

使用精确的多视图 3D 重建 适用于大型场景的基于补丁的立体

沉书涵

抽象的—在本文中, 我们提出了一种针对大规模场景的基于深度图合并的多视图立体方法, 该方法同时考虑了**准确性**和**效率**考虑到。在所提出的方法中, 使用基于块的高效立体匹配过程来在每个图像上生成具有可接受误差的深度图, 然后进行深度图细化过程以强制相邻视图的一致性。与最先进的方法相比, 所提出的方法可以以高计算效率重建相当准确和密集的点云。此外, 所提出的方法可以很容易地在图像级别并行化, 即每个深度图都是单独计算的, 这使得它适合于高分辨率图像的大规模场景重建。该方法的准确性和效率在基准数据上进行定量评估, 在大数据集上进行定性评估。

索引/术语—3D 重建、深度图、多视图立体 (MVS)。

我、我简介

瓦现代数码相机的快速发展, 如今可以轻松捕获大量高分辨率图像。许多应用迫切需要从这些图像中提取 3D 结构, 例如建筑遗产保护、城市规模建模等。多视图立体 (MVS) 重建是基于图像的 3D 采集的关键步骤, 最近受到越来越多的关注。尽管人们在 MVS 方面做出了巨大的努力, 并提出了一些有效的算法, 特别是对于小型和紧凑的物体, 但使用高分辨率图像 (6 兆像素及以上) 处理大规模场景仍然是一个悬而未决的问题。

根据[1], MVS算法可以分为四类, 称为基于体素的方法[2]-[4]、基于表面进化的方法[5]-[8]、基于特征点生长的方法[9]-[13], 以及基于深度图合并的方法[14]-[22]。在这些类中, 基于体素的方法仅适用于紧密封闭的盒子内的小型紧凑物体; 基于表面演化的方法需要可靠的初始猜测, 这对于大规模场景来说很难获得; 特点

稿件于2012年6月19日收到; 2012年12月20日修订; 2012年12月28日接受。发表日期2013年1月11日; 当前版本日期为2013年3月14日。这项工作得到了中国自然科学基金委 61105032、国家 973 重点基础研究计划 2012CB316302 和中国科学院战略性先导研究计划的部分支持科学, 根据拨款 XDA06030300。Bulent Sankur 教授负责协调该手稿的审查并批准其出版。

作者单位: 中国科学院自动化研究所模式识别国家实验室, 北京 100190 (e-mail: shshen@nlpr.ia.ac.cn)

本文中一幅或多幅图的彩色版本可在线获取: <http://ieeexplore.ieee.org>。
数字对象标识符 10.1109/TIP.2013.2237921

基于点生长的方法将纹理区域中重建的点传播到无纹理区域, 这可能会在最终结果中留下漏洞; 基于深度图的方法已被证明更适合大规模场景, 但在准确性和完整性方面其性能通常低于其他方法。

在本文中, 我们提出了一种针对大规模场景的基于深度图合并的 MVS 方法, 该方法同时采用**准确性**和**效率**考虑到。我们方法的关键是基于深度图细化过程的高效基于补丁的立体匹配, 该过程增强了多个视图的一致性。与最先进的方法相比, 所提出的方法具有三个主要优点: 1) 它可以重建非常准确和密集的点云, 因为基于补丁的立体能够生成具有可接受误差的深度图, 并且可以进一步细化通过深度图细化过程。2) 它是一种计算高效的方法, 在达到相似精度的同时, 比最先进的方法[11]快约 10 至 20 倍。3) 它可以很容易地在图像级别并行化, 即每个深度图都是单独计算的, 这使得它适合于高分辨率图像的大规模场景重建。

二. 磷回顾瓦奥克斯

根据[1]中给出的分类法, 本节回顾了四类 MVS、基于体素的方法、基于表面演化的方法、基于特征点生长的方法和基于深度图合并的方法。

基于体素的方法计算 3D 体积上的成本函数, 该体积是对象的边界框。塞茨等人。[2] 提出了一种体素着色框架, 该框架以广义深度顺序遍历离散 3D 空间, 以识别具有独特颜色的体素, 该颜色在场景的所有可能解释中保持不变。沃吉亚齐斯等人。[3]使用图割优化来计算包含最大可能体积的最小表面, 其中表面积只是该光一致性场中的表面积分。由于这些方法的准确性受到体素网格分辨率的限制, Sinha 等人。[4]提出了一种方法, 不需要表面位于视觉船体周围的有限带内。该方法使用照片一致性来指导包围体的粗网格的自适应细分, 它生成一个多分辨率体积网格, 该网格在可能包含未知表面的部分中密集地镶嵌。然而, 这种方法仅适用于容纳紧密封闭框的紧凑对象, 并且其计算和内存成本对于大规模场景来说变得过高。

基于表面演化的方法迭代地演化初始猜测以改进照片一致性测量。

1057-7149/\$31.00 © 2013 IEEE

福格拉斯等人。[5] 实现水平集来求解一组偏微分方程, 这些偏微分方程用于使一组初始表面变形, 然后将其移向要检测的对象。埃尔南德斯等人。[6]提出了一种基于纹理和轮廓信息的方法, 并将轮廓力融合到snake框架中。该方法使用纹理和轮廓驱动力演化出足够接近物体表面的初始表面。希普等人。[7]使用最小 st 切割来生成粗略的初始网格, 然后使用变分方法对其进行细化以捕获小细节。这种方法的一个主要缺点是需要可靠的初始猜测, 而这对于室外场景来说很难获得。为此, Cremers 等人。[8] 将重建问题表述为凸函数最小化, 其中精确的轮廓一致性被强加为限制可行函数域的凸约束。该方法不依赖于初始化, 并且可以提供位于最优解的误差范围内的解。然而, 该方法依赖于空间的体素表示, 因此不能用于大规模场景。

基于特征点生长的方法首先重建纹理区域中的点, 然后将这些点扩展到无纹理区域。卢利尔等人。[9]提出了一种准密集的 3D 表面模型获取方法。该方法首先初始化稀疏对应兴趣点, 然后从准稠密视差图中重新采样准稠密点, 以稠化特征点以克服兴趣点的稀疏性。戈塞勒等人。[10] 提出了一种使用全局和局部视图选择以及来自重建 SIFT [23] 特征的区域生长过程来处理包含障碍物的互联网照片集的方法。基于这些方法, Furukawa 等人。[11] 提出了一种相当准确的基于补丁的 MVS (PMVS) 方法, 该方法从一组稀疏的匹配关键点开始, 并重复扩展它们, 直到调用可见性约束来过滤掉错误匹配。该方法现在被认为是最先进的 MVS 方法。基于 PMVS, Wu 等人。[12]提出了一种用于准密集3D重建的基于张量的 MVS (TMVS) 方法, 该方法结合了3D张量框架下MVS中光一致性、可见性和几何一致性执行的互补优势。这些基于特征点增长的方法旨在通过同时使用所有可用图像来重建全局 3D 模型, 因此随着图像数量的增长, 它们会遇到可扩展性问题。虽然这个问题可以通过将输入图像分解为具有小重叠的簇来部分解决[13], 但对于大规模场景来说, 计算复杂度仍然相当高。

基于深度图合并的方法是从双目立体到多视图的自然延伸。此类方法首先计算每个视图的深度图, 然后考虑可见性将它们合并到单个模型中。戈塞勒等人。[14]使用基于归一化互相关 (NCC) 的像素窗口匹配技术来生成深度图, 然后将它们与体积积分合并。斯特雷查等人。[15]将深度和可见度联合建模为隐马尔可夫随机场, 并使用 EM 算法来优化模型参数。梅雷尔等人。[16]首先使用计算成本低的立体算法来生成潜在的

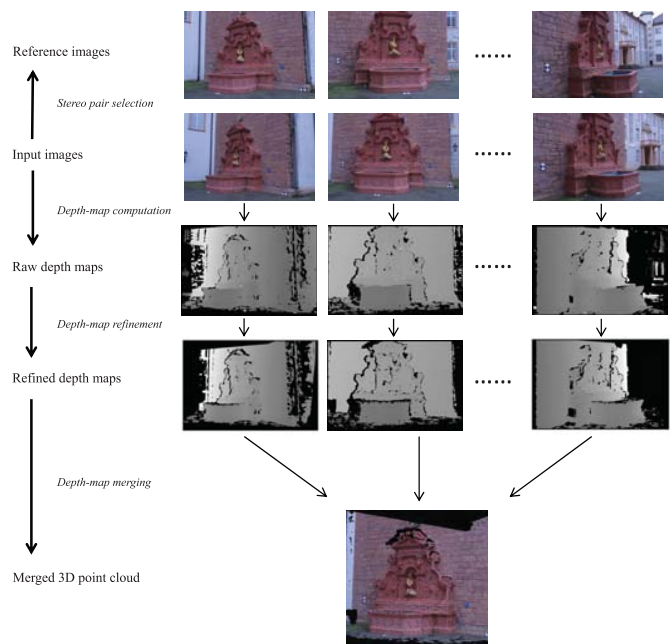


图. 1。所提出方法的框架。

噪声、重叠的深度图, 然后融合这些深度图以获得基于点之间的可见性关系的集成表面。扎克等人。[17]提出了一种全局优化能量泛函的方法, 该能量泛函由总变分正则化力和 L_1 数据保真度术语。布拉德利等人。[18]提出了一种方法, 该方法使用具有缩放匹配窗口的鲁棒双目立体, 然后对合并点云进行自适应基于点的过滤。坎贝尔等人。[19]为每个像素存储多个深度假设, 并使用空间一致性约束来提取离散马尔可夫随机场框架中的真实深度。刘等人。[20]使用变分光流生成的连续深度图产生高质量的 MVS 重建结果。但该方法需要视觉外壳作为初始化。李等人。[21]使用DAISY [24]功能生成深度图, 并使用两个阶段的束调整来优化3D点的位置和法线。托拉等人。[22]还使用DAISY功能来生成深度图, 然后通过相邻视图的一致性检查来合并它们。

三. MVS U唱歌磷阿奇乙东盟经济发展局S特雷奥

该方法由四个步骤组成: 立体对选择、深度图计算、深度图细化和深度图合并。该方法的框架如图1所示。对于输入图像集中的每个图像, 我们选择一个参考图像来形成立体对以进行深度图计算。由于立体视觉生成的这些原始深度图可能包含噪声和错误, 因此我们通过使用其相邻深度图进行一致性检查来细化每个深度图。最后将所有细化的深度图合并在一起

