《Instant 3D Photography》读书报告

李熙恺 21951052

1. 研究动机

虚拟现实（VR）是一种令人着迷的新兴技术，可以在沉浸式虚拟环境中创造栩栩如生的体验，并且高端耳机现已广泛可用。当今VR中消耗的大部分内容都是合成的，需要由专业艺术家创建。对于消费者而言，没有任何可行的方式来捕捉和分享他们自己的现实生活，从而使VR技术可以充分利用现实环境。当考虑到此问题的可成形性时，这可能不足为奇：对于不需要新颖的用户就不需要昂贵的硬件且易于使用的方法，即使对于新手用户也很容易使用。然而，它应该创建高质量，真正令人身临其境的内容，即支持双目视觉和头部运动视差的3D表示法，几乎是一秒钟。全景影像和视频现在可以使用消费类360°摄像机轻松捕获。虽然周围的影像提供了一定的沉浸感，但由于缺乏深度和视差，因此逼真度受到限制。

在本文中，作者提出了一种新算法，该算法可根据从小型基线立体双摄像头手机（例如最近的iPhone）生成的彩色和深度照片序列构造3D全景图。

1. 论文算法解决方案

文章中的3D全景建筑算法在以下四个阶段进行：

**1.捕获和预处理**

彩色和深度图像对对齐的算法总序列的输入，它使用自定义的突发捕获应用程序从有关点继续摄像头的捕获。论文中使用iPhone 7 Plus并使用自定义的基本捕获应用程序进行四次捕获，在捕获会话期间，它会在1秒钟的时间间隔内自动触发捕获彩色和深度照片（使用当时的iOS立体算法）。 捕获的动作类似于今天人们捕获的全景照片：将照相机指向外面，同时将设备保持在半臂的长度上，并以任意角度向上，向下或侧向移动来扫描场景。特征提取和匹配：作为以下对齐算法的输入，我们使用标准方法计算成对特征匹配。

**2. 变形深度对齐**

由于相机的基准线较小且导致三角测量的不确定性，因此输入深度图并非十分准确，并且无法使用全局转换来调整它们的深度。作者使用无角度优化方法解决了这个问题，该方法联合估计了相机的姿势以及适用于深度的各种变化的调整图。

首要目标是对齐深度图。 由于图像是从不同的角度拍摄的，因此无法在2D图像中处理视差。该步骤恢复外部相机的姿势（方向和位置），以便在投影深度图时将其对准3D。

**3. 拼接**

接下来，将针脚对准的颜色和深度照片缝合到铝马赛克上。通常，这是通过离散优化方法解决的标记问题。但是，即使问题缩小，使用MRF求解器的平滑度优化也非常缓慢。 我们使用经过精心设计的数据项和高质量的深度对齐方式，以深度引导的边缘感知方式对数据项进行滤波后，通过独立优化每个像素来替代标签平滑度优化。这可以在视觉上获得相似的结果，并且幅度提高了一个数量级。

首先，我们通过向后追踪摄像机前向向量并找到最小的3D点来计算全景投影的中心，然后从该中心视点开始绘制所有颜色和深度图，以获取矩形的全景图。将针迹公式化为离散的标记问题，希望在全景图像的源图像α p中选择每个像素p，然后从中提取颜色和深度。

**4.多层网络生成**

在最后一个阶段，作者将panora主体转换为可以使用标准图形引擎在任何设备上使用的多层纹理网格。 我们在很深的深度边缘撕开网格，将背面扩展到封闭的区域，产生新的幻觉，最后，作者简化主题手计算固定图集。

最后的算法步骤将潘诺拉网转换为可以使用标准图形引擎在任何设备上渲染的三角形网格。通过将所有像素连接到4个相邻的像素点，从而在视点变化时露出的深度较大的三角形边缘上创建了一个三角形网格。文章中的解决方案类似于 “两层合并”算法。作者发现，在一个基本不变的领域上，有很大不同的观点表明，针脚所占的比例却很小。如果较小的摄像机，则图像之间的重叠很少。如果使用了更好的基线或更广角摄像机来捕获场景，则上面提到的两层缝合合并算法可以适应运行时间较短的情况。

不规则算法中，每个顶点的顶点表示对应的像素位置，即“弹出”以达到一定深度。每个顶点都连接到四个基本方向的菱形附近。由于我们的目标生成多个层，因此在不同深度上可能存在多个顶点，因此单个像素位置不同。 作者通过为每个全景像素创建顶点并将其连接到4个邻居来初始化网格。

1. 论文创新点

1.由于双摄手机的每个较小的基线深度估计值都高度不确定，因此需要对边缘进行过滤以平滑产生的噪声。但是，这会导致深度图中的低频错误，从而避免使用全局变换进行简单对准。该论文提出了一种新的优化方法，将其深度值始终保持一致，以便将其值重新分配到各个位置。

2.现有的使用离散优化的图像融合方法速度很慢。论文作者利用经过精心设计的数据项和高质量的深度校正来满足标签平滑度优化的需要，并在过滤深度向导的边缘数据后独立地优化了每个像素标签，从而使幅度提高了两倍以上。

1. 论文应用效果

作者已使用iPhone7 Plus捕获并处理了数十种场景，其中包括此提交内容中的25个。这些场景跨越了很多种不同的环境（室内和室外，城市和自然），并捕获了整个场景（从360°到120°的整个白天和黑夜），从大约20个场景到整个场景。

1. 论文的局限性

论文的结果显示出与其他3D重建系统类似的伪像。特别是几何形状的海反射块，无纹理区域的深度不正确，动态对象上的伪影。与现有系统相比，这些问题有所减少，但仍然存在。 要仔细检查这些文物，建议在补充材料中观看视频比较。同时算法具有一定的局限性，可以为未来的工作带来有趣的途径。 捕获：iPhone摄影机在深度捕获模式下具有狭窄的视野，这是因为一个广角镜和热成像镜头。如果广角镜头需要捕获较少的图像以实现相同的重叠量。同时基线会增加，从而使构造问题变得更容易。

1. 总结

在本文中，作者提出了一种从彩色和深度图像序列生成3D全景图的快速端到端算法。尽管这些深度图包含相当大的流量频率误差，但可变形的对齐方式的优化能够精确对齐它们。这为用数据的深度引导边缘感知滤波替换订书机中的离散平滑度优化以及独立优化每个像素（实现幅值加速的二阶）提供了可能性。为进一步改进和研究这项工作的许多途径感到可能。考虑到性能讨论，可以直接与电话交互进行交互实现。

论文已经知道如何快速捕获和重建系统的可用性相对于3D场景捕获而改变。捕获场景将变得更具机会主义和冲动性。几乎所有提供的场景都在没有计划的情况下被捕获，例如在旅行时 人们尤其对使用单图像深度图获得有希望的第一结果。这些深度图的质量仍然很低。但对齐算法能够将它们合并，并进一步缝合减少了伪影的共识数据。改进了这些结果，为进一步的研究提供了令人感兴趣的方向，并保持了将3D摄影技术带到数十亿部带有单目相机的普通手机的承诺。