

基于图像和分布的大规模数据体绘制 (Image and Distribution Based Volume Rendering for Large Data Sets)

作者: Jiang Zhang 日期: 2018年4月21日

随着高性能计算技术的发展，领域科学家可以使用超级计算机对一些物理现象进行建模，并进行高分辨率的模拟。通过对模拟输出结果进行可视化和分析，科学家可以对这些物理现象有更深入的理解。但是，由于数据规模的不断增大，网络和存储设备的有限带宽和容量成为了一个重要的瓶颈。针对这一问题，研究者提出了原位数据可视化和分析的方法。这种方法使用相同的超级计算机资源，可以在不用移动原始数据的情况下生成紧凑型的数据代理(proxy)。之后，用户只需要对这些数据代理在后处理机器上进行分析。这里面，一类有效的方法是基于图像的方法。通过预先选择几个有意义的视角，其可以从原始数据中生成图像用于后处理分析。然而，这些方法在使用传递函数发现和分析被遮挡特征真一方面的能力非常有限，并且产生的基于图像的代理也会造成不可避免的信息损失，加大了对数据进一步分析的复杂度。为了解决这个问题，今年PacificVis会议上的一篇文章[1]提出了一种基于图像和分布的大规模数据表示方法，将原始数据存储为基于图像的数据代理，可以在资源有限的后处理机器上进行传递函数探索，并对其中引入的误差进行量化和可视化。