以获得最终的重建。接下来我们详细说明每个步骤。

A. 立体对选择

对于图像集中的每个图像，我们需要为其选择一个参考图像来进行立体计算。立体图像对的选择不仅对于立体匹配的准确性很重要，而且对于最终的MVS结果也很重要。对于车辆上的街边摄像头 [25]-[28] 或受控环境（如 Middlebury 基准数据 [1]）中的摄像头来说，立体对选择是一项相对容易的任务，但需要针对无序图像进行仔细设计。一个好的候选参考图像应具有与目标图像相似的观看方向，并且具有合适的基线，既不能太短而导致重建精度下降，也不能太长而导致场景的共同覆盖范围较小。

我们使用类似于[21]的方法来选择符合条件的立体对。假设我们有 n 图像，并且对于我-第一个，我们计算 $\theta_{ij}, j=1, \dots, n$ 这是相机主视角方向之间的角度我和 j 。如果使用运动结构（SfM）算法[29]-[31]来校准相机姿态，一组稀疏的3D点及其可见性作为SfM的副产品生成，那么更好的 θ_{ij} 可以分别计算为相机的可见点和相机中心之间的角度的平均值我和 j 。除了 θ_{ij} 我们计算另一个参数 $d_{ij}, j=1, \dots, n$ 对于每个图像 i ，这是相机光心之间的距离我和 j 。然后对于满足 5° 的图像 $< \theta_{ij} < 60^\circ$ ，我们计算中位数 d 他们的 d_{ij} 并删除其图像 $d_{ij} > 2d$ 或者 $d_{ij} < 0.05d$ 。经过这些计算，如果剩余图像的数量小于 k_1 ，它们被视为图像的相邻图像 i ，表示为（你）。否则，剩余图像按照升序排序 $\theta_{ij} \cdot d_{ij}$ 和第一个 k_1 图像由相邻图像组成（你）（在本文中我们设置 $k_1=10$ ）。最后，最小的一个 $\theta_{ij} \cdot d_{ij}$ 之中（你）被选为我-图像的参考图像形成立体对。

B. 深度图计算

对于每个符合条件的立体对，我们遵循[32]中的想法来计算深度图。核心思想是，对于输入图像中的每个像素，我们尝试找到一个与参考图像具有最小聚合匹配成本的良好支持平面，如图2所示。

支撑平面 F 本质上是场景表面的局部切平面，由3D点表示 X_i 和它的正常现象 n_i 在相关相机的坐标系中，如图3所示。

为了我-第一个输入图像 i 我在图像集中，给定其参考图像 j ，以及相关的相机参数 $\{K_i, R_i, C_i\}$ 和 $\{K_j, R_j, C_j\}$ ，在哪里 K 是内在参数， R 是旋转矩阵，并且 C 是相机中心，我们首先分配每个像素 p 在我-到随机3D平面。认为

p 的齐次坐标为：

$$p = \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

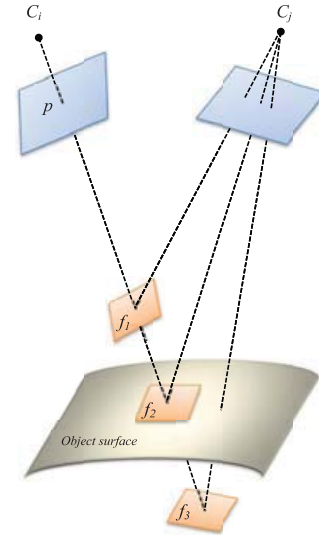
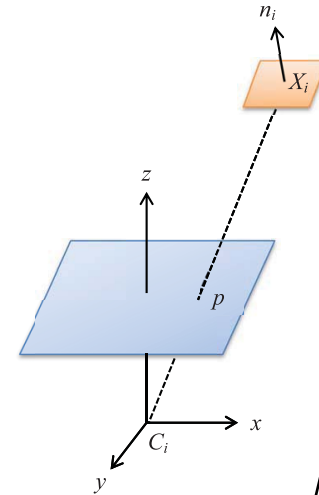


图2. 对于每个像素 p 在输入图像中，我们估计其对应的3D平面。 C_i 和 C_j 分别是输入图像和参考图像的相机中心， $F1$ 、 $F2$ 和 $F3$ 是三个3D平面 p 的观察光线。明显地 $F2$ 具有最小的聚合匹配成本。



像素坐标转换为物理坐标？

图3. 支撑平面由3D点表示 X_i 和它的正常现象 n_i 在相机中 C_i 's坐标，其中 C_i 是相机的中心我-第一个输入图像，以及 C_i -XYZ是相机的坐标。

3D点 X_i 我必须位于视线内 p ，我们选择一个随机深度 λ 在深度范围内 $\lambda \in [\lambda_{\text{min}}, \lambda_{\text{max}}]$ ，然后 X_i 计算为 C_i 's坐标为：

$$X_i = \lambda K_i^{-1} p \quad (2)$$

然后我们在相机中随机分配平面的法线 C_i 's球坐标为：

$$n_i = \begin{bmatrix} \cos \theta \sin \phi \\ \sin \theta \sin \phi \\ \cos \phi \end{bmatrix} \quad (3)$$

在哪里， θ 是 $[0^\circ, 360^\circ]$ 范围内的随机角度， ϕ 是范围 $[0, 60^\circ]$ 。这些范围设置来自一个简单的假设，即图像中的补丁是可见的 i 我

当补丁法线之间的角度 n_i 和 z 轴

(1)相机 C_i 's坐标系低于某个阈值（本文我们将这个阈值设置为 60° ）。

上述随机初始化过程很可能对图像中的每个场景平面至少有一个好的猜测, 特别是对于每个场景平面包含更多像素的高分辨率图像, 这意味着比低分辨率图像有更多的猜测。我们应该注意的是, 一旦图像的深度图*我*计算后, 我们可以改进计算深度图时的纯随机初始化过程*我*参考图像*我₀*。在这种情况下, 每个像素的深度和补丁法线*我_s*深度图可以扭曲为*我*作为计算时的初步估计*我_s*深度图和像素*我_j*之间没有映射*我*和*我_j*

仍然使用随机初始化。通过这种方式, 我们可以将每个扭曲像素分配给我最初的平面比随机猜测更好, 因为该平面对于立体声来说是一致的*我*和*我_j*。

根据[33], 给定两个相机的投影矩阵*磷*=[*我*_{3×3}|**0**₃]和*磷*=[*右*_{3×3}|*右*_{3×1}], 和一个定义为的平面*π*:*时间**X*=0 与 *π*= (*V*_{*时间*}, 1) *时间*, 那么单应性*H*由平面引起的为:

$$H = \text{右} \cdot \text{电视} \quad (4)$$

这里, *我*_{3×3}是 3×3 单位矩阵和**0**₃是一个零 3 向量, 这表明选择世界坐标与相机一致*磷*。本文中, 图像对的相机参数为{*K*_{*我*}, *R*_{*我*}, *C*_{*我*}}和{*K*_{*右*}, *R*_{*右*}, *C*_{*右*}}, 和飞机 *F_p*={*X*_{*我*}, *n*_{*我*}}是在相机中定义的*C*_{*我*}'s 坐标。因此, 投影矩阵和平面参数可以转换为标准形式 (将世界原点放在*C*_{*我*}), 作为:

$$\text{磷} = K[\text{我}_{3 \times 3} | \mathbf{0}_3], \text{磷} = K[\text{右}_{3 \times 3} | \text{右}_{3 \times 1} \quad \text{我}_{3 \times 1} (C_{\text{右}} - C_{\text{我}})],$$

$$V_{\text{时间}} = - \frac{n_{\text{时间}}}{n_{\text{时间}} \text{我}}$$

根据方程. 4、相机的单应性*P*_{*我*}=[*我*_{3×3}|**0**₃]和*P*_{*右*}=[*右*_{3×3}|*右*_{3×1}]*我* [*右*_{3×1} (*C*_{*右*} - *C*_{*我*})] 是:

$$H_{ij} = \text{右}_{j \text{ 右}-1 \text{ 我}} + \frac{\text{右}_j (\text{我} - C_{\text{我}}) n_{\text{时间}}}{n_{\text{时间}} \text{我}}$$

应用转换*K*_{*我*}和*K*_{*右*}对于我们给定范围的图像{*λ*, *θ*, *φ*}, 我们1) 选择一个随机平面获取相机*磷*_{*我*}=*K*_{*我*}*P*_{*我*}, *磷*_{*右*}=*K*_{*右*}*P*_{*右*}以及由此产生的范围{*λ*, *θ*, *φ*}在范围中*λ* ∈ [*λ* - *θ*, *λ* + *θ*], *θ* ∈ [*θ* - *φ*, *θ* + *φ*], *φ* ∈ [*φ* - *φ*, *φ* + *φ*]. 2) 计算飞机*F*_{*我*}={*X*_{*我*}, *n*_{*我*}}使用等式2和等式. 3.3) 如果米(*p*, *f_p*) < 米(*p*, *f_p*), 我们接受 *F_p*=*F*_{*我*}并设置*λ*=*λ*, *θ*=*θ*, *φ*=*φ*. 4) 我们将范围减半{*λ*, *θ*, *φ*}. 5) 返回第一步。重复上述过程为*k*₃次。在本文中, 我们将初始范围和重复时间设置为:

$$H_{ij} = K_j (R_{j \text{ 右}-1 \text{ 我}} + \frac{\text{右}_j (C_{\text{我}} - C_j) n_{\text{时间}}}{n_{\text{时间}} \text{我}}) K_{\text{我}} \quad (5)$$

我们设置一个方形的窗口*乙*以像素为中心*p*, 在哪里*乙*=*w* × *w* (在本文中我们设置*w*=7 像素)。对于每个像素*q*在*乙*我们在参考图像中找到其对应的像素*我*使用单应性映射*H_{ij}*(*q*)。然后聚合匹配成本米(*p*, *f_p*)对于像素*p*计算为 1 减去归一化互相关 (NCC) 分数*q*和*H_{ij}*(*q*), 作为:

$$\text{米} (p, f_p) = 1 - \frac{\sum_{q \in \text{乙}} (q - \bar{q})(H_{ij}(q) - \overline{H_{ij}(q)})}{\sqrt{\sum_{q \in \text{乙}} (q - \bar{q})^2 \sum_{q \in \text{乙}} (H_{ij}(q) - \overline{H_{ij}(q)})^2}} \quad (6)$$

