

硕 士 研 究 生 读 书 报 告



题目 使用核回归的高效光度立体法

作者姓名 赵中源

作者学号 22051049

指导教师 金小刚

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**

光度立体测量法从使用固定摄像机在不同光线方向下拍摄的多个图像中估计表面法线。为了处理非朗伯体反射，最近的光度立体测量法采用了迭代或优化框架，计算成本很高。本文提出了一种使用核回归的高效光度立体测量方法，它可以被转化为一个特征分解问题。核参数对每个表面点来说都是可变的，因此它可以应对各种一般的反射率。最佳核参数是通过留一交叉验证自动确定的。为了提高计算效率，通过快速的矩阵计算和适当的法线初始化来加速留一的过程。该文提出的光度立体法在具有不同反射率的合成和真实表面上进行了广泛的评估。实验结果表明该方法在计算上是高效的，并且在表面法线估计方面达到了先进的精度。

**关键词**：光度立体法 核回归 法线生成 特征分解

Abstract

Photometric stereo estimates surface normals from multiple images captured under different light directions using a fixed camera. To deal with non-Lambertian reflections, the recent photometric stereo methods employ iterative or optimization frameworks that are computationally expensive. This paper proposes an efficient photometric stereo method using kernel regression, which can be transformed to an eigendecomposition problem. The kernel parameter is variable for each surface point so that it can cope with the variety of general reflectances. The best kernel parameter is automatically determined by leave-oneout cross validation. To improve computational efficiency, the leave-one-out process is accelerated by fast matrix computation and proper normal initialization. The proposed photometric stereo method is extensively evaluated on synthetic and real surfaces with various reflectances. Experimental results validate that the method is computationally efficient and achieves the state-of-the-art accuracy in surface normal estimation.

**Keywords：**Photometric stereo kernel regression surface normal eigen decomposition

**1引言**

自Woodham[1]的开创性工作以来，光度立体学已被广泛地应用于表面重建。在光度立体技术中，物体表面被多个方向的光源照射，其图像由固定视角的相机拍摄[2]。在理想的朗伯反射的假设下，表面法线只需通过三幅图像就可以计算出来。对于非朗伯反射的物体表面，总是需要更多数量的图像。通过将高光和阴影像素视为离群值，在朗伯反射的假设下，可以从剩余的像素计算出表面法线。

然而，由于粗糙镜面高光甚至非朗伯斯漫反射，对于许多真实的物体表面而言，高光检测并不是一件容易的事情。近年来，一些工作通过采用参数反射模型或利用反射特性来处理各种双向反射分布函数（BRDFs）。一个限制是，由于大多数方法采用了迭代方案或优化框架，正常估计的计算成本很高。当大量的光源被用于光度立体技术时，这个限制就成为一个实际问题。

该文提出了一种高效的光度测量立体方法，适用于具有一般反射率的表面，在已知的光线方向条件下。在该方法中，法线估计被表述为核回归，进一步表述为特征分解。尽管核回归在图像修复中得到了应用，但是核回归还没有被引入到光度测量立体中。为了适当地描述各种材料的反射率，最佳的核参数是通过留一（LOO）交叉验证来确定的。在LOO过程中，通过快速的矩阵操作加速了反格拉姆矩阵的计算，并通过适当的法线初始化加快了特征分解的速度。通过这种方式，所提出的测光立体方法的计算效率得到了很大的提高。