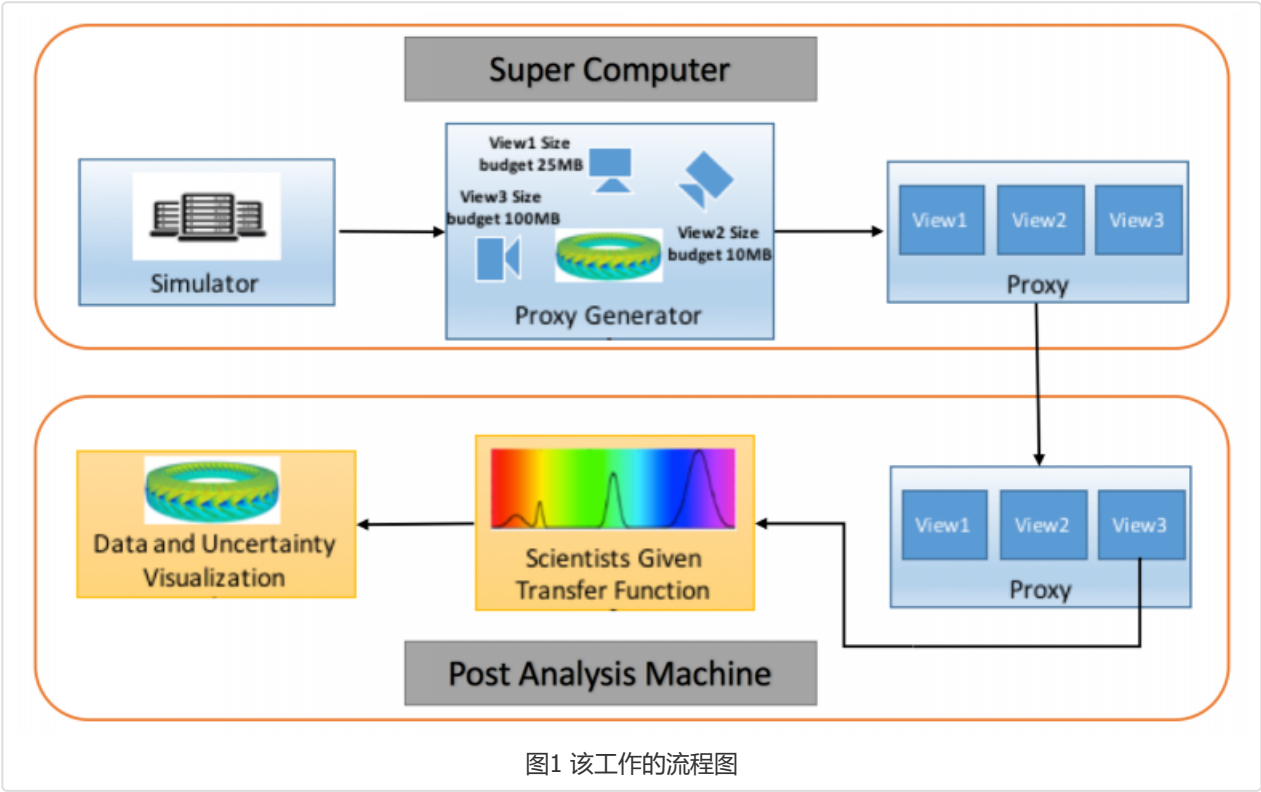
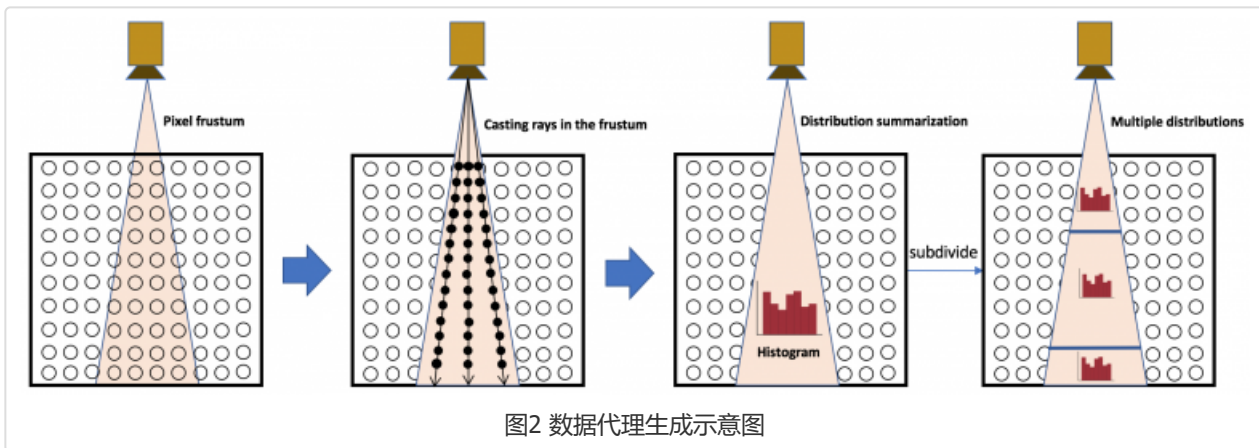
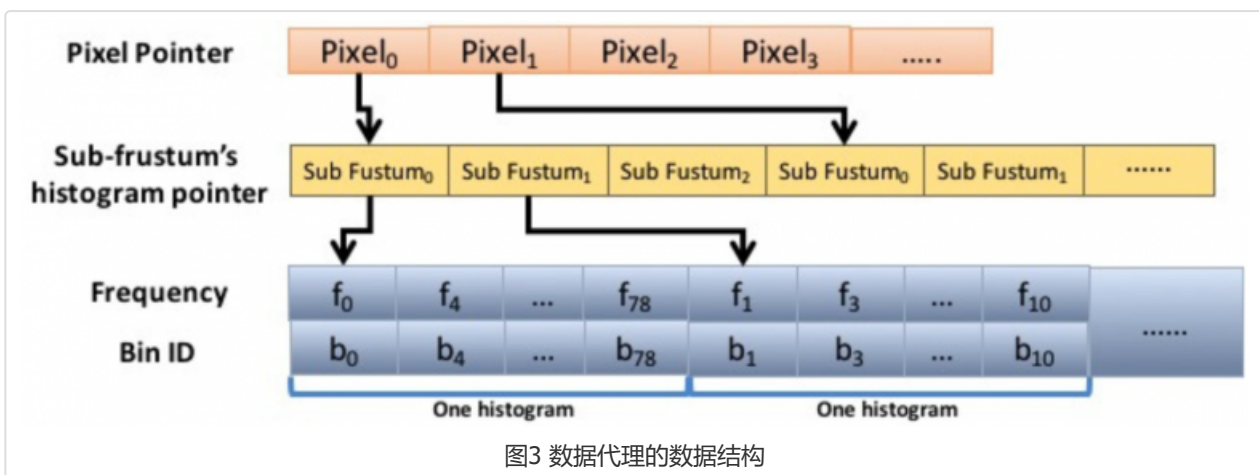