请注意, 一些更复杂和更强大的聚合技术, 如 [34]-[36], 可以用来生成比 NCC 更可靠的结果。然而, 高分辨率图像可以

提供比低分辨率匹配更可靠的匹配, 并且简单的 NCC 足以可靠地测量光度一致性。此外, NCC 生成的大多数不可靠像素可以在深度图细化步骤中被去除, 这使得 NCC 的最终重建结果几乎等同于其他复杂聚合方法的重建结果。因此, 在本文中, 我们使用简单的 NCC 作为聚合匹配成本, 与[11]相同。

初始化后, 图像中的每个像素*我*与 3D 平面相关联。然后我们处理像素*我*逐一细化其中的飞机*m*迭代。在奇数次迭代中, 我们从左上角像素开始, 按行顺序遍历, 直到到达右下角像素。在偶数迭代中, 我们反转访问像素的顺序, 从右下角到左上角像素, 也是按行顺序。在本文中, 我们将平面细化的数量设置为*k*₂= 3。

在每次迭代中, 每个像素都有两个操作, 称为*空间传播*和*随机分配*。空间传播用于将相邻像素的平面与当前像素的平面进行比较和传播。在奇数迭代中, 相邻像素是左、上和左上邻居, 而在偶数迭代中是右、下和右下邻居。让*p*_{*我*}表示当前像素的邻域*p*, 和*F_p*_{*我*}表示*p*_{*我*}'s 平面, 我们使用式6中的匹配成本来检查条件米(*p*, *f_p*) < 米(*p*, *f_p*)。如果满足这个条件, 我们认为*F_p*_{*我*}是一个更好的选择*p*与当前平面相比*F_p*, 并传播*F_p*_{*我*}到*p*, 即设置*F_p*=*F_p*_{*我*}。这种空间传播过程依赖于这样一个事实: 相邻像素很可能具有相似的 3D 平面, 尤其是对于高分辨率图像。理论上, 即使是一个好的猜测也足以在第一次和第二次空间传播之后将该平面传播到该区域的其他像素。

对于每个像素*p*, 在空间传播之后, 我们进一步细化平面*F_p*使用随机分配。随机分配的目的是通过测试几个随机平面参数来进一步降低等式6中的匹配成本。

对于每个像素*p*, 在空间传播之后, 我们进一步细化平面*F_p*使用随机分配。随机分配的目的是通过测试几个随机平面参数来进一步降低等式6中的匹配成本。

空间传播和随机分配的思想已经成功应用于补丁匹配立体方法[32]和混合递归匹配 (HRM) 方法[37]。本文将[32]的思想扩展到多视图

使用高分辨率图像实现大范围场景的立体。因此, 所提出方法的新颖之处在于以适当的方式修改路径匹配立体算法[32], 使其对于大规模MVS问题更加强大和高效。

所提出的方法与[32]中的方法之间的主要区别在于平面是在图像坐标中定义的

在[32]中,但是在我们工作中的相机坐标中,因为立体对在我们的论文中没有得到纠正。此外,通过充分利用多个高分辨率图像,我们与[32]相比进行了三项简化,以减少计算费用。首先,聚合匹配成本 $\mu(p, f_p)$ 我们使用的是简单的归一化互相关,而不是[32]中更复杂的自适应支持权重版本,因为 NCC 对于高分辨率图像来说足够可靠,并且可以在接下来的深度图细化步骤中消除剩余的误差。其次,与[32]中的空间和视图传播相比,我们只使用空间传播,因为我们只计算深度图 d_i ,不在两者上 d_i 和 d_j 如[32]中所示。第三,[32]中的方法包含另一个后处理步骤,该步骤通过左/右一致性检查应用遮挡处理并填充无效像素。我们的方法没有这个过程,因为下面的深度图细化步骤可以达到类似的效果。

在空间传播和随机分配过程之后,我们删除深度图中聚合匹配成本高于特定阈值的不可靠点 τ_1 (在本文中我们设置 $\tau_1 = 0.3$)。

C. 深度图细化

由于深度误差,原始深度图可能在公共区域上彼此不完全一致,因此执行细化过程以强制相邻视图的一致性。对于每个点 p 在图像中 d_i ,我们将其投影为 3D

利用其深度 λ 以及相机参数,如:

$$X = \lambda R_i + C_i$$

在哪里 p 是方程中定义的齐次坐标。1、 X 是世界坐标中的 3D 点。然后我们投影 X 到 d_j 的相邻图像 N_j 。它们是在立体声对选择阶段生成的。认为 N_k 是个 k 第一个相邻图像 N_j , 我们表示 $d(X, N_k)$ 作为深度 X 相对于相机 N_k 并表示 $\lambda(X, N_k)$ 作为在 X 的投影处计算的深度值 λ 使用 N_k 的深度图。如果 $\lambda(X, N_k)$ 足够接近 $d(X, N_k)$, 即, $|d(X, N_k) - \lambda(X, N_k)| < \tau_2$, 在哪里 τ_2 是一个阈值 (在本文中我们设置 $\tau_2 = 0.01$), 我们说 X 是一致的 d_i 和 N_k 。如果 X 至少在以下时间内是一致的 N_k 中的相邻图像 N_j (在本文中我们设置 $k_4 = 2$), 视为可靠的场景点及其对应像素 p 在 d_i 的深度图被保留, 否则被删除。

经过上述细化过程后,可以消除大多数错误,从而在每个视图上产生相对干净的深度图。

D. 深度图合并

细化后,可以合并所有深度图来表示场景。然而,直接合并深度图可能包含大量冗余,因为不同的深度图可能具有共同的场景覆盖范围,特别是对于相邻图像。为了消除这些冗余,通过相邻深度图测试进一步减少深度图。如图4所示,对于相机中的每个像素

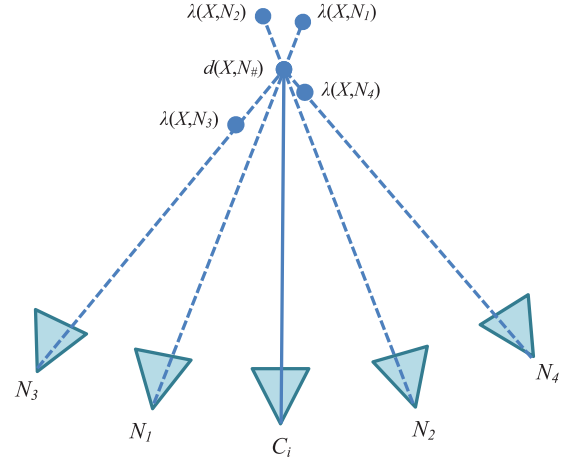


图4. 通过深度图测试去除冗余的图示。对于每个相机中的像素 s 深度图,我们将其反向投影为 3D X 使用方程式 (7) 并重新投影 X 到 s 的相邻相机 $N_{\#}$ 。定义 $d(X, N_{\#})$ 由深度 X 相对于相机 $N_{\#}$ 并定义 $\lambda(X, N_{\#})$ 通过在投影处计算的深度值 λ 在 $N_{\#}$ 使用 $N_{\#}$ 的深度图。如果 $d(X, N_{\#}) < \lambda(X, N_{\#})$, 我们说的投影 X 在 $N_{\#}$

被遮挡并将其从中移除 $N_{\#}$ 的深度图, 如 N_1 和 N_2 。如果 $|d(X, N_{\#}) - \lambda(X, N_{\#})| < \tau_2$, 我们说 X 重复于 s 和 $N_{\#}$ 并删除它从 $N_{\#}$ 的深度图, 如 N_4 。不满足上述两个条件的点将保留在其深度图中, 例如 N_3 。

C_i 的深度图, 我们将其反向投影为 3D X 使用方程式 7、重新投影 X 到 C_i 的邻近的摄像机。如果深度为 X 相对于相邻相机的深度值小于在投影处计算的深度值 λ 在相邻相机的深度图中, 就像相机一样 N_1 和 N_2 在

(7) 图4, 我们说的投影 X 在该相邻相机中被遮挡并将其从该相邻相机的深度图中删除。如果这两个深度值足够接近, 就像相机一样 N_4 在图 4 中, 我们称投影 X 在这个相邻的相机中代表相同的点 X 这是一个冗余, 并将其从相邻相机的深度图中删除。

最后, 所有深度图都被反向投影为 3D, 并合并为单个点云。最终的点云通常非常密集, 尤其是在使用高分辨率图像时。如果我们想让它变得稀疏, 我们可以简单地将投影点返回到深度图中的稀疏位置。例如, 仅使用图像位置处的点 $(2, 2n)$ 深度图中的点云大小大约会将点云的大小减小到使用所有点的大小的四分之一。这为我们提供了一种根据内存和存储限制来控制点云大小的方法。

四. 乙实验右结果

A. 两种最先进的比较方法

我们将我们的方法与两种最先进的方法进行了比较[11]、[22]。PMVS [11] 方法是一种基于特征点生长的方法, 它重复扩展一组稀疏的匹配关键点, 并使用可见性约束来过滤掉错误匹配。基于 DAISY 的方法 [22] 是一种用于超高分辨率图像集的基于深度图合并的方法。这个方法和我们的类似, 但是它立体

表一
磷参数S的会议磷罗波塞德中号方法

范围	部分	描述	价值
k_1	三、A	相邻图像的最大数量	10
w	三、B	匹配窗口的大小, $w \times w$ 平面细化的迭	7像素
k_2	三、B	代次数 随机平面分配的深度范围 随机平面	3
λ	三、B	分配的角度范围 随机平面分配的角度范围	$\frac{\text{最大}-\text{最小}}{4}$
θ	三、B		90°
ϕ	三、B		15°
k_3	三、B	随机分配的重复时间 可靠点的匹配成本	6
τ_1	三、B	阈值	0.3
τ_2	III.C 和 III.D	测量深度接近度的阈值 一致相邻图像的最小	0.01
k_4	三、C	数量	2