**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。

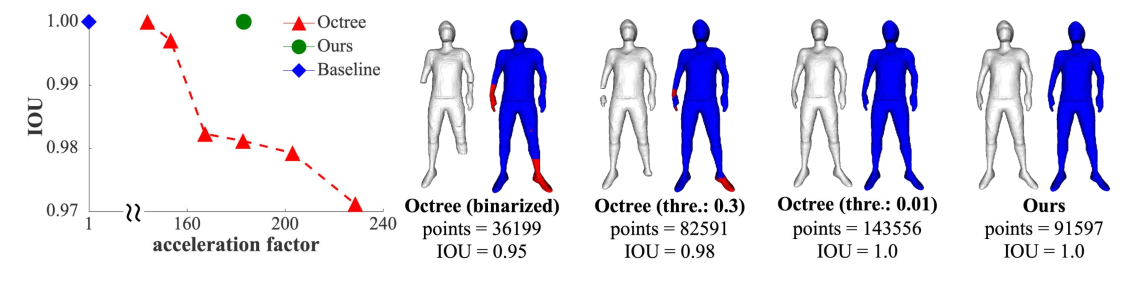
**2.1像素一致隐式函数（PIFu）**

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**

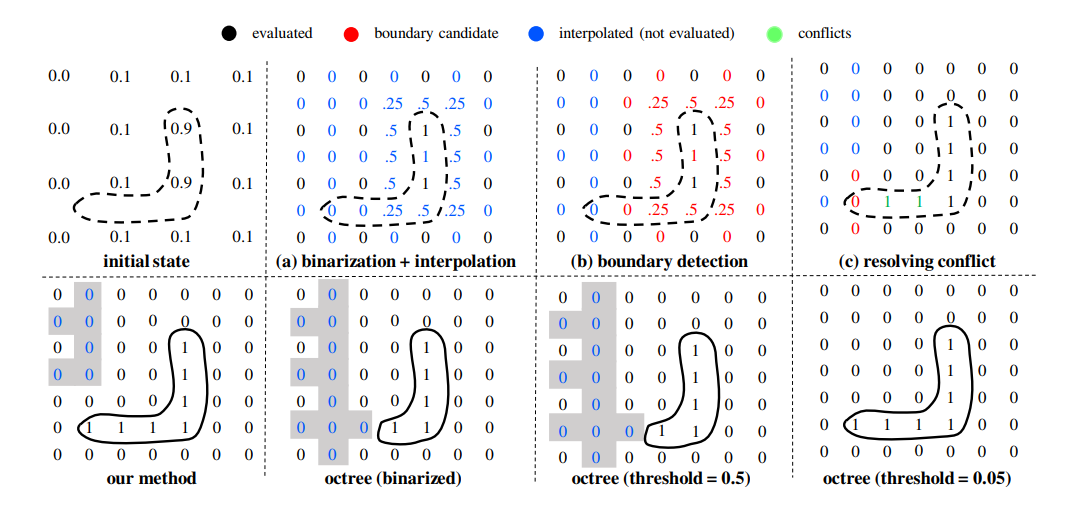
为了减少实时性能捕获所需的计算量，我们引入了两种新的加速技术。首先，我们提出了一种有效的表面定位算法，该算法保留了蛮力重构的精度，且复杂度与基于基础八叉树的重构算法相同。此外，由于我们的最终输出是从新的角度进行渲染，因此我们通过直接从PIFu生成新视图渲染来绕过显式网格重建阶段。通过结合这两种算法，我们成功实现了任意角度的实时渲染。

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）**

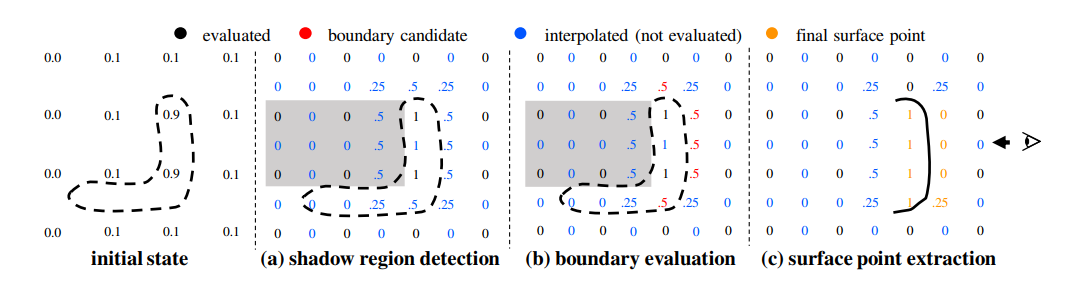
还有一种算法是当邻近粗网格的最大绝对偏差超过了某个阈值时细分网格，这种算法能够平衡好重建的速度和准确度，但却计算了更多不必要的点。



**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.