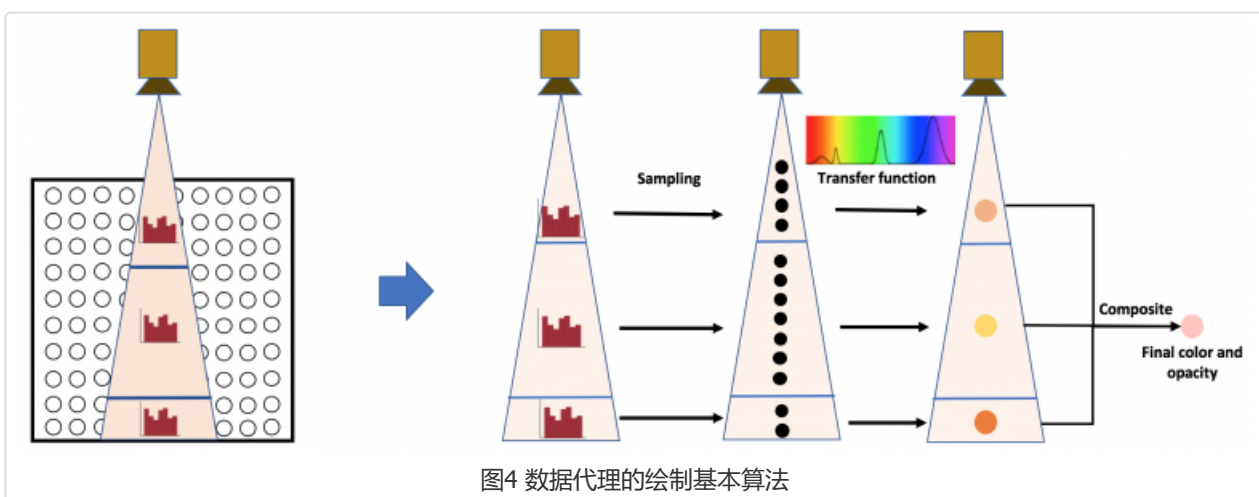
图1 该工作的流程图



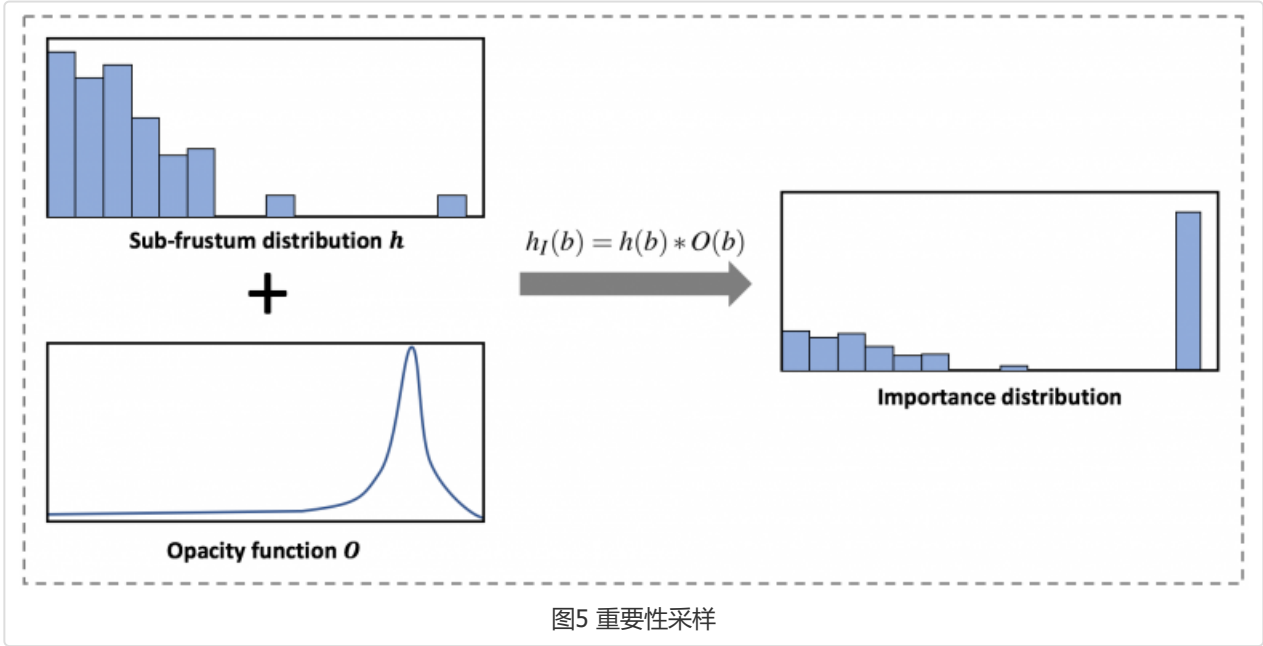
如图1所示, 该方法的流程主要分为两个部分。第一部分是在超级计算机上生成数据代理。在该方法中, 基于图像的数据代理在每个选定的视角下进行构建 (如图2所示)。对于每个视角, 其定义一个目标像素分辨率的图像面。在每个像素上, 进行多条射线的采样, 这些射线的范围形成一个像素平截体(frustum)。在该平截体中, 采集沿着这些射线的所有数据点, 然后使用统计直方图将其进行汇总, 得到一个分布。但是, 仅仅使用一个分布来总结像素平截体中的数据点会丢失数据的深度信息。为了保留深度信息, 像素平截体被进一步细分为多个子平截体。每个子平截体包含一个分布, 汇总较小范围内的信息。每个像素中的子平截体互相不重叠并且按深度排序。