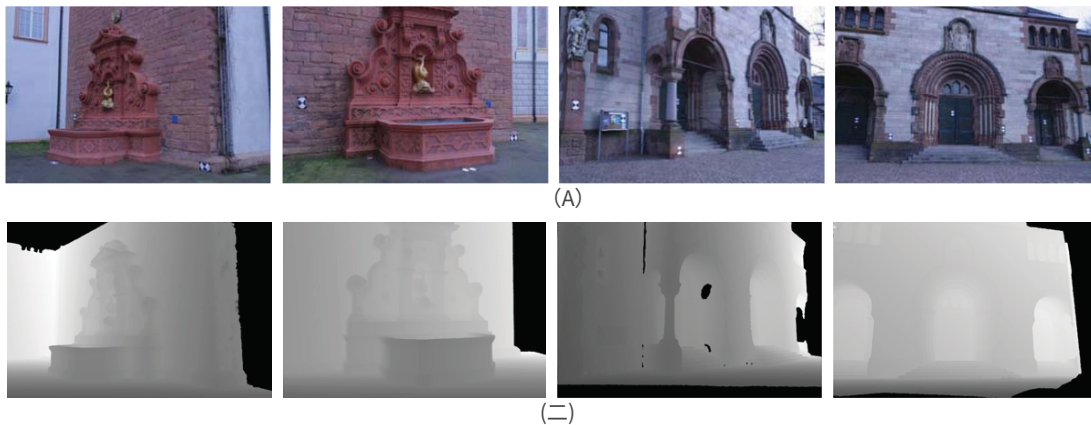


图 5. 基准数据集的样本图像及其地面真实深度图。在 (a) 和 (b) 中, 左边两幅图像来自 Fountain-P11, 右边两幅图像来自 Herz-Jesu-P8。

匹配过程基于每个像素的 DAISY [24] 描述符。

这两种方法现在被认为是最先进的 MVS 方法, 可以用于像我们这样的大规模场景。在接下来的三小节中, 所提出的方法以及 PMVS 和 DAISY 方法在不同的数据集上进行了定量和定性评估。所有实验均在具有 16G RAM 的 Intel 2.8GHz 四核 CPU 上实现。

B. 参数设置

所提出的方法有九个参数, 我们已经在第三节中讨论了它们的值设置。表一是一个总结。

DAISY方法的关键是DAISY描述符, 我们将DAISY参数设置为: $r=8$ 、 $m=2$ 、 $时间=4$ 、和 $H=4$ 。这些参数的详细信息可以参考[22]、[24]。[22]的作者建议首先计算图像稀疏位置的深度, 以限制相邻像素的搜索范围。因此, 在实验中, 我们首先使用图像上 10 个像素的采样步长选择控制像素, 并计算它们的深度。然后, 我们计算其他像素的深度, 其深度范围受其最接近的四个相邻控制像素的限制。

对于大量高分辨率图像, PMVS 方法可能会耗尽内存, 因此我们使用集群

PMVS [13] 的版本, 首先将输入图像分解为一组大小可管理的图像簇, 然后在每个簇上分别运行 PMVS。[11]、[13]的作者提供了PMVS和clustered-PMVS的源代码, 我们将其参数设置为: $乐v埃尔=0$ 、 $CS/泽=1$ 、 $临界点=0.7$ 、 $w/尺寸=7$ 、 $最小图像数量=3$ 。 $乐v埃尔$ 指定用于计算的图像金字塔中的级别, 以及 $乐v埃尔=0$ 表示使用原始分辨率图像。 $CS/泽$ 控制重建的密度, 并且 $CS/泽=1$ 表示软件尝试在每个像素中重建一个补丁。 $临界点=0.7$ 是光度一致性测量的阈值, 与 1 相同 - n 在我们的方法中。 $最小图像数量=3$ 意味着每个 3D 点必须在至少 3 张图像中可见才能重建, 这是作者建议的。这些参数设置确保 PMVS 方法尝试在全分辨率图像的每个像素处重建 3D 点, 这与我们的相同。有关这些参数的完整描述, 可以参考[11]、[13]、[38]了解详细信息。

所提出的方法以及 DAISY 方法可以在图像级别轻松并行化, 即单独计算每个深度图。因此, 在这两种方法中我们使用四个核心进行并行计算。对于PMVS方法, 我们设置其参数 $中央处理器=4$ 表示在我们的四核 CPU 上使用四个核心的代码。

由于内存消耗是大规模重建的关键问题, 我们分析了

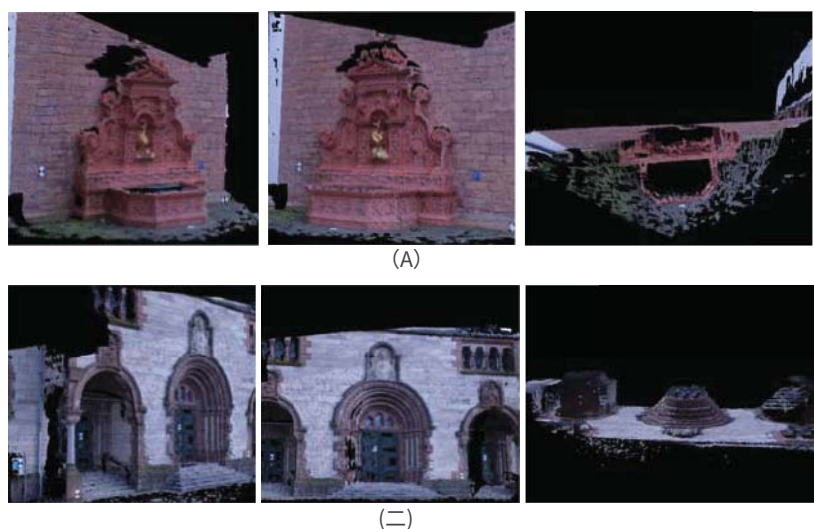


图6.所提出的方法在基准数据集上的最终重建结果（彩色点云渲染）。在（a）和（b）中，结果都是从三个不同的视点呈现的（右侧的视点是从顶视图看到的）。

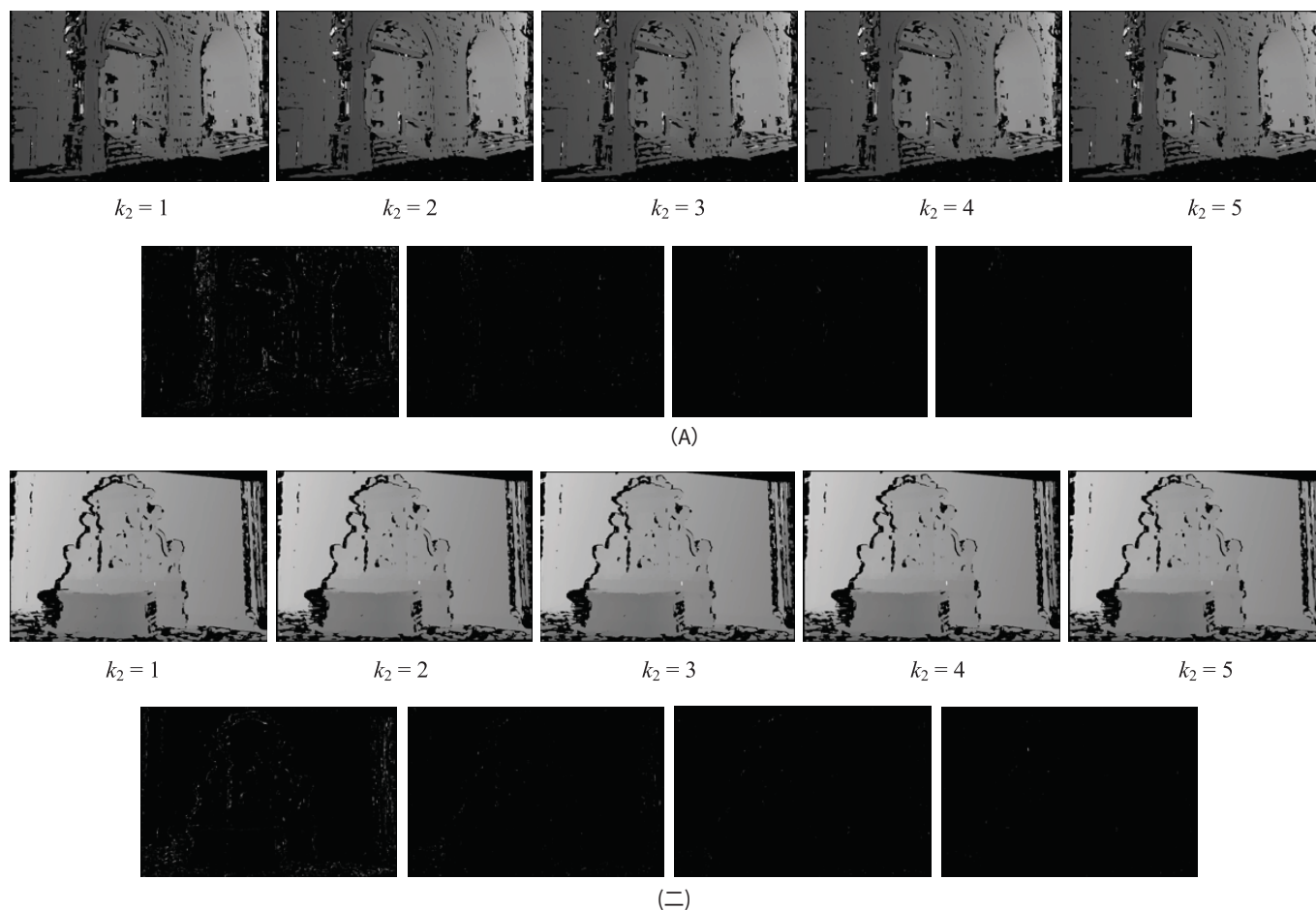


图7. 之后生成的深度图 $k_2 = 1, 2, \dots$ ，深度图计算过程中的5次迭代。在（a）和（b）中，顶行分别是一到五次迭代后生成的深度图，底行是相邻迭代之间的绝对深度差。

三种评价方法。PMVS方法需要同时加载所有图像（或集群中的图像），这意味着它需要 $H \times W \times 4 \times n$ 的记忆，在哪里 H 和 W 分别是图像的高度和宽度，4表示

