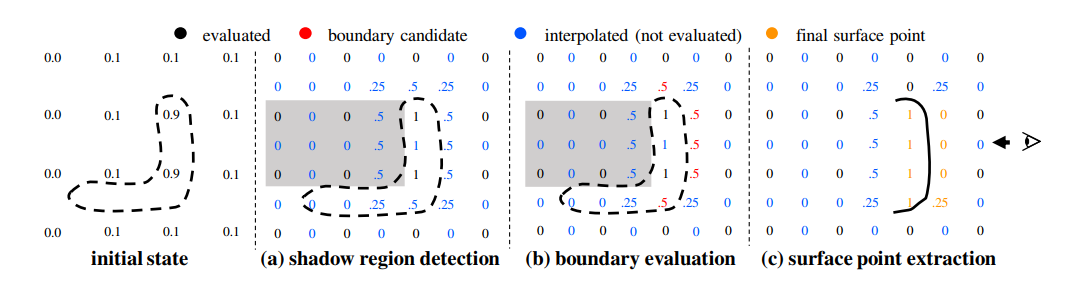
还有一种算法是当邻近粗网格的最大绝对偏差超过了某个阈值时细分网格，这种算法能够平衡好重建的速度和准确度，但却计算了更多不必要的点。



**图2 表面定位算法比较。虚线和实线分别代表了真实表面和重建表面的结果，灰色网格代表了没有包含进计算的点**

从上图可以看出来，由于只有表面附近的节点才对重建有影响，团队提出的算法就是仅定位和表面间隔在一个节点之内的网格，因此使用了一种由粗到精的网格选取策略，并在逐渐更新的计算过程中剔除不必要的点。

当定位算法成功加速了表面定位过程后，我们最终的目标就是从新的角度渲染物体，包括那些在视频中看不到的地方。这里再一次借用PIFu的能力，PIFu允许我们直接推算出3D空间中任意一点的纹理，这可以用来替代需要显式网格表示的传统渲染管线，如果表面位置是确定的，我们可以直接生成一个全新视角的图片。基础PIFu的表现，团队提出了一种基于视角的剔除算法和一个从隐式表示数据直接渲染的方法，目前已经有可微球追踪（differentiable sphere tracing）和光线步进（ray marching）算法来直接渲染隐式数据，但由于这些方法为了基于图片的监督任务的可微性而牺牲了计算速度，并不适合实时渲染。



**图3 免网格渲染（mesh-free rendering）。虚线为真实表面，实线为重建表面。**

**2.3 OHEM数据采样**

在形状学习中，重要性点采样要比在整个包围盒中采样更有效率，但是团队发现这种采样策略仍然会在一些具有挑战性的动作和视角重建时失败，当然这些只占训练数据的很小一部分，一种解决办法是增加更多的复杂场景，但这种做法不仅费时费力，而且更多参数时可能会引入更多的错误。

为了解决该问题，团队采取让系统面对复杂场景时自动更新采样率的方法，在数据偏差问题上，Online Hard Example Mining（OHEM）策略已经被应用在很多案例上，比如图片分类、物体识别和图片描述符学习等，每个都根据特定的任务使用一种挖掘策略，因此，将他们的算法扩展到另一个问题并非易事。相反，我们的表述是一般性的，因为它不需要特定领域的知识，因此可以应用于任何问题领域。

**3 结果与展望**

该方法在没有任何超参数的情况下实现了最佳的加速度，在保持原始重建精度的同时，将表面重建从30秒加速到0.14秒（7 fps）。通过将其与我们的无网格渲染技术相结合，我们只需2563的体积分辨率就可以以每帧0.06秒（15 fps）的速度进行新视图的渲染，从而首次实现了单眼视频实时体积性能捕获。

在实现了实时渲染的同时，作者也提到了该系统的缺点：

1. 泛化性不足。作者提到不论是什么奇怪的动作或者服饰，只要是训练数据中没涉猎到的，重建效果都不能保证，所以如果能做成半监督无监督系统就更好了。
2. 缺乏先验知识。隐式表达本身缺乏先验信息，所以会出现重建出奇怪东西的现象。
3. 背面纹理模糊。大部分衣服都是对称纹理的，所以模型也会倾向于对称渲染，但不是所有衣物都满足对称的条件。
4. 不可导。
5. 不支持多人

**参考文献**

[1] Ruilong Li, Yuliang Xiu, Shunsuke Saito. etc. Monocular Real-Time Volumetric Performance Capture. ECCV 2020.

[2] Shunsuke Saito, Zeng Huang, Ryota Natsume. etc. PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization. ICCV 2019.