为了减少数据代理的误差, 在该工作中, 像素平截体是根据其标量值的“随机性”进行细分的。这种“随机性”是通过评估所关联的直方图的Shannon熵所得到的。只要熵超过某一个阈值, 像素平截体便会被细分, 开始形成一个新的子平截体。由此, 不同像素平截体之间的细分的数量也可能会不同。经过细分得到所有的子平截体后, 每个视角和时间步下的数据代理会被组织成如图3所示的数据结构。该结构由四个数组组成, 包括图像面上的像素列表, 每个像素的子平截体列表, 以及每个子平截体所关联的直方图分布。



在生成数据代理后，该工作流程的第二部分即是在后处理机器上对这些数据代理进行可视化。如图4所示，其基本算法是，通过按照前后顺序绘制相关的像素子平截体来计算每个像素处的颜色。对于每个像素子平截体，使用Monte Carlo采样，数据采样点会被重构，之后每个重采样的值会被用于传递函数并最终合成得到像素的颜色和不透明度。



但是，使用Monte Carlo采样从分布中重构的采样点偏向于频率高的那些标量值，但这些频率高的标量值并不一定在科学家感兴趣的区域。因此，为了突出科学家感兴趣的标量值，该方法也使用了一种重要性采样方法。也即是，在采样过程中同时结合标量值的频率和其不透明度。这样，根据给定的不透明度函数和像素子平截体的分布可以计算得到一个重要性分布（如图5所示）。在这种情况下，只有同时具有高出现概率和高不透明度的标量值才更加重要。

	Size		Time (Average)	
	Raw data (total)	Proxy (total)	Proxy generation	Proxy rendering
Isabel	385GB	14.2GB	73.3s	0.12s
Plume	189GB	8.9GB	166.1s	0.075s
Combustion	407GB	15.2GB	156.5	0.078s
Turbine	271GB	15.3GB	275s	0.19s