四个字节，因为彩色图像转换为单精度浮点灰度图像，并且 n 是图像集的大小。显然，随着图像数量的增长，PMVS方法可能会遇到可扩展性问题（内存不足）。在

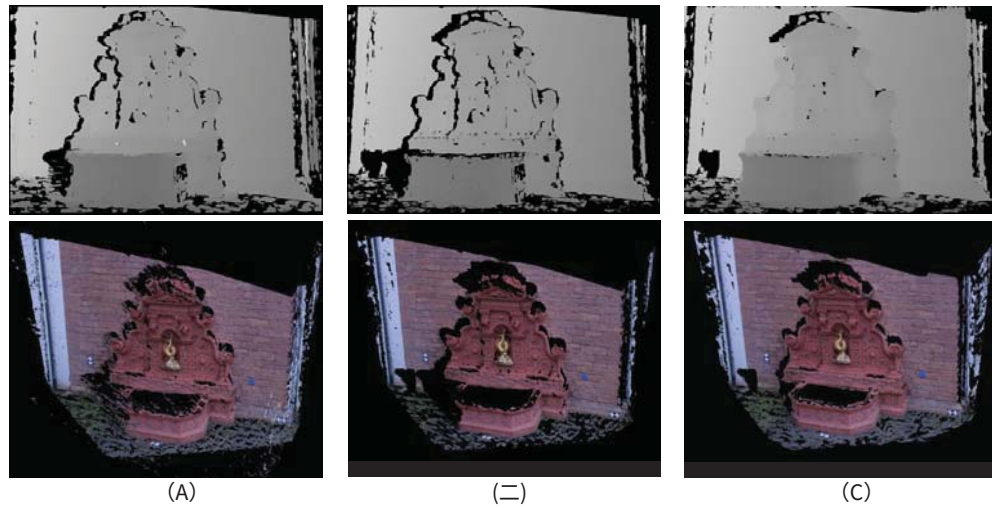


图 8. Fountain-P11 中第四张图像每个步骤后的深度图和反向投影 3D 点（彩色渲染）。在 (a)-(c) 中，顶部图像是深度图，底部是反向投影的 3D 点。

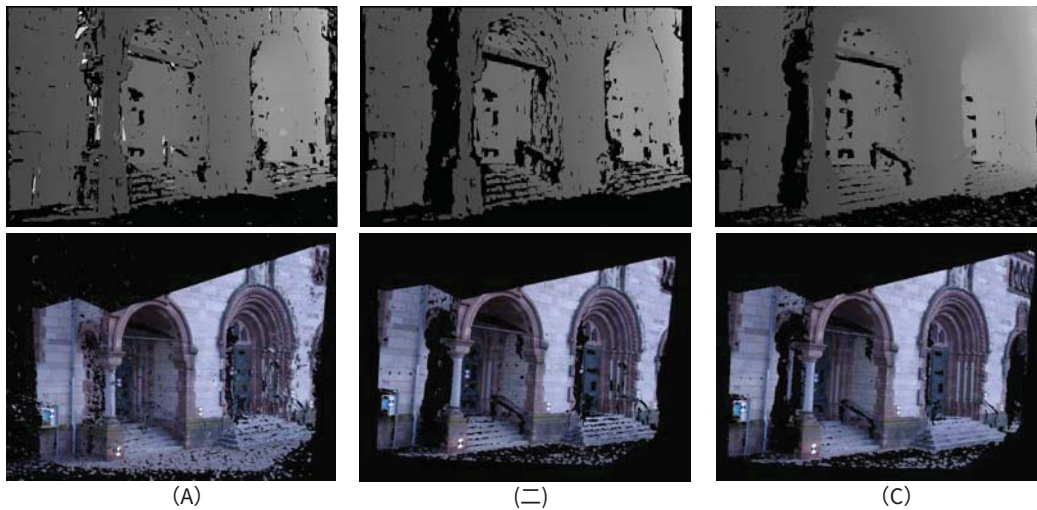


图 9. Herz-Jesu-P8 中第二张图像每一步后的深度图和反向投影 3D 点（彩色渲染）。在 (a)-(c) 中，顶部图像是深度图，底部是反向投影的 3D 点。

相反，所提出的方法和 DAISY 方法可以避免这种可扩展性问题，因为每个深度图都是单独计算和细化的。DAISY 方法加载两个图像并计算每个像素上的描述符以进行深度图计算。每个描述符需要 36 个浮点数，这意味着 DAISY 方法需要 $H \times W \times 36 \times 4 \times 2 = H \times W \times 288$ 内存。该方法还加载两个图像进行深度图计算，并且需要 $H \times W \times 4 \times 2 = H \times W \times 8$ 内存。显然，与 DAISY 和 PMVS 方法相比，该方法的内存需求要低得多。

C. 基准数据的定量评估

为了定量评估我们的方法，Strecha 等人提供了两个基准数据集 Fountain-P11 和 Herz-Jesu-P8。使用[39][40]。Fountain-P11和Herz-Jesu-P8分别有11张和8张图像，数量为 3072×2048 分辨率（6 兆像素）。图5（a）显示了数据中的一些样本图像

表二
氮数以百计的C正确与乙错误率磷IXELSA富特乙乙酰碱酯酶抑制剂S为
F奥恩我法师进来F奥坦-P11以及S第二次我法师进来
HERZ-J埃苏-P8

步	喷泉-P11		赫兹-耶稣-P8	
	正确像素	错误像素	正确像素	错误像素
深度图 计算	5 174 163	737211	4 477 824	990058
深度图 细化	4 603 032	165868	3 992 684	176 736
深度图 合并	5 254 683	260 351	4 306 332	266224

套。地面实况是通过激光扫描（LIDAR）获得的，这是一个单一的高分辨率三角形网格模型。基准数据站点[40]可以用地面实况定量评估三角网格模型中的MVS结果。这

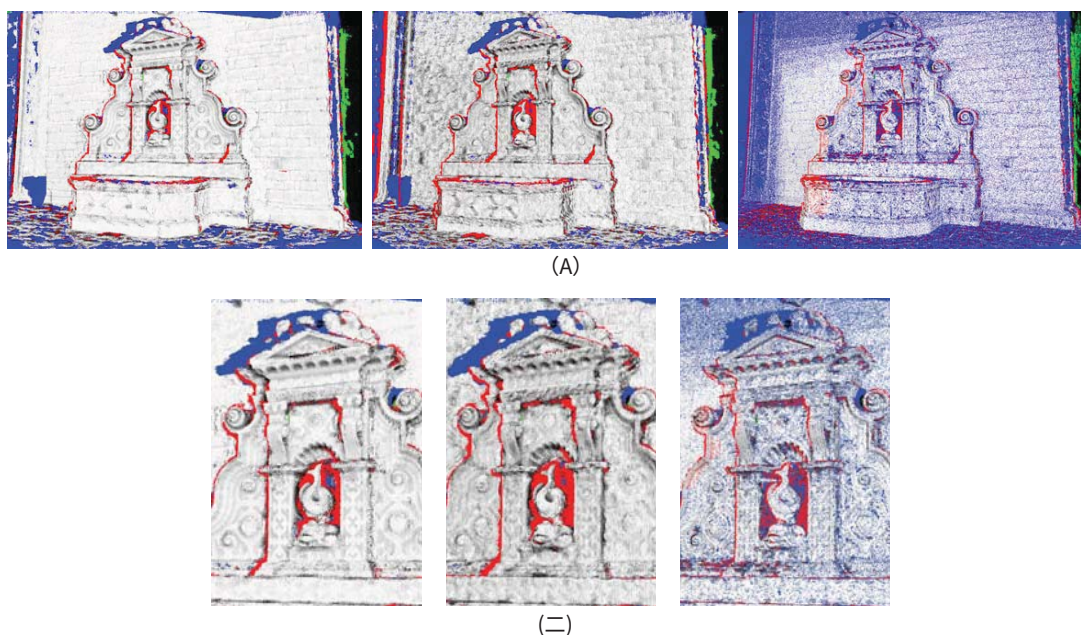


图 10. (a) 和 (b) 中，从左到右：分别使用所提出的方法、DAISY 方法和 PMVS 方法在 Fountain-P11 中的第四张图像的深度误差图。在所有图像中，蓝色像素通过 MVS 方法编码缺失的深度值，绿色像素编码缺失的地面真实数据，红色像素编码错误 e 比大 t_e ，以及误差在 0 和 0 之间的像素 t_e 以灰色编码 [255,0]。

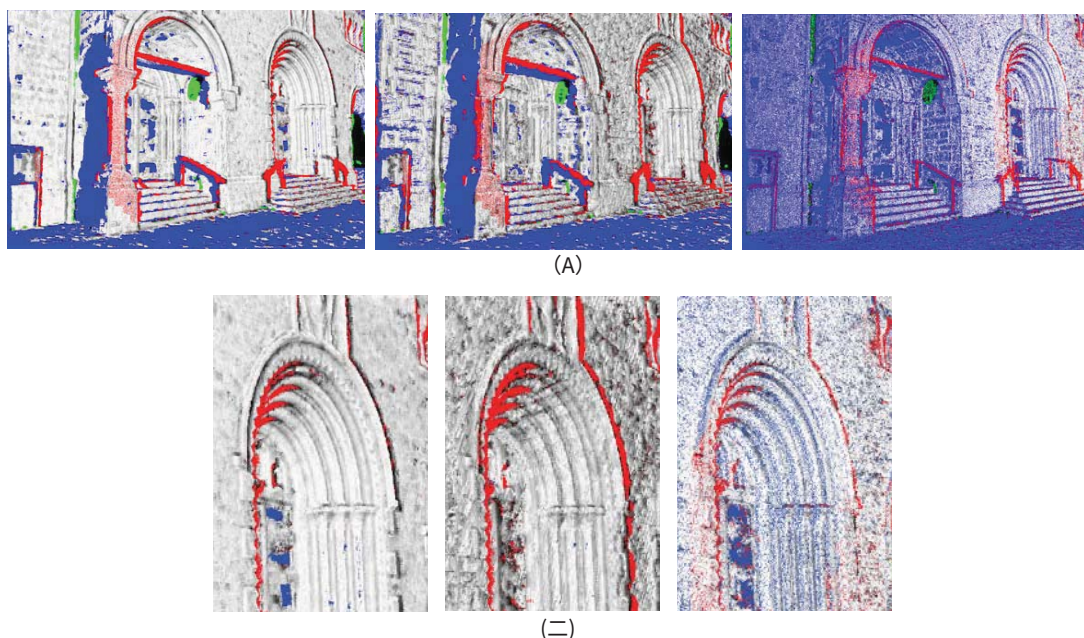


图 11. (a) 和 (b) 中，从左到右：分别使用所提出的方法、DAISY 方法和 PMVS 方法在 Herz-Jesu-P8 中第二张图像的深度误差图。在所有图像中，蓝色像素通过 MVS 方法编码缺失的深度值，绿色像素编码缺失的地面实况数据，红色像素编码错误 e 比大 t_e ，以及误差在 0 和 0 之间的像素 t_e 以灰色编码 [255,0]。

使用某种网格划分算法可以轻松地从点云生成网格模型 [41]。然而，本节中的三种评估方法都是输出 3D 点，因此更直接的方法是以点形式而不是以细化网格形式比较原始输出。为了使这种比较可行，我们首先将地面实况投影到每个图像以生成地面实况深度图。由于地面实况模型是 3D 三角网格形式，因此地面实况

每个像素的深度是通过计算第一个三角形与穿过该像素的相机光线相交的深度从 3D 三角形网格获得的。在此过程之后，生成地面真实深度图。图 5(b) 显示了一些地面真实深度图示例。

三种评估的 MVS 方法用于使用第 IV.B 节中给出的参数在基准数据上重建点云，并且由

表三

氮数以百计的C正确与乙错误率磷IXELSU唱歌时间重稀土元素乙估值中号方法F奥思我法师进来F奥坦-P11以及S第二次
我法师进来HERZ-J埃苏-P8

方法	喷泉-P11			赫兹-耶稣-P8		
	正确的像素	错误像素	错误/正确	正确的像素	错误像素	错误/正确
建议的方法	5 254 683	260 351	4.9%	4 306 332	266 224	6.2%
维菊	5 163 432	263 544	5.1%	4 107 572	382 104	9.3%
PMVS	3 853 304	246 696	6.4%	2 838 744	346 312	12.2%

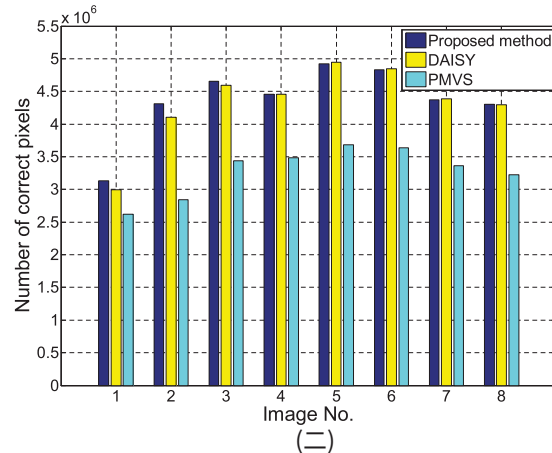
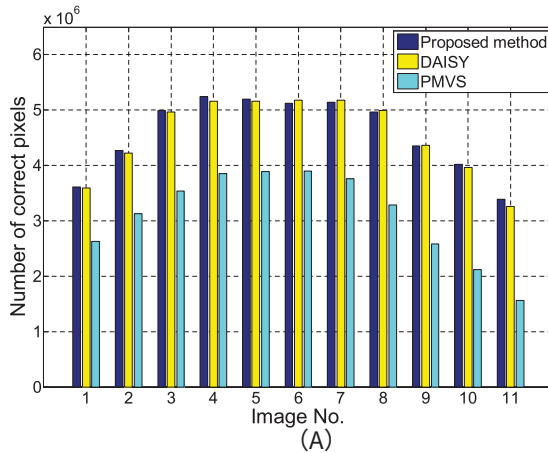


图 12. 使用三种评估方法的正确像素数。对于每个像素, 如果深度误差 e 在下面 $\tau_e = 0.01$ 。(a) 是 Fountain-P11 的结果, 其中包含 11 个图像。(b) 是 Herz-Jesu-P8 的结果, 其中包含八张图像。

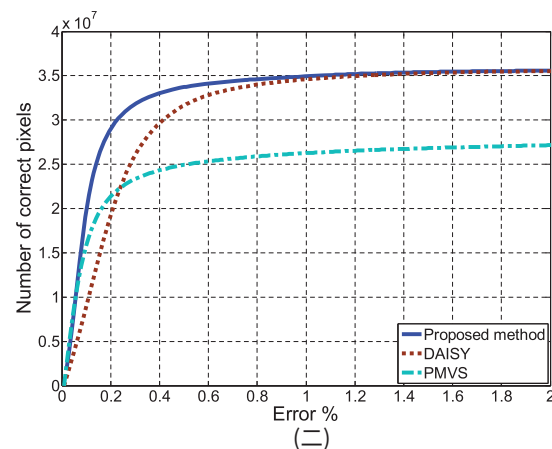
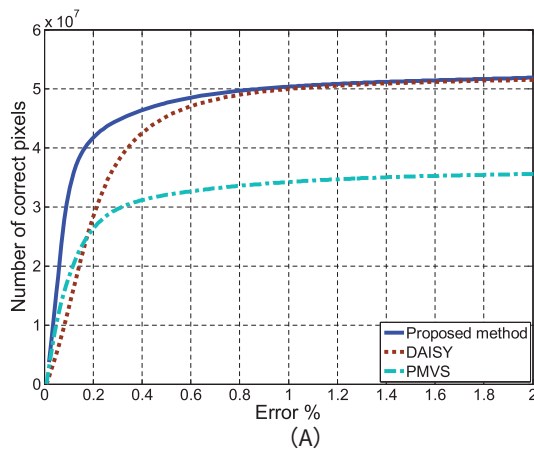


图 13. 所有图像中正确像素的总数作为错误阈值的函数。(a) Fountain-P11 的结果。(b) Herz-Jesu-P8 的结果。

所提出的方法如图6所示。然后我们将通过不同方法计算的点云投影到每幅图像上, 以与地面真实深度图进行定量评估。

对于图像中的每个像素, 我们将 MVS 方法计算的深度表示为 d 并表示真实值的深度 d_{GT} , 计算深度和地面真实值之间的深度误差可以测量为:

$$e = \frac{\|d - d_{GT}\|}{d_{GT}}$$

如果深度误差 e 低于阈值 τ_e , 深度 d 被认为是正确的 (在本文中我们设置 $\tau_e = \tau_2 = 0.01$)。

等式. 8 是关于如何进行定量测量准确的重建深度为, 以下评价均基于该测量。

第一个实验说明了迭代次数的影响 k_2 用于深度图计算过程中的平面细化, 并说明了我们选择的原因 $k_2 = 3$ 为迭代次数。我们选择 Fountain-P11 中的第四张图像和 Herz-Jesu-P8 中的第二张图像, 并使用不同的值计算它们的深度图 k_2 ($k_2 = 1, 2, \dots, 5$) 如图所示

(8) 图7。在图7(a)和图7(b)中, 顶行是深度-分别经过一到五次迭代后生成的图, 底行是相邻迭代之间的绝对深度差。结果表明, 相当多的深度



图 14. 大数据集的示例图像。左边两张图片来自赫尔，右边两张图片来自生命科学大楼。

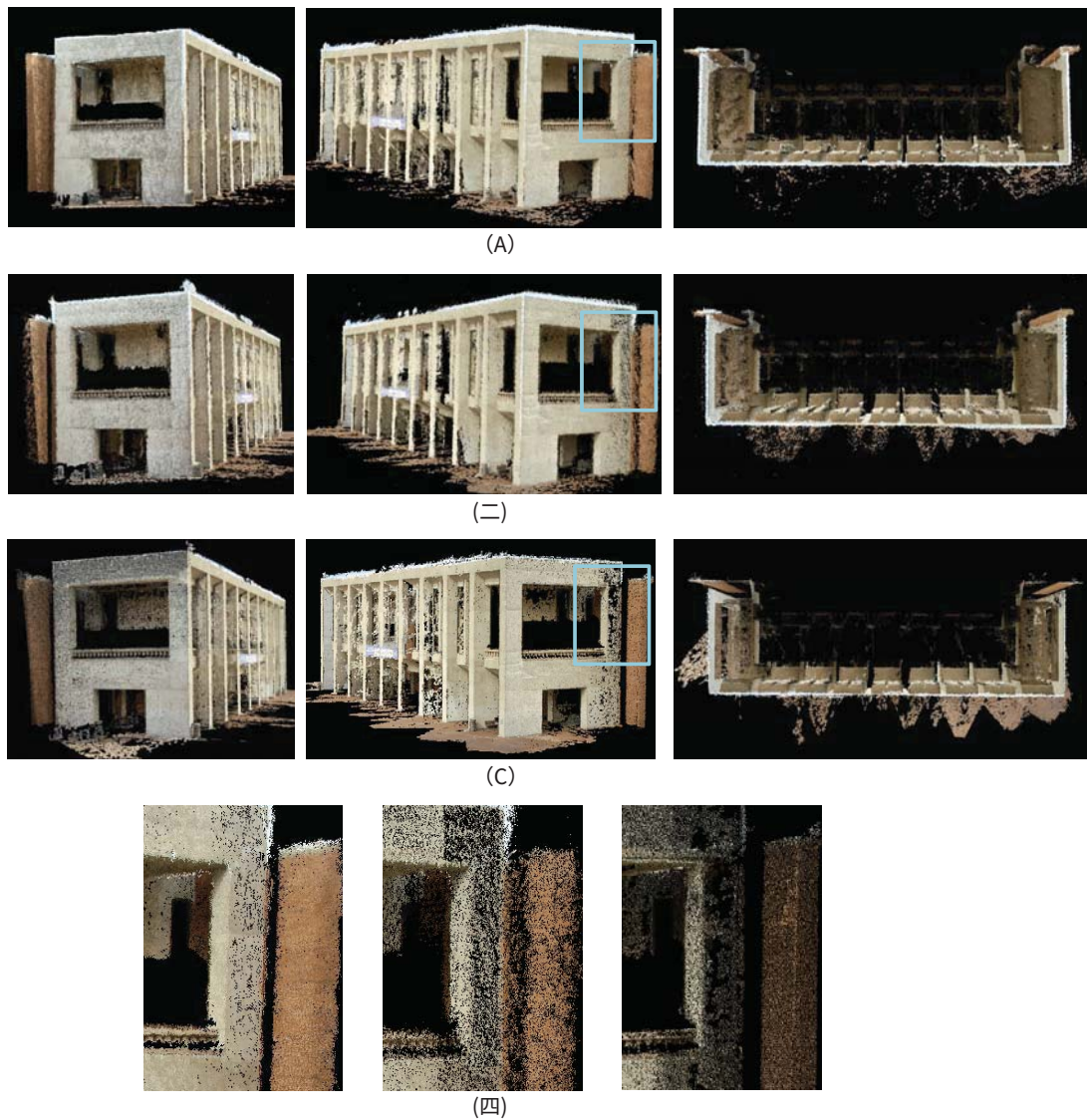


图 15. 三种评估方法在 Hull 数据集上的最终重建结果（彩色点云渲染）。在 (a) - (c) 中，结果是从三个不同的视点呈现的（右侧是从顶视图看到的）。

3次迭代后可以发现变化，因此设置 $k_2=3$ 可以在两者之间提供良好的平衡准确性和效率。

该方法分为三个步骤：深度图计算、深度图细化和深度图合并。为了说明每个步骤的效果，我们显示深度图

分别是图 8 和图 9 中 Fountain-P11 中的第四个图像和 Herz-Jesu-P8 中的第二个图像在每个步骤后的反向投影 3D 点和背投影 3D 点。除了视觉结果之外，表 II 中还给出了每个步骤之后基于等式 1 中给出的测量的正确和错误像素的数量。8. 结果表明，在深度图计算之后，基于补丁的立体