图6 四种数据集的性能参数

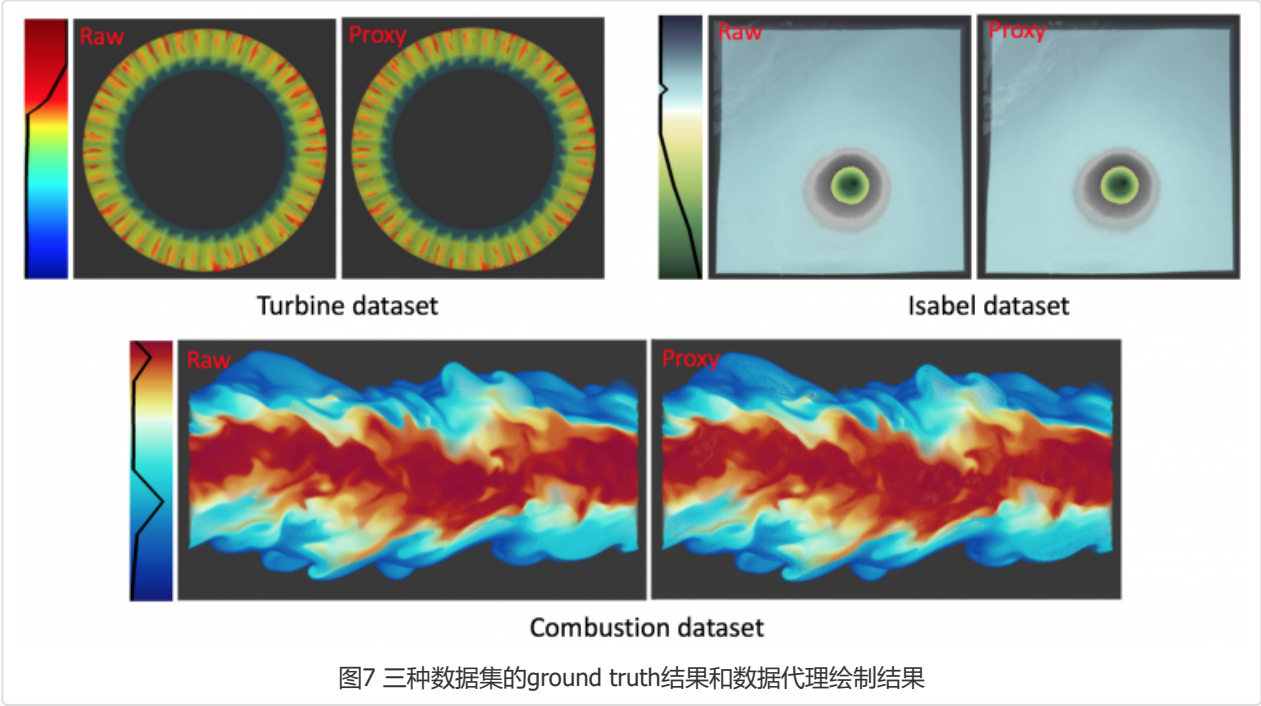


图6给出了该工作所使用的四种数据集的性能参数，包括数据代理的大小，数据代理生成和绘制的平均时间。从中可以看到，其生成的数据代理的规模远小于原始数据。图7展示了其中三种数据集的ground truth结果和从数据代理绘制的结果，可以看到在每个数据集中两种结果是非常相似的。

总结来讲，该工作提出了一种基于图像和分布的数据代理生成方法，并采用了基于重要性的采样，可以在资源有限的后处理分析机器上高效地进行体绘制。不过，这种方法比较简单，作者也并没有将其与其他方法进行对比，因而很难评估其优势。对于其中的视角选择和数据代理生成，我们也许可以考虑使用机器学习的方法，在不需要科学家先验知识的情况下自动进行处理，提高数据代理的质量。

Reference

[1] Ko-Chih Wang, Naeem Shareef, and Han-Wei Shen. Image and Distribution Based Volume Rendering for Large Data Sets. In Proceedings of IEEE PacificVis 2018, 2018.

论文报告 体绘制, 图像和分布, 大规模数据, 数据代理

← One Reality: 通过结合多种混合现实方式增强物理世界体验 (One Reality: Augmenting How the Physical World is Experienced by combining Multiple Mixed Reality Modalities)

Data Illustrator: 使用延迟数据绑定增强矢量设计工具以实现可视化创作 (Data Illustrator: Augmenting Vector Design Tools with Lazy Data Binding for Expressive Visualization Authoring) →

评论关闭。

RSS 订阅

功能

- 登录
- 文章RSS
- 评论RSS
- WordPress.org