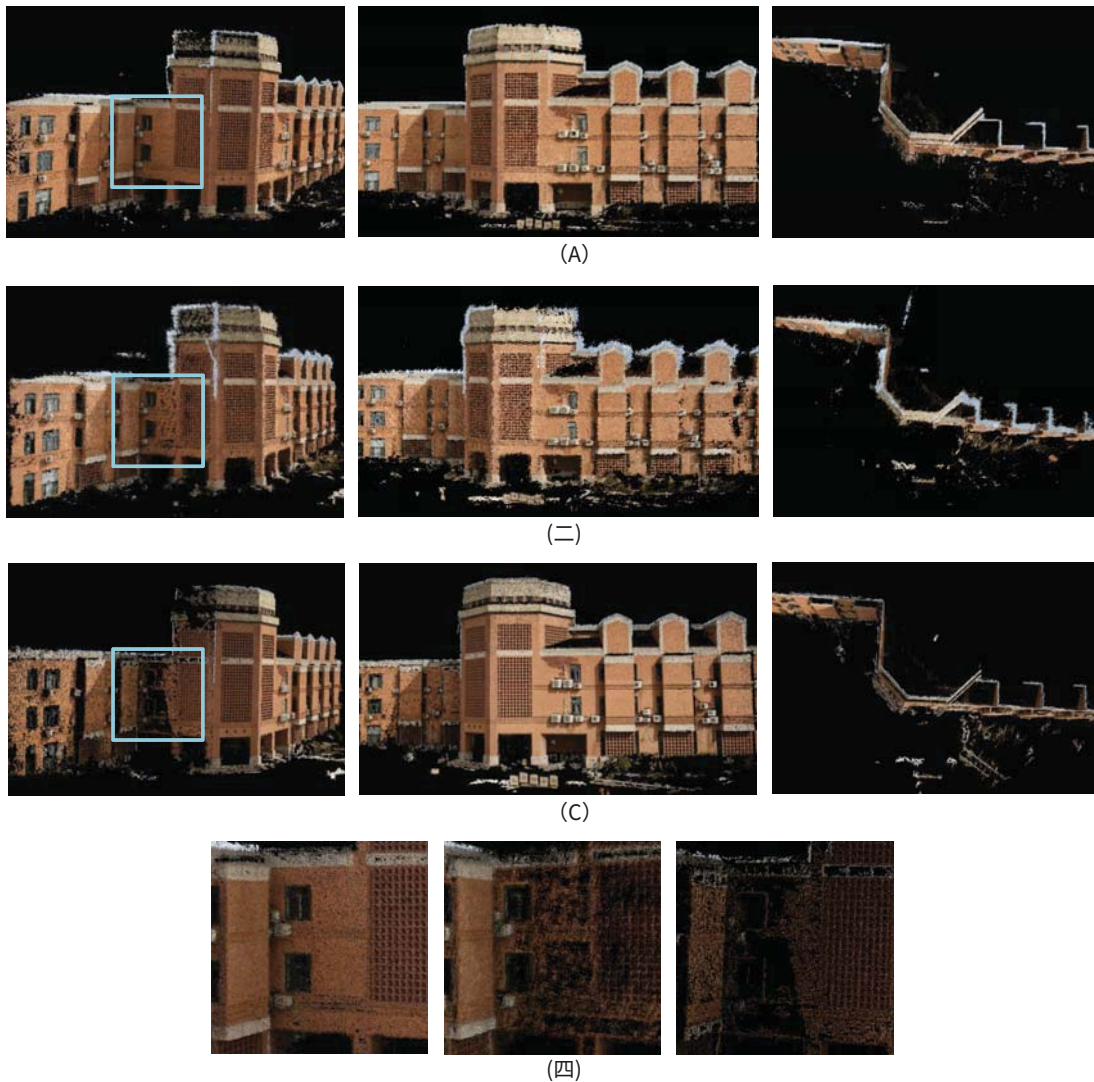


图 16. 三种评估方法在生命科学大楼数据集上的最终重建结果（彩色点云渲染）。在（a）-（c）中，结果是从三个不同的视点呈现的（右侧是从顶视图看到的）。

可以生成可接受的深度图，但仍然包含某些可见错误。经过深度图细化后，Fountain-P11 和 Herz-Jesu-P8 中分别去除了 77% 和 82% 的错误像素，从而得到相对干净的点云。最后，在深度图合并后，一些孔被填充，如图 8 中喷泉底部的左侧部分。

图10和图11分别显示了使用三种评估方法的Fountain-P11中的第四图像和Herz-Jesu-P8中的第二图像的深度误差图。在这些图中，蓝色像素通过 MVS 方法对缺失的深度值进行编码，绿色像素对缺失的地面实况数据进行编码，红色像素对错误进行编码 e 比大 τ_e ，以及误差在 0 和 0 之间的像素 τ_e 以灰色编码 [255,0]。结果表明，我们的方法和 DAISY 方法可以生成比 PMVS 方法更密集的点。尽管 PMVS 的参数已设置为尝试在每个像素处重建 3D 点，但它仍然留下大量没有深度的像素。与 DAISY 方法相比，该方法的误差图更小，可以获得更准确的结果。更亮

比 DAISY 方法的方法，并且更亮意味着深度误差更小。表 III 通过计算正确和错误像素的数量对图10和图11所示的结果进行定量评估。结果表明，与 DAISY 和 PMVS 方法相比，该方法不仅产生更多的正确像素，而且具有更低的错误/正确率。

除了单个图像之外，我们还计算数据集中所有图像的深度误差，并评估三种评估方法生成的重建结果的整体质量。对于每个图像，我们计算深度误差低于的正确像素的数量 τ_e 结果如图 12 所示。结果表明，在所有图像中，本文方法和 DAISY 方法的正确像素数几乎相同，约为 1。比 PMVS 方法大 5 倍。

为了进一步评估重建精度，我们将所有图像中正确像素的总数计算为误差阈值的函数 τ_e 。我们设置 τ_e 至 200 个值的均匀性

表四

C计算型时间输入法时间重稀土元素乙估值中号方法（米伊努特斯）

数据集	提议的方法	雏菊	PMVS
Fountain-P11 (11张图片)	9. 5	11. 9	127
Herz-Jesu-P8 (8张图片)	7. 1	7. 5	106

表五

C计算型时间输入法时间重稀土元素乙估值中号方法（米伊努特斯）

数据集	拟议论文	雏菊PMVS
船体 (61 张图片)	45. 3	49. 2 621
生命科学大楼 (102 张)	81	77. 9 1579

分布在 $[0, 0.02]$ ，结果如图13所示。结果表明，当误差阈值 τ_e 相当小（低于0.002），所提出的方法和PMVS方法比DAISY方法可以获得更多的正确像素，这表明如果我们关注高精度的重建结果，所提出的方法和PMVS都优于DAISY方法。这个结果来自三种评估方法的性质。所提出的方法和PMVS方法依赖于连续域中平面位置和法线的细化，但是DAISY方法在沿着极线的离散像素位置处匹配DAISY描述符，这导致空间中的离散深度，并且这种离散性质会影响它的准确性。

最后，我们评估不同方法的速度。首先我们分析所提出方法的计算复杂度。深度图计算、细化和合并的复杂性是 $O(HWBD)$, $O(HWk_1)$ ，和 $O(HWk_1)$ 分别是，其中 H 和 W 分别是图像的高度和宽度， B 是匹配窗口的大小， D 是要测试的深度数，并且 k_1 是相邻图像的最大数量。在本文中， $B=w \times w=49$ 、 $k_1=10$ 。显然，深度图细化和合并的复杂度远低于深度图计算的复杂度。对于深度图计算步骤中的每个像素，我们在开始时计算一次其成本聚合，然后进行三次迭代以细化平面。在每次迭代中，我们使用其三个相邻像素的平面参数计算像素的成本聚合三次以进行空间传播，并计算六次以进行随机平面分配。这给出了每个像素要测试的深度数： $D=1+(3+6) \times 3=28$ 。与需要测试大量深度假设的传统立体匹配方法相比，所提出的方法可以显著降低计算复杂度。三种评估方法的速度如表IV所示。结果表明，该方法的运行速度与DAISY方法相似，比PMVS快约13至15倍。

D. 大数据集的定性评估

在本节中，我们在大数据集上测试所提出的方法。这里使用了两个数据集，一个是[38]提供的Hull数据集，包括61张图像， 3008×2000 （6兆像素）分辨率，另一个是

清华大学生命科学大楼数据集，包含102张图像， 3456×2304 （8兆像素）分辨率。这两个数据集的一些样本图像如图14所示。

我们在这些大数据集上定性评估三种不同的方法。一些重建结果如图15和图16所示，计算时间如表V所示。结果表明，三种方法都可以达到可接受的重建结果，但所提出的方法和PMVS方法的执行精度比从图16中可以更清楚地看到DAISY方法。与PMVS方法相比，我们的方法可以获得更完整的结果，如图15（d）和图16（d）中的墙壁。所提出的方法和DAISY方法在船体上的运行速度比PMVS快约13倍，在生命科学大楼上的运行速度快20倍。

VC结论

在本文中，我们提出了一种针对大规模场景的基于深度图合并的MVS方法，该方法同时采用**准确性**和**效率**考虑到。该方法的关键是基于高效补丁的立体匹配加上深度图细化过程，以增强多个视图的一致性。与最先进的MVS方法相比，所提出的方法具有三个主要优点：1）重建的点云非常准确和密集，因为基于补丁的立体能够生成具有可接受误差的深度图，并且可以进一步通过深度图细化过程进行细化。2）所提出的方法非常有效，在达到相似精度的同时，比PMVS方法快约10至20倍。3）它可以很容易地在图像级别并行化，即每个深度图都是单独计算的，这使得它适合于高分辨率图像的大规模场景重建。

右参考文献

- [1] SM Seitz, B. Curless, J. Diebel, D. Scharstein and R. Szeliski, “多视图立体重建算法的比较和评估”，载于**过程。IEEE 计算。苏克。会议。计算。维斯。模式识别。**，2006年6月，第519–528页。
- [2] SM Seitz and CR Dyer, “通过体素着色重建真实场景”，**国际。J. 计算机。维斯。**，卷。35，没有。2，第151–173页，1999年11月。
- [3] G. Vogiatzis, C. Hernandez, PH Torr 和 R. Cipolla, “通过体积图形切割和遮挡稳健的照片一致性实现多视图立体” **IEEE 传输。模式识别。马赫。英特尔。**，卷。29，没有。12，第2241–2246页，2007年12月。
- [4] SN Sinha, P. Mordohai and M. Pollefeys, “通过自适应四面体网格对偶上的图形切割实现多视图立体”，载于**过程。IEEE 国际。会议。计算。维斯。**，2007年10月，第1–8页。
- [5] O. Faugeras and R. Keriven, “变分原理、表面演化、偏微分方程、水平集方法和立体问题”，**IEEE 传输。图像处理。**，卷。7，没有。3，第336–344页，1998年6月。
- [6] C. Hernandez and F. Schmitt, “3D 对象建模的轮廓和立体融合”，**计算。维斯。形象理解。**，卷。96，没有。3，第367–392页，2004年12月。
- [7] VH Hiep, R. Keriven, P. Labatut and J.-P. 庞斯, “迈向高分辨率大规模多视图立体”，载于**过程。IEEE 计算。苏克。会议。计算机视觉。模式识别。**，2009年6月，第1430–1437页。
- [8] D. Cremers and K. Kolev, “通过凸域上的凸函数实现多视图立体和轮廓一致性” **IEEE 传输。模式识别。马赫。英特尔。**，卷。33，没有。6，第1161–1174页，2011年6月。
- [9] M. Lhuillier and L. Quan, “从未校准图像进行表面重建的准密集方法”，**IEEE 传输。模式识别。马赫。英特尔。**，卷。27，没有。3，第418–433页，2005年3月。

- [10] M. Goesele, N. Snavely, B. Curless, H. Hoppe 和 SM Seitz, “社区照片集的多视图立体”, 载于过程. *IEEE 国际. 会议. 计算. 维斯.*, 2007 年 10 月, 第 1-8 页。
- [11] Y. Furukawa 和 J. Ponce, “准确、密集且稳健的多视图立体视觉”, *IEEE 传输. 模式识别. 马赫. 英特尔.*, 卷. 32、没有. 8, 第 1362-1376 页, 2010 年 8 月。
- [12] T.-P. 吴, S.-K. Yeung, J. Jia 和 C.-K. Tang, “使用基于张量的多视图立体进行准密集 3D 重建”, 载于过程. *IEEE 计算. 苏克. 会议. 计算机视觉. 模式识别.*, 2010 年 6 月, 第 1482-1489 页。
- [13] Y. Furukawa, B. Curless, SM Seitz 和 R. Szeliski, “迈向互联网规模多视图立体”, 载于过程. *IEEE 计算. 苏克. 会议. 计算. 维斯. 模式识别.*, 2010 年 6 月, 第 1434-1441 页。
- [14] M. Goesele, B. Curless 和 SM Seitz, “多视图立体重温”, 载于过程. *IEEE 计算. 苏克. 会议. 计算. 维斯. 模式识别.*, 2006 年 10 月, 第 2402-2409 页。
- [15] C. Strecha, R. Fransens 和 LV Gool, “多视图立体中的组合深度和异常值估计”, 载于过程. *IEEE 计算. 苏克. 会议. 计算. 维斯. 模式识别.*, 2006 年 10 月, 第 2394-2401 页。
- [16] P. Merrell, A. Akbarzadeh, L. Wang, P. Mordohai 和 J.-M. Frahm, “基于实时可见性的深度图融合”, 过程. *IEEE 国际. 会议. 计算. 维斯.*, 2007 年 10 月, 第 1-8 页。
- [17] C. Zach, T. Pock 和 H. Bischof, “鲁棒 tv-l1 范围图像集成的全局最优算法”, 载于过程. *IEEE 国际. 会议. 计算. 维斯.*, 2007 年 10 月, 第 1-8 页。
- [18] D. Bradley, T. Boubekeur 和 W. Heidrich, “使用稳健的双目立体和表面网格进行精确的多视图重建”, 载于过程. *IEEE 计算. 苏克. 会议. 计算. 维斯. 模式识别.*, 2008 年 6 月, 第 1-8 页。
- [19] ND Campbell, G. Vogiatis, C. Hernandez 和 R. Cipolla, “使用多种假设改进多视图立体深度图”, 载于过程. *欧元. 会议. 计算. 维斯.*, 2008 年 10 月, 第 766-779 页。
- [20] Y. Liu, X. Cao, Q. Dai 和 W. Xu, “多视图立体的连续深度估计”, 载于过程. *IEEE 计算. 苏克. 会议. 计算. 维斯. 模式识别.*, 2009 年 6 月, 第 2121-2128 页。
- [21] J. Li, E. Li, Y. Chen, L. Xu 和 Y. Zhang, “多视图立体的捆绑深度图合并”, 载于过程. *IEEE 计算. 苏克. 会议. 计算. 维斯. 模式识别.*, 2010 年 8 月, 第 2769-2776 页。
- [22] E. Tola, C. Strecha 和 P. Fua, “高效的大规模多视图立体 [40] 多视角评估”, 卷. 23、没有. 5, 第 903-920 页, 2012 年。
- [23] D. Lowe, “来自尺度不变关键点的独特图像特征” *国际. J. 计算机. 维斯.*, 卷. 60, 没有. 2, 第 91-110 页, 2004 年。
- [24] E. Tola, V. Lepetit 和 P. Fua, “Daisy: 一种应用于宽基线立体声的高效密集描述符” *IEEE 传输. 模式识别. 马赫. 英特尔.*, 卷. 32、没有. 5, 第 815-830 页, 2010 年 5 月。
- [25] D. 盖洛普, J.-M. Frahm, P. Mordohai, Q. Yang 和 M. Pollefeys, “具有多个扫描方向的实时平面扫描立体声”, 载于过程. *IEEE 计算. 苏克. 会议. 计算. 维斯. 模式识别.*, 2007 年 6 月, 第 1-8 页。
- [26] M. Pollefeys, D. Nister, J.-M. 弗拉姆, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, SJ Kim, P. Merrell, C. Salmi, SN Sinha, B. Talton, L. Wang, Q. Yang, H. Stewenius, R. Yang, G. Welch 和 H. Towles, “根据视频进行详细的实时城市 3D 重建”, *国际. J. 计算机. 维斯.*, 卷. 72、没有. 2, 第 143-167 页, 2008 年。
- [27] 张刚, 贾建, T.-T. Wong 和 H. Bao, “从视频序列中恢复一致的深度图” *IEEE 传输. 模式识别. 马赫. 英特尔.*, 卷. 31、没有. 6, 第 974-988 页, 2009 年 6 月。
- [28] D. 盖洛普, J.-M. Frahm 和 M. Pollefeys, “用于城市场景重建的分段平面和非平面立体”, 载于过程. *IEEE 计算. 苏克. 会议. 计算. 维斯. 模式识别.*, 2010 年 6 月, 第 1418-1425 页。
- [29] N. Snavely, SM Seitz 和 R. Szeliski, “根据互联网照片集建模世界”, *国际. J. 计算机. 维斯.*, 卷. 80, 没有. 2, 第 189-210 页, 2008 年。
- [30] S. Agarwal, N. Snavely, I. Simon, SM Seitz 和 R. Szeliski, “一日之内建设罗马”, 载于过程. *IEEE 国际. 会议. 计算. 维斯.*, 2009 年 9 月, 第 72-79 页。
- [31] J.-M. 弗拉姆, P. 乔治, D. 盖洛普, T. 约翰逊, R. 拉古拉姆, C. 吴, Y.-H. Jen, E. Dunn, B. Clipp, S. Lazebnik 和 M. Pollefeys, “在万里无云的日子里建设罗马”, 载于过程. *欧元. 会议. 计算机视觉.*, 2010 年 9 月, 第 368-381 页。
- [32] M. Bleyer, C. Rhemann 和 C. Rother, “Patchmatch 立体声 — 具有倾斜支撑窗口的立体声匹配”, 载于过程. *英国人. 马赫. 维斯. 会议.*, 八月至九月. 2011 年, 第 14.1-14.11 页。
- [33] R. Hartley 和 A. Zisserman, *计算机视觉中的多视图几何*, 第二版. 英国剑桥: 剑桥大学. 出版社, 2004 年。
- [34] K.-J. 尹和 I.-S. Kweon, “用于视觉对应搜索的局部自适应支持权重方法”, 载于过程. *IEEE 计算. 苏克. 会议. 计算. 维斯. 模式识别.*, 2005 年 6 月, 第 924-931 页。
- [35] A. Hosni, M. Bleyer, M. Gelautz 和 C. Rhemann, “使用测地线支撑权重进行局部立体匹配”, 载于过程. *IEEE 国际. 会议. 图像处理.*, 2009 年 11 月, 第 2093-2096 页。
- [36] Q. Yang, L. Wang, R. Yang, H. Stewenius 和 D. Nister, “具有颜色加权相关性、分层置信传播和遮挡处理的立体匹配”, *IEEE 传输. 模式识别. 马赫. 英特尔.*, 卷. 31、没有. 3, 第 492-504 页, 2009 年 3 月。
- [37] N. Atzpadin, P. Kauff 和 O. Schreier, “立体门”。通过实时沉浸式视频会议的混合递归匹配,” *IEEE 传输. 电路系统. 视频技术.*, 卷. 14、没有. 3, 第 321-334 页, 2004 年 3 月。
- [38] 基于补丁的多视图立体软件 (PMVS - 版本 2)。 (2010) [在线]. 可用: <http://grail.cs.washington.edu/software/pmvs/>
- [39] C. Strecha, W. von Hansen, LV Gool, P. Fua 和 U. Thoennessen, “关于高分辨率图像的基准相机校准和多视图立体”, 载于过程. *IEEE 计算. 苏克. 会议. 计算. 维斯. 模式识别.*, 2008 年 6 月, 第 1-8 页。
- (2008) [在线]. 可用: <http://cvlab.epfl.ch/~strecha/multiview/denseMVS.html>
- [41] M. Kazhdan, M. Bolitho 和 H. Hoppe, “泊松曲面重建”, 载于过程. *第四届欧洲图表. 症状. 几何过程.*, 2006 年 7 月, 第 61-70 页。



沉书涵分别于2003年和2006年在中国成都西南交通大学获得学士和硕士学位, 并于2006年在中国成都获得博士学位。2010年毕业于上海交通大学, 获博士学位。

现任中国科学院自动化研究所模式识别国家实验室助理研究员。他目前的研究兴趣包括 3D 重建和基于图像的